issue 1/2026

10n Journal

Translation in Software. Software in Translation



L10N Journal: Translation in Software, Software in Translation. 1/2026

Translation, Interpreting & Culture 2025: Translators, Interpreters, and Society

Editor-in-Chief
Marián Kabát

Deputy Editor-in-Chief
Mária Koscelníková

Editorial Assistants
Zuzana Hudáková and Milan Velecký

Technical Editors
Milan Regec and Milan Velecký

Language Editors
Authors and Zuzana Hudáková

Cover design
Zuzana Hudáková

Editorial
editorial@l10njournal.net

Publisher
STIMUL: the publishing and advisory center of the Faculty of Arts, Comenius University, Gondova 2, Bratislava, Slovakia, fif.stimul@uniba.sk, https://fphil.uniba.sk/stimul/

© *STIMUL and the authors of the articles, 2026*



This work is published under Creative Commons CC BY-NC-ND 4.0 international license. This license permits distribution of this work in its original, unaltered form for non-commercial purposes, with the appropriate credit given to the author. For more information about the license and the use of this work, see: https://creativecommons.org/licenses/by-nc-nd/4.0/

The authors of individual articles are responsible for translating quotes from foreign languages.

ISSN 2730-0757

Sponsors

The publication of this issue was financially supported by the Scientific Grant APVV-23-0539: Future-Proofing Translators and Interpreters: Establishing a Knowledge Platform for Language Professionals in Slovakia, and by the Scientific Grant Agency VEGA, under project No 1/0113/23 Localization of Software Products Specifics in the Slovak – English Language Pair.

Scientific Committee

Marianna Bachledová, Matej Bel University, Slovakia
Miguel Á. Bernal-Merino, University of Las Palmas de Gran Canaria, Spain
Oliver Carreira Martinez, Pablo de Olavide University, Spain
Dragoș Ioan Ciobanu, University of Vienna, Austria
Mikołaj Deckert, University of Lodz, Poland
Renée Desjardins, Université de Saint-Boniface, Canada
Martin Djovčoš, Constantine the Philosopher University, Slovakia
Ugo Ellefsen, Concordia University, Canada
Miguel A. Jiménez-Crespo, Rutgers University, USA
Hendrik J. Kockaert, KU Leuven, Belgium
Ralph Krüger, University of Applied Sciences, Germany
Dominik Kudła, University of Warsaw, Poland
Carme Mangiron, Universitat Autónoma de Barcelona, Spain
Pascaline Merten, Université Libre de Bruxelles, Belgium
Kristijan Nikolić, University of Zagreb, Croatia
David Orrego-Carmona, University of Warwick, United Kingdom
Emília Perez, Constantine the Philosopher University, Slovakia
Tomáš Svoboda, Charles University, Czech Republic

Contents

Mária Bakti, Márta Lesznyák, and Eszter Sermann 5

Explicitation operations in from-scratch human translation and in the post-editing of neural machine translation

Kateryna Bondarenko and Oleksandr Bondarenko 24

How language industry defines localization: a corpus-based study

Stavroula Paraskevi Vraila 39

Political Translation in the AI Era: Journalists as Dual Agents and Systemic Vulnerabilities in Cross-Linguistic Political Communication

Barbara Quaranta 62

Molisano to English: a qualitative case study of AI-translated culture-specific items in migration

Lucia Paulínyová 100

Šveda, Djovčoš, Perez (eds.). 2026. Confronting Digital Dilemmas in Translator and Interpreter Training. Routledge.

Bakti, Mária; Lesznyák, Márta; Sermann, Eszter. 2026. Explicitation operations in from-scratch human translation and in the post-editing of neural machine translation. In: L10N Journal. 1(5), pp. 5-23.

Explicitation operations in from-scratch human translation and in the post-editing of neural machine translation

Mária Bakti, Márta Lesznyák and Eszter Sermann

University of Szeged, University of Szeged; University of Szeged

bakti.maria@szte.hu, lesznyakm@t-online.hu, eszter.sermann.lingo@gmail.com

Abstract

Explicitation is considered a translation universal (Laviosa-Braithwaite 1998) and encompasses obligatory and optional operations that translators carry out in order to make what is implicit in the source language text explicit in the target language text (Englund Dimitrova 2005; Vinay & Darbelnet 1958/1995). Explicitation has become an umbrella term referring both to these operations and the resulting structures in the target language (Englund Dimitrova 2005). To date, there has been a limited body of research carried out into explicitation in machine translation or into the post-editing of machine translation (Khoury 2024, Krüger 2020). The aim of this paper is to compare and contrast explicitation operations carried out during from-scratch human translation and during the post-editing of neural machine translation output.

10 MA students of translation post-edited and 10 MA students of translation translated the same source language text, a research article abstract on bilingualism from English into Hungarian.

Our results show that both post-editors and translators carried out more explicitation operations than the neural machine translation engine. In addition, human translators carried out significantly more explicitation operations than post-editors. Finally, the findings revealed a work-mode-specific distribution of explicitation operations.

Keywords: explicitation, from-scratch human translation, post-editing of machine translation, target text error numbers, time on task

1 Introduction

Research into explicitation is one of the most vibrant fields of investigation in Translation Studies. Explicitation has become a widely researched, albeit controversial topic. It is considered a translation universal (Laviosa-Braithwaite 1998); however, some authors also question its universality (Baumgarten et al. 2008; Krüger 2014).

Although explicitation has attracted considerable research attention in Translation Studies, there is a limited number of investigations into machine-translated or post-edited target language texts (Khoury 2024; Krüger 2020).

This article aims to analyze explicitation operations in from-scratch human translation and in the post-editing of neural machine translation.

First, an overview is given of the Szeged Translation and MT Post-Editing Competence Research Centre's investigations into MT post-editing competence. This is followed by a literature review on explicitation, types of explicitation, and research relevant for our topic, namely investigations on reviser's explicitation operations and on explicitation in MT. In the next section, the methodology of our investigations is outlined. After that, the results are presented, followed by a summary and the implications of the research.

2 Overview of the investigations of the University of Szeged's Translation and MT Post-Editing Competence Research Centre

The University of Szeged's Translation and MT Post-Editing Competence Research Centre's investigations focus on the role different elements of competence play in from-scratch human translation and post-editing of MT. The theoretical foundation for the investigations is provided by PACTE's model of translation competence (2011, 2020). The Centre has carried out research into how the target language output of MA students of translation is influenced by source language (English) competence, domain knowledge, beliefs about translation, text type, expertise, task type (from-scratch human translation or post-editing of MT), and experience. Data collection was carried out in eight waves between May 2020 and May 2024. The data collected made product- and process-oriented analyses possible, the results of which have been published previously (Lesznyák et al. 2023, 2024, 2025).

Initially, 113 students entered the project, but some of them dropped out. In total, 106 students completed at least one task and produced 170 texts, 64 of which formed the longitudinal dataset. Of the 113 students who initially entered the study, 54 students (47.8%) had English as language B and 59 (52.2%) as language C. Independent samples t-tests showed no significant differences between the language competence of the two groups (Lesznyák et al. 2024; Lesznyák 2026), which means that they could be handled as a homogeneous group. Consequently, no further references will be made to students' working languages.

The research center's findings indicate that post-editing of machine translation output improves the performance of translation trainees only if they are familiar with the content, the genre and the register of the SL text. As concerns the link between

competence elements and performance, the research center's findings suggest that language competence appears to be a decisive factor, both in from-scratch human translation and the post-editing of machine translation, if the trainees have to deal with demanding and complicated texts.

Results related to translation students' perceptions of post-editing revealed a significant association between more realistic task assessment and better performance for unfamiliar genres and texts.

3 Explicitation

Explicitation is considered one of the most researched topics in Translation Studies. According to a broad definition, based on Vinay and Darbelnet (1958/1995), Klaudy states that explicitation is the technique of making explicit in the target text information that is implicit in the source text (Klaudy 1999). To fine-tune this circular definition, Klaudy adds that explicitation is a translation transfer operation in the course of which the translator expresses a source language unit in the target language more openly, more clearly or using more words than the author of the source language text (Klaudy 1999). The construct of explicitation was introduced into translation studies by Vinay and Darbelnet (1958/1995).

Research into explicitation has proliferated since the publication of Blum-Kulka's seminal article on the explicitation hypothesis (1986/2000), according to which, during translation, the level of explicitness of the TL text increases. The hypothesis was fine-tuned and complemented by Klaudy's asymmetry hypothesis (2001), which claims that translators favor explicitation over implicitation. A possible explanation provided by Klaudy is that Grice's (1975) cooperative principle operates in translation as well, but slightly differently from a conversation between two parties. In translation, the cooperative principle operates; however, the TL reader is not physically present in the communication situation. Thus, the translator cannot directly check whether the information rendered in the target language text is sufficient for the target reader to understand the message. As a result, translators might decide "to play it safe", in other words, they try to make sure that the target reader understands the target language text, using explanations, additions and extra information (Klaudy 2001).

Gradually, the meaning of explicitation in Translation Studies has expanded to refer to "a host of phenomena with certain aspects in common, are grouped together under the term explicitation, which tends to be used as a kind of umbrella term to label certain phenomena of differences between the ST and the TT which seem to be permissible in translation" (Englund Dimitrova 2005, 40).

In Translation Studies, different typologies exist to describe explicitation operations carried out during translation. Klaudy, for instance, differentiates between obligatory, optional, pragmatic, and translation-inherent types of explicitation operations (Klaudy 1999), while Englund Dimitrova works with the categories of norm-driven and strategic explicitation. Robin (2019), in her work on reviser's explicitation operations, identifies rule-based, norm-based and editorial explicitation. In her review of explicitation, Gumul (2006) lists twelve explicitation operations. These include (1) adding connectives, (2) the categorical shift of cohesive devices, (3) shifts from referential cohesion to lexical cohesion, (4) shifts from reiteration in the form of paraphrase to reiteration in the form of identical / partial repetition, (5) filling out elliptical constructions, (6) lexical specification, (7) adding modifiers and quantifiers, (8) adding a proper name to a generic name, (9) distributing the meaning of a ST unit over several units in the target text, (10) replacing nominalizations with verb phrases, (11) disambiguating metaphors, and (12) including additional explanatory remarks.

Explicitation operations in MT and the post-editing of MT have received limited research attention. One possible source to better understand explicitation operations in the post-editing of MT is research related to explicitation operations carried out by revisers. Robin (2019), in her investigation using qualitative and quantitative methods, analyzed ten English novels, their Hungarian draft translations and their revised versions, using the categories of rule-based, norm-based, and strategy-based explicitation and implicitation operations. The results of her contrastive text analysis indicate that translators, all in all, perform more explicitation and implicitation operations than revisers do. This leads us to hypothesize that post-editors would also carry out fewer explicitation operations than translators. Robin (2019) also found that translators favor explicitating over implicitating operations, which confirms Blum-Kulka's (1986/2000) explicitation hypothesis and Klaudy's (2001) asymmetry hypothesis. The same applies to revisers as well, especially in the case of lexical operations. Revisers seemed to employ specification operations; this way, they enriched the vocabulary of the target text, and also tried to curb redundancy. On the sentence level, revisers broke up sentences in order to help target readers to understand them. This operation was classified by Robin (2019) as a case of explicitation: "to enhance the readability by dividing complex sentences, thus employing explicitation" (Robin 2019, p. 135).

Krüger (2020) examined explicitation in neural machine translation, using the categories of lexical insertion, lexical specification, and relational specification. He compared the German renditions of an English source language text, namely the summary of a special report by the Intergovernmental Panel on Climate Change on carbon dioxide capture and storage published in 2005. A human translation was prepared of the source language text, together with a machine translated target text

produced by DeepL. Results show that the human translator carried out about three times as many explicitation operations as DeepL. There were also qualitative differences between the MT output and the from-scratch human translation, in terms of lexical specification and lexical insertion. It appears that DeepL has learnt to carry out some explicitation shifts, for instance explicitating source text structural ambiguities occurring on the syntagmatic level; however, DeepL struggled to insert autonomous lexical units into the target text when there is no explicit linguistic trigger provoking such a shift. In summary, Krüger states that the differences in explicitating behavior are due to the fact that translators follow strategies, while DeepL's explicitation performance is more arbitrary.

Khoury (2024) investigated optional pragmatic explicitation operations in the English-to-Arabic machine-translated and post-edited scientific texts. The neural machine translation engine employed in the study was Google Translate. The source language text was a report by the World Meteorological Organisation. The investigation focused on the translation of acronyms and scientific terms. The results indicate that Google Neural Machine Translation performed less reliably when translating specific terms and acronyms, while the post-edited target language texts were more readable.

To our knowledge, previous research has not examined the relationship between the frequency of explicitation operations and time on task either in from-scratch human translation or in post-editing. The question is important because post-editing is often seen as a means of reducing translation time. A significant correlation between time on task and the frequency of explicitation operations may suggest that explicitation requires extra effort on the part of the translator or the post-editor.

Another issue that might be of interest is the relationship between the number of errors in the target text and the number of explicitation operations. Data on this relationship could provide insight into how explicitation operations contribute to the quality of the TL texts during from-scratch human translation and post-editing.

Given the limited research on explicitation in post-editing, our study set out to answer the following research questions:

1. What is the frequency of explicitation operations in HT and PE?
2. How do post-editors change the explicitness of the MT output?
3. Is there a correlation between time on task and explicitation operations?
4. Is there a correlation between error numbers and explicitation operations?

4 Methodology

The participants of the investigation reported here were 20 first-year MA students in Translation and Interpreting at the University of Szeged. Their task was to translate an

abstract of an article on bilingualism; the abstract contained 127 words. The language direction was from English into Hungarian. The students were divided into two groups; 10 students carried out from-scratch human translation (FSHT), and 10 students post-edited a machine-translated text (PE). Students in the two groups were selected from a larger pool of participants and were matched on overall performance, based on the mean error numbers in the target language texts. Mean error numbers were as follows: FSHT = 24.4, PE = 24.35. In this way, we could answer the question of how translated and post-edited texts of comparable quality differed in terms of explicitation.

Data collection was carried out in 2020, 2021, and 2022. On all three occasions, the students completed the task in class. The post-editors were invited to post-edit a Hungarian MT output of the English source language text produced by DeepL. The MT output was generated in 2020 and was used throughout the study to ensure comparability of data collected at different points in time. The post-editing task was full post-editing, without the use of TM software. Students had no previous experience with PE and were not introduced to what full and light PE entailed. Nevertheless, they were instructed to produce a publishable text. In addition, they did not participate in any specific training in post-editing prior to the task. It should be noted here that at the time of the data collection, post-editing was not integrated into the MA program yet. In order to protect students' privacy, pseudonyms were used.

5 Taxonomy used

The Hungarian target language texts were analyzed using a taxonomy developed by our research group, based on the taxonomies of Gumul (2006) and the categories identified during the pilot analysis.

The categories taken from Gumul (2006), from her comprehensive analysis of explicitation operations, included adding connectives, categorical shift of cohesive devices, shifts from referential to lexical cohesion, filling out elliptical constructions, adding modifiers and quantifiers, distributing the meaning of a source-text unit over several units in the target language text, and adding explanatory remarks.

The pilot analysis revealed that, mainly because of the specificities of Hungarian, in the target language, certain TL-specific explicitation operations were also carried out during from-scratch translation and post-editing. These include word class change, verbalization with specification, personalization, and personalization with specification.

Personalization was defined as transforming a passive sentence into an active one, while explicitly naming the agent, that is, the subject of the active sentence. Personalization with specification was defined as transforming a passive sentence into

an active one, while explicitly naming the agent, that is, the subject of the active sentence, and using a more specific verb in the target language text than the verb in the source language text. Verbalization with specification was used to describe cases when the linking verb 'to be' in the source language texts was replaced by verbs with specific meaning in the target language texts. Word class changes, as the name suggests, were used to describe cases when a different word class was used in the target language to express the source language meaning. This is a somewhat broader category than Gumul's (2006) 'replacing nominalizations with verb phrases', which is considered part of the category of word class change. This new category was introduced because there are some language-specific word class changes in English to Hungarian that make the TL text more explicit. These were instances where, without word class change, the sentence was not only awkward but relatively hard to understand, too. That is what justified handling these changes as explicitation.

Table 1 below shows the categories used in our analysis of the text, with examples.

Table 1. ***Categories of explicitation operations used in our text analysis.***

Explicitation operation	**Example**
Categories adapted from Gumul (2006)	
adding connectives (ADD CONNECT)	0 > **and, so, thus, although**
categorical shift of cohesive devices	and they decided to wait > **so** they decided to wait
shift from referential cohesion to lexical cohesion (e.g. lexicalization of pro-forms) (SHIFT REF TO LEX)	overlap between them > overlap between **these sections**
filling out elliptical constructions	(...) some of the other consequences, and there were many of them, some **of the consequences were** very important
lexical specification (LEX. SPEC.)	say > **accuse**
adding modifiers and quantifiers (ADDING MOD/QUANT)	serious psychological damage a **strange** mixture
distributing the meaning of a source-text unit over several units in the target text	this double focus > this **division into two urban centres**
including additional explanatory remarks (EXPLAN REMARKS)	**the web page** liberty.org
Categories developed by the research group following the pilot analysis	
personalization	reported here > itt beszámolunk [here report we]
personalization with specification (PERS + SPEC)	were given > töltettek ki [they made them to fill out]

verbalization with specification	(the purpose) was (to investigate) > azt kívánta [it wished to]
word class change, including replacing nominalizations with verb phrases	conversational understanding> 1. megérteni 2. a beszélgetéseket [to understand conversations]

The explicitation operations in the Hungarian target language texts were each analyzed by two evaluators, aiming for 100% interrater agreement. The raters first categorized the data independently. In the next step, disagreements were discussed and resolved by consensus. The consensus classifications were used in the final analysis. Interrater statistics were not performed.

It is important to stress here that no quality check was performed. In other words, it was not evaluated whether the explicitation operation improved or harmed the target text. Explicitation operations were coded in QDA Miner Lite, and statistical analyses were performed with SPSS v. 24.

6 Results

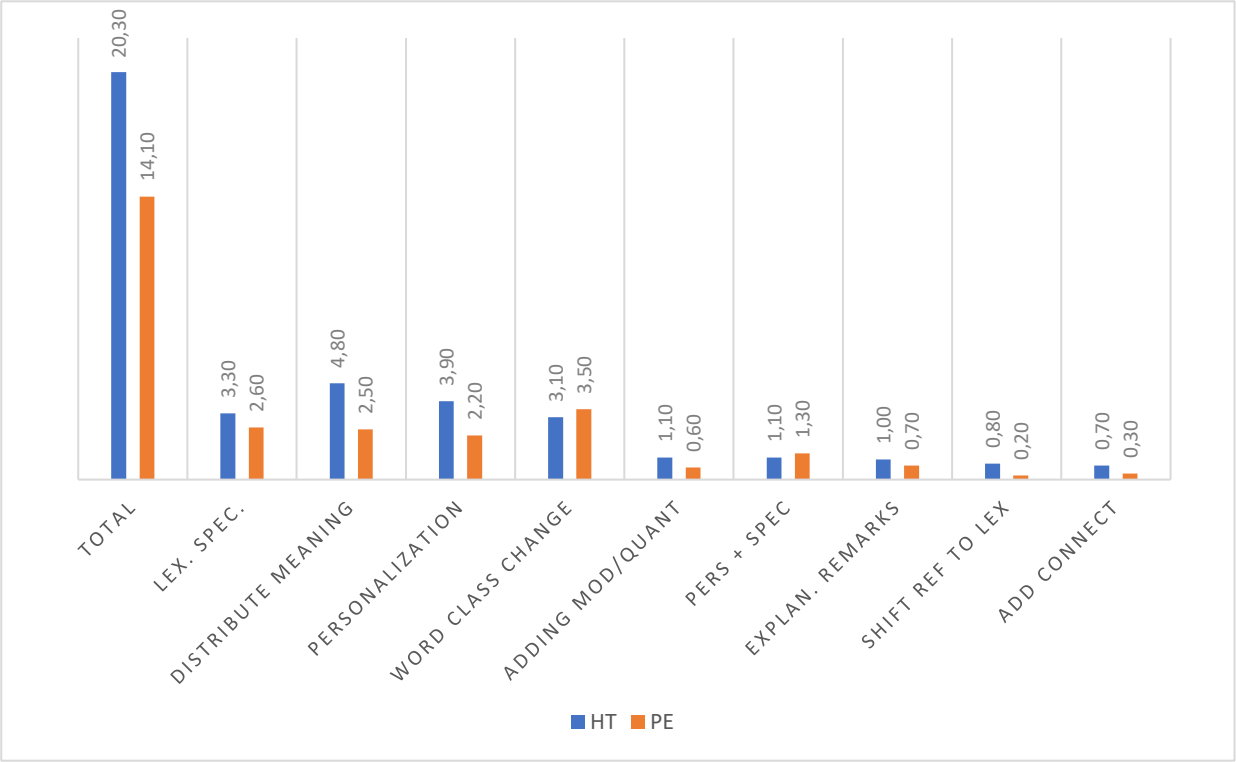
6.1 The frequency of explicitation operations in human translation and post-editing

Our first research question focused on finding out how often translators and post-editors use explicitation operations and whether there are differences between the two groups.

Figure 1 shows that, overall, human translators carried out significantly more explicitation operations than post-editors (Human translation: M = 20.30, SD = 5.755; Post-editing: M = 14.10, SD = 4.909; t(18) = 2.592, p = .018). If one looks at the different categories, it can be seen that the differences are statistically significant for the categories of distributing the meaning of one SL for several TL units (Human translation: M = 4.80, SD = 1.813; Post-editing: M = 2.50, SD = 1.509; t(18) = 3.083, p = .006) and shift from referential to lexical cohesion (Human translation: M = 0.80, SD = 0.632; Post-editing: M = 0.20, SD = 0.421; t(18) = 2.496, p = .022). In both categories, human translators engaged in significantly more operations than post-editors. On the other hand, post-editors carried out more word class changes (Human translation: M = 3.10, SD = 2.079; Post-editing: M = 3.50, SD = 1.434; t(18) = -0.501, p = .623) and personalization with specification operations (Human translation: M = 1.10, SD = 0.568;

Post-editing: M = 1.30, SD = 1.059; t(18) = -0.526, p = .605), but the differences did not reach statistical significance for these categories.

Figure 1. ***The most common explicitation operations in post-editing and from-scratch human translation. The coding labels are explained in Table 1.***



Furthermore, our data suggest that a work-mode-specific explicitation operation pattern exists. In human translation, the most frequent explicitation operations are distributing meaning, followed by personalization, word class change and lexical specification. In contrast, in the post-editing working mode, the frequency of explicitation operations is somewhat different, with word class change ranking first, followed by lexical specification, distributing meaning and personalization. See Table 2. It must be noted, however, that because of low frequencies in the individual categories, no statistical analysis was attempted.

These results are partly in line with Robin's (2019) findings, according to which revisers favor specification operations. In addition to personal preferences or style, we assume that differences between human translation and machine translation may also contribute to the differences found in the frequency of explicitation operations. Although post-editors did introduce new explications, they relied heavily on the MT output, keeping a relatively large proportion of MT explicitations (see next section). Consequently, the figures in the Post-editing column in Table 2 show a mix of MT-

introduced and human-introduced explicitation operations, whereas the Human translation column contains only operations introduced by human translators.

Our data suggests that human translators are meaning-focused and, at the same time, tend to spread out meaning to make sure that the reader understands the text (distributing meaning). In contrast, the post-edited texts contain a higher proportion of one-to-one solutions, like word class change and lexical specification. This pattern may partly reflect the influence of the MT-output rather than the preference of the post-editors themselves, as post-editors may simply retain solutions provided by the MT.

As for personalization, its large number in the human translation condition may reflect the tendency of native Hungarians to almost automatically turn English passive constructions into active Hungarian sentences, as there is no passive in modern Hungarian. In the post-editing condition, however, the MT system may have produced at least some impersonal solutions that are correct in Hungarian, and the post-editors may have decided to keep them.

It must be noted, however, that these attempts to explain our data should be seen as hypotheses for further research rather than conclusions. The unedited MT output should be analyzed qualitatively together with the post-edited texts before stronger claims are made about the source of the differences.

Table 2. ***The most frequent explicitation operations per work-mode (mean values).***

Human translation	Post-editing
distributing meaning (4.8)	word class change (3.6)
personalization (3.9)	lexical specification (2.6)
word class change (3.1)	distributing meaning (2.4)
lexical specification (3.0)	personalization (2.1)

6.2 Comparison between the explicitness of the MT output and the post-edited target texts

The second research question aimed to reveal how post-editors change the explicitness of the MT output. To answer this research question, first, explicitation operations in the target text created by DeepL were identified. The analysis showed that the DeepL output contained 9 explicitation operations.

The second step was to perform one-sample t-tests, using the number of explicitations in the MT output as the reference value. The aim was to determine whether the translators' and post-editors' target texts differed significantly from the MT output in

terms of the number of explicitation operations completed. The test results indicated that both the human translators and the post-editors carried out significantly more explicitation operations (one-sample t-test: FSHT $t = 6.21$, $p < .001$, PE $t = 3.3$, $p = .009$). This finding suggests that humans, whether translators or post-editors, may have a strong motive to make translated texts explicit and clear. This tendency may result in more explicitation operations than those observed in the MT output. It must be noted, however, that a higher number of explicitations will not necessarily lead to better quality target texts, as, unfortunately, errors may be introduced when attempting to make the text clearer. In this case, an error does not necessarily mean wrong information but may also involve any other type of error, like register, terminology, word order etc.

When comparing the NMT output and the post-edited target texts, the following types of post-editing behavior were identified. In some cases, the NMT explicitation operation was kept ($M = 5$, min = 1, max = 9). In some cases, the original explicitation operation was modified ($M = 1.8$, min = 0, max = 4). In other cases, the post-editors introduced new explicitation operations ($M = 7.3$, min = 0, max = 15). Overall, these numbers seem to indicate that students were not satisfied with the explicitness of the MT output and most of the time, they tried to make the target text more explicit. It is also interesting to see that students were reluctant to modify the segments already explicitated by DeepL. These findings are in line with previous empirical studies suggesting that NMT output has a priming effect (Castilho et al. 2017; Vardaro et al. 2019; Wang et al. 2024, 2026). In other words, previously seen information, in this case, the MT output affects how translators process subsequent information and what decisions they make. Another concept that may help describe what is happening here is anchoring bias. Anchoring bias refers to people's tendency to make decisions based on the initial information they are presented with (Furnham and Boo 2011). In our case, the MT output serves as an anchor. Post-editors may become cognitively anchored to the initial solution offered by the MT and may find it difficult to move away from it. Anchoring effects have also been reported in translation studies research (Cao and Zou 2026).

6.3 Correlation of explicitation operations with time on task

The third research question focused on the relationship between the frequency of explicitation operations and time on task both in the from-scratch human translation and in the post-editing condition. The question is important because a significant correlation may suggest that explicitation requires extra effort on the part of the translator or the post-editor.

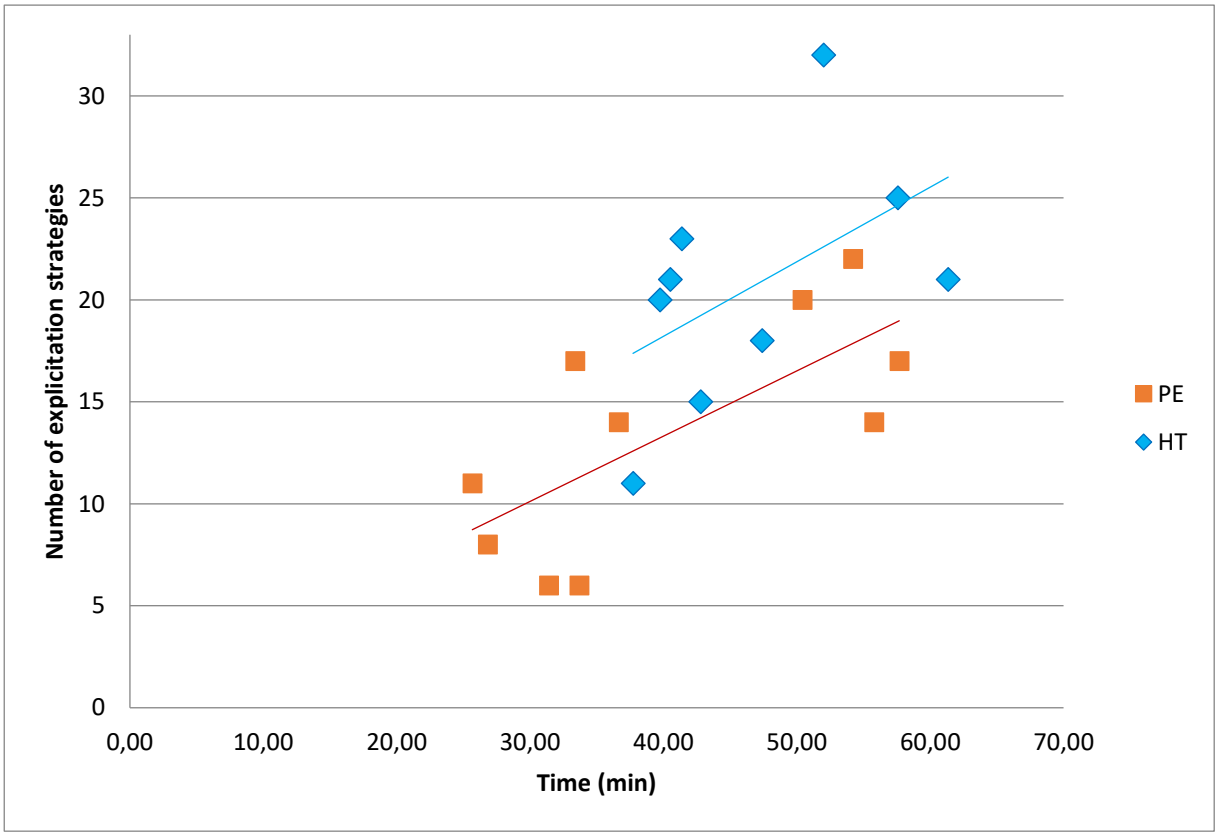
Our findings show that in the case of post-editing, there was a strong, significant positive correlation between the number of explicitation operations carried out and time on task ($r = .71$, $p = .021$), while in the case of human translation, no significant correlations could be observed between these two variables ($r = .52$, $p = .154$).

Figure 2 visualizes the relations between explicitation operations and time on task in each condition. As can be seen, individual cases are closer to the hypothetical trendline for post-editing than for from-scratch human translation, indicating a stronger relationship between the two variables.

The strong correlation observed in post-editing means that the more explicitation operations were carried out, the longer the time on task was. This seems to suggest that explicitation is not (always) an automatic operation but requires extra cognitive processing and extra time when post-editing an MT output. One possible reason for this extra time is that information must be retrieved from the context, which may take longer than processing information that is already explicit.

The lack of significant correlation in the from-scratch translation condition is somewhat surprising, but the complexity of the translation process itself may explain this phenomenon. It is very likely that there are other mental operations involved in translation that demand a considerable amount of time and, thus, intervene in the relations between time and explicitation. In light of this, the strong correlation found in post-editing is ever more striking. It may indicate that 'slower' post-editors spend a considerable amount of time on explicitation operations, because they find it important to make the MT output clear and comprehensible. This is, however, only a hypothesis that should be tested with qualitative interviews.

Figure 2. ***Correlation of explicitation operations with time on task.***



Further statistical tests were carried out to explore whether there were significant correlations between time on task and specific types of explicitation operations. Interestingly, only one operation, personalization, showed significant associations with time on task in both human translation ($r = .74$, $p = .023$) and post-editing ($r = .77$, $p = .009$). Personalization, as explained above, refers to rendering an English passive sentence as an active Hungarian sentence. The strong positive correlations in the data set indicate that extra time may be needed for identifying the agent in the SL text and for transforming the structure of the sentence.

6.4 Correlation between the number of explicitation operations and error numbers

In the analysis, we found no correlations between the overall quality of translations (total number of errors) and the number of explicitation operations carried out in either group. This is in contrast with the findings of Khoury (2024); however, it has to be noted that he worked with a different MT engine and focused on only acronyms and scientific terms.

Some significant correlations, however, were found in our data.

In human translation, there was a strong, significant, negative correlation between personalization operations and accuracy error numbers (r = -.772, p = .009). This suggests that making an effort to transform English passive sentences into active Hungarian ones is associated with lower error numbers. On the one hand, it is possible that the correct transformations themselves play a direct role in reducing error numbers. On the other hand, higher numbers of active-to-passive transformations involving personalization may be a manifestation of underlying factors, such as language awareness, which may directly contribute to reducing other error types.

In the case of post-editing, there were significant correlations between accuracy and the distribution of meaning operation (r = .835, p = .003), and between accuracy and adding quantifiers/modifiers (r=.743, p= .014). The observed positive correlations are potentially problematic, because they indicate that error numbers increased together with some explicitation operations. There might be several explanations for this; the most obvious one is that unnecessary changes and even errors were introduced into the target texts during post-editing.

6.5 Changing the MT output: associations with time on task and error numbers

As we have seen, the MT output contained significantly fewer explicitated language elements than either the human translated or the post-edited texts. This finding led us to examine an extra problem, originally not included in the research questions: how the willingness to explicitate the MT output is associated with time on task and error numbers. Our assumption was that modifying the MT output, that is, introducing new explicitation operations or changing the existing ones, (a) requires more time; and (b) leads to reduced error counts.

To answer our research question, first, each post-edited text was compared to the original DeepL output and explicitation operations in the post-edited texts were grouped into three categories: (1) completely new explicitation operations (not present in the MT-output); (2) modified explicitation operations (explicitation present in the MT-output modified by the post-editor); (3) no change (accepting the explicitation operation performed by DeepL, not changing it at all).

Next, Pearson correlation analyses were conducted to assess the strength of the correlations between the above factors. The results of these analyses showed that there was a strong, positive significant correlation between performing new explicitation operations (category 1) and time on task (r = .80, p =.006), and also between modifying the explicitation operations carried out by MT (category 2) and time on task (r =.62, p

=.056 – significance is trending). Not changing the MT solution (category 3) had no significant correlation with time on task.

These findings support the assumption that a decision to make the MT output more explicit is a critical one as it will extend post-editing time. An important question is whether such a temporal investment is worth the effort; that is, whether such changes increase the quality of the target text. As a result, we decided to carry out a series of statistical tests that examined correlations between the frequency of explicitation operations and error counts.

Concerning error numbers, strong, significant negative correlations were found between not changing the MT output and overall error count ($r = -.70$, $p = .024$), and between fluency errors ($r = -.66$. $p = .039$). In other words, the more MT explicitation operations students left unchanged, the lower the error count in their post-edited target language texts.

When it comes to modifying MT output and overall error count, the correlation is significant and positive ($r = .64$, $p = .48$), in other words, the more modifications to MT explicitation operations carried out by post-editors, the more errors in the post-edited target language texts. This supports the assumption that students may introduce errors when post-editing MT output.

There were no correlations between new explicitation operations and error numbers. This suggests that introducing new explicitation operations to MT output by post-editors might be idiosyncratic and may depend on individual preferences.

To summarize, correlations with error numbers indicate that students would be better off not modifying the MT output. Their well-intentioned efforts to explicitate the target text seem to have reduced the quality of the MT output. Moreover, this led to longer preparation times, suggesting potential waste of time and energy. This pattern likely reflects a situation where students feel the need to improve the texts, but because of a lack of post-editing competence, they end up doing more harm than good. As a word of caution, it must be noted that the error counts used in the analysis were overall error numbers; the correctness of the explicitation operations themselves was not evaluated. Accordingly, our conclusions are based on indirect data.

7 Summary

In this article, our aim was to answer research questions related to explicitation operations carried out during from-scratch human translation and the post-editing of neural machine translation, namely DeepL output.

Our first research question was about the frequency of explicitation operations in HT and PE. Our findings show that human translators carry out significantly more

explicitation operations than post-editors. This is in line with the findings of Robin (2019), who found that translators carry out more explicitation operations than revisers.

In addition, there seems to exist a work-mode-specific explicitation operation pattern. Human translators favored the operations of distributing the meaning of one source text unit over several units in the target text and the shift from referential to lexical cohesion, while post-editors carried out mostly word class changes and personalization with specification operations.

Our second research question focused on how post-editors changed the explicitness of the MT output. Compared with the machine-translated output, both the translated and post-edited target texts contained more explicitation operations. This is consistent with the findings of Krüger (2020).

Our third research question concerned the correlation between time on task and explicitation operations. The results indicate that, in post-editing, the more explicitation operations were carried out, the longer the time on task was. Both in the case of from-scratch human translation and post-editing, there was a significant positive correlation between time on task and the number of personalization explicitation operations carried out, which might be explained by the extra time needed to identify the agent in the passive sentences in the SL and to transform the sentence structure.

Our fourth research question investigated correlations between error numbers and explicitation operations. We found no correlations between the overall error numbers and the total numbers of explicitation operations carried out by post-editors or human translators. This is in contrast with the findings of Khoury (2024); however, he used the output of a different MT engine and focused on explicitation operations in the translation of acronyms and scientific terms. Still, there were some significant correlations between the frequency of certain explicitation operations and error numbers. This may indicate that some explicitation operations may play a key role in affecting the quality of the target text while others can likely be preferential. Nevertheless, this assumption should be tested in targeted research.

A focused analysis of how students changed the explicitness of the MT output and how these changes correlated with time on task and the quality of the post-edited texts yielded some surprising results. It was found that changing or adding explicitation operations to the MT output was accompanied by increased time on task and reduced target text quality, indicating largely wasted effort. As the participants were students, the findings underline the need for extensive training in post-editing.

Although, to our knowledge, this study is among the few to investigate explicitation operations in post-edited texts, it has some limitations. First, the sample size was small,

which restricts the generalizability of the findings. Another limitation is that we did not have the opportunity to assess whether the explicitation operations themselves improved or harmed the text. In the future, studies addressing these issues should be conducted. Given the importance of explicitation operations in ensuring the comprehensibility of target texts, it is vital that research on this topic continues.

Bakti, Mária; Lesznyák, Márta; Sermann, Eszter. 2026. Explicitation operations in from-scratch human translation and in the post-editing of neural machine translation. In: L10N Journal. 1(5), pp. 5-23.

Bibliography

Baumgarten, Nicole, Meyer, Bernd and Özçetin, Demet. 2008. Explicitness in Translation and Interpreting: A Critical Review and Some Empirical Evidence (of an Elusive Concept). In: Across Languages and Cultures. 9(2): pp. 177-203.

Blum-Kulka, Shoshana. 1986/2000. Shifts of Cohesion and Coherence in Translation. In: Venuti, Lawrence (ed.). 2000. The Translation Studies Reader. London and New York: Routledge. pp. 298-313.

Cao, Jing and Zhou, Tianli. 2026. Human or Machine? Do source beliefs shape cognitive bias in post-editing? Frontiers in Psychology. 17: Article 1819659. https://doi.org/10.3389/fpsyg.2026.1819659

Castilho, Sheila, Moorkens, Joss, Gaspari, Federico, Calixto, Iacer, Tinsley, John and Way, Andy. 2017. Is Neural Machine Translation the New State of the Art? In: The Prague Bulletin of Mathematical Linguistics. 108(1): pp. 109-120. https://doi.org/10.1515/pralin-2017-0013.

Englund Dimitrova, Birgitta. 2005. Expertise and Explicitation in the Translation Process. Amsterdam: John Benjamins Publishing Company.

Gumul, Ewa. 2006. Explicitation in Simultaneous Interpreting: A Strategy or By-product of Language Mediation? In: Across Languages and Cultures. 7(2): pp. 171-190.

Furnham, Adrian and Boo, Hua Chu. 2011. A literature review of the anchoring effect. In: The Journal of Socio-Economics. 40 (1): pp. 35-42. https://doi.org/10.1016/j.socec.2010.10.008

Khoury, Ogareet Y. 2024. Reflection of Explicitation in Scientific Translation: Neural Machine Translation vs. Human Post-Editing. In: Journal of Language Teaching and Research. 15(5): pp. 1510-1517. https://doi.org/10.17507/jltr.1505.12

Klaudy, Kinga. 1999. Az explicitációs hipotézisről. In: Fordítástudomány. 1(2): pp. 5-22.

Klaudy, Kinga. 2001. Az aszimmetria hipotézis. In: Bartha, M. (ed.). 2001. A X. Magyar Alkalmazott Nyelvészeti Konferencia előadásai. KJF. pp. 371-378.

Krüger, Ralph. 2014. From S-explicitation to T-explicitation? Tracing the Development of the Explicitation Concept. In: Across Languages and Cultures. 15(2): pp. 153-175.

Krüger, Ralph. 2020. Explicitation in Neural Machine Translation. In: Across Languages and Cultures. 21(2): pp. 195-216. DOI: 10.1556/084.2020.00012

Laviosa-Braithwaite, Sara. 1998. Universals of Translation. In: Baker, Mona (ed.). 1998. The Routledge Encyclopedia of Translation Studies. London and New York: Routledge. pp. 288-294.

Lesznyák, Márta. 2026. A Longitudinal Investigation of Translation Trainees' Translation and Post-editing Competence Development I: Translating an Academic Abstract. Budapest: Akadémiai Kiadó. https://mersz.hu/dokumentum/m1364alittt__1/

Lesznyák, Márta, Bakti, Mária and Sermann, Eszter. 2023. Translation and Post-editing Performance of Translation Students – A Cross-Sectional Analysis. In: L10N Journal. 2(2): pp. 7-23.

Lesznyák, Márta, Bakti, Mária and Sermann, Eszter. 2024. Reading Takes it All? – The Role of Language Competence and Subject Knowledge in Legal Translation. In: The Interpreter and Translator Trainer. 18(2): pp. 192-211. https://doi.org/10.1080/1750399X.2024.2344177

Lesznyák, Márta, Sermann, Eszter and Bakti, Mária. 2025. The Role of Beliefs about Translation in Translation Performance. In: Schmidhofer, Astrid and Recio Ariza, María Ángeles (eds.). 2025. Zukunftsperspektiven in der Translationswissenschaft / Future Perspectives in Translation and Interpreting Studies. Innsbruck: Innsbruck University Press. pp. 71-92. https://doi.org/10.15203/99106-165-6-06

Bakti, Mária; Lesznyák, Márta; Sermann, Eszter. 2026. Explicitation operations in from-scratch human translation and in the post-editing of neural machine translation. In: L10N Journal. 1(5), pp. 5-23.

PACTE. 2011. Results of the Validation of the PACTE Translation Competence Model: Translation Project and Dynamic Translation Index. In: O'Brien, Sharon (ed.). 2011. Cognitive Explorations of Translation. London: Continuum. pp. 30-53.

PACTE Group. 2020. Translation Competence Acquisition. Design and Results of the PACTE Group's Experimental Research. In: The Interpreter and Translator Trainer. 14(2): pp. 95-233.

Robin, Edina. 2019. The Modification of Translation Universals in Revised Texts. In: Sohár, Anikó, Limpár, Ildikó and Galambos, Dalma (eds.). 2019. Getting Translated. Budapest: Tinta Könyvkiadó. pp. 125-147.

Vardaro, Jennifer, Schaeffer, Moritz and Hansen-Schirra, Silvia. 2019. Comparing the Quality of Neural Machine Translation and Professional Post-Editing. In: 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE. pp. 1-3. https://doi.org/10.1109/QoMEX.2019.8743218.

Vinay, Jean-Paul and Darbelnet, Jean. 1958/1995. Comparative Stylistics of French and English: A Methodology for Translation. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wang, Feng, Xie, Hong and Zhang, Xiang. 2026. Association Between Machine Translation Post-Editing and Post-Editors' Cognition: A Three-Level Meta-Analysis Based on Eye-Tracking Evidence. In: Behavioral Sciences. 16(3): Article 365. https://doi.org/10.3390/bs16030365.

Bondarenko, Kateryna and Bondarenko, Oleksandr. 2026. How language industry defines localization: a corpus-based study. In: L10N Journal. 1(5), pp. 24-38.

How language industry defines localization: a corpus-based study

Kateryna Bondarenko and Oleksandr Bondarenko

Borys Grinchenko Kyiv Metropolitan University, Translatel Ltd

katerinabond@gmail.com, abondarenko@translatel.com

Abstract

The relationship between *translation* and *localization* remains conceptually contested in both academic literature and professional discourse. While theoretical models often attempt to draw clear boundaries between the two, industry usage suggests a more fluid and practice-oriented relationship. This study adopts a corpus-based approach to examine how *localization* is explicitly defined and how its relationship with *translation* is constructed in professional industry discourse. The analysis is based on a 2.4-million-token corpus compiled from *MultiLingual* magazine (2014–2024). Methodologically, the study combines the extraction of definitions, collocation analysis, and the examination of coordinated constructions. Explicit definitional contexts reveal that *localization* is conceptualized as a process involving *translation* and adaptation. Collocations demonstrate substantial overlap between *translation* and *localization*, particularly in domains related to service provision, technology, and quality. The analysis of coordinated constructions further shows that *translation* and *localization* are most frequently paired in service- and organization-oriented contexts, reflecting the realities of language service companies as primary discourse producers. Taken together, the findings indicate that contemporary industry discourse increasingly converges with the long-standing position of Translation Studies by framing *localization* as a *translation*-centered practice rather than as an independent ontological category. The study highlights the value of a corpus-based approach for clarifying terminological relationships in rapidly evolving professional domains.

Keywords: localization, translation, terminology, industry discourse, corpus linguistics, collocation analysis, language services industry

1 Introduction

The relationship between *localization* and *translation* has been extensively discussed in both academic literature and professional discourse (Kabát and Koscelníková 2022; Jiménez-Crespo 2024); however, no uniform understanding of *localization* has yet emerged. Numerous studies point to persistent conceptual ambiguity between the two terms (Pym 2004; Zethsen 2007; Achkasov 2016; Floros and Jiménez-Crespo 2016;

Charalampidou 2019; Kassawat 2021). Despite sustained scholarly attention, a clear and universally accepted definition of *localization* in relation to *translation* remains elusive. Several translation scholars (Hatim 2009; Chesterman 2009; Jiménez-Crespo 2024) conceptualize *translation* as an umbrella term extending beyond linguistic transfer to encompass publishing practices, ethical, political, and ideological dimensions, with *localization* framed as a specialized instantiation of *translation*. In industry discourse, *localization* is typically defined as the process of adapting content, products, or services to the linguistic, cultural, and technical requirements of specific target markets. Within the language industry, it has traditionally been argued that, unlike *translation*, which primarily focuses on linguistic transfer, *localization* involves a wider range of adjustments aimed at ensuring functional and cultural appropriateness. These adjustments may include the adaptation of textual and visual elements, as well as modifications to formats, conventions, and system-level features, such as dates, currencies, contact information, and payment methods (Cockrell 2024). This view positions *translation* as subordinate to *localization*, even though, paradoxically, *translation* remains indispensable to the existence of the other components. *Localization* is frequently conceptualized through the *translation*-plus model, attributed to B. Esselink, which frames *localization* as the *translation* and adaptation of a software product (Esselink 1998). Jiménez-Crespo identifies several features that distinguish *localization* from other types of *translation* and adaptation, including a strong reliance on technology, the use of translation management systems, text processing and programming, testing, and quality management (Jiménez-Crespo 2024). This distinction has also been linked to the organization of *localization* workflows. *Translation* often constitutes only one component of a broader *localization* process that encompasses engineering, testing, content management, and other technical tasks. In practice, the linguistic component is frequently outsourced to freelance translators, while language service providers coordinate the overall *localization* workflow. Consequently, the distinction between *translation* and *localization* may reflect differences in project organization and service provision as much as differences in the linguistic activity itself (DePalma 2021; Jiménez-Crespo 2024).

Taken together, prior research indicates that the conceptual boundaries between *localization* and *translation* remain fluid and contested. This persistent lack of definitional stability points to the need for empirical, usage-based approaches capable of capturing how *localization* is actually defined and understood in professional discourse – an aim to which corpus-based analysis is particularly well suited.

This study adopts a corpus-based approach to examine how *localization* is defined and conceptualized in professional industry discourse. Specifically, it addresses the following research questions:

RQ1. How is *localization* explicitly defined in industry discourse as represented in *MultiLingual* magazine (2014–2024)?

RQ2. What do corpus-derived collocations reveal about the relationship between *localization* and *translation* in professional practice?

2 Language industry publicistic discourse as empirical evidence

Industry discourse, in both oral and written forms, reflects established professional practice and contributes to its ongoing formation. The *MultiLingual* journal has, for several decades, functioned as a central communication platform for a broad range of stakeholders, including language service providers, localization teams, technology vendors and policy-oriented industry bodies. Its editorial scope consistently bridges operational practice, technological innovation and strategic reflection, making it particularly well suited for capturing mainstream professional discourse. The representativeness of *MultiLingual* is further reinforced by the diversity of its contributors, who include senior industry practitioners, localization managers, technologists, and, to a lesser extent, academic researchers and industry analysts. This plurality of voices ensures that the discourse reflects both hands-on operational concerns and higher-level conceptualization, while remaining anchored in professional practice. As a result, definitions and conceptual framings emerging in *MultiLingual* can be interpreted as indicative of widely circulated and socially validated understandings within the localization field, rather than isolated individual perspectives. The authorship profile remained broadly stable across the decade. Industry representatives consistently dominated the magazine, contributing 54% of articles in 2014 and 50% in 2024, while scholars accounted for 21% and 27%, industry researchers for 17% and 11%, and journalists for 8% and 12%, respectively. At the same time, the journal predominantly represents the perspective of language service providers and related industry stakeholders, whereas the voices of freelance translators and localizers, who perform much of the linguistic work itself, are comparatively less prominent. Consequently, the discourse captured in *MultiLingual* should be understood as reflecting an influential segment of the localization industry rather than the views of all professional participants.

As such, the journal provides a valuable empirical window into how *localization* is understood and discussed by practitioners, complementing previous empirical research based on interviews and focus groups with professionals (Dam and Zethsen 2024) by capturing naturally occurring professional media discourse in dynamics. Analyzing industry discourse allows researchers to capture definitions, conceptual boundaries, and evaluative stances as they are produced in situ, within the everyday

communicative practices of the localization industry. This is particularly relevant in a domain characterized by rapid technological change and evolving workflows, where professional usage often precedes and sometimes challenges conceptualization. Industry texts thus function not merely as secondary commentary, but as primary evidence of how key concepts are constructed and redefined over time.

3 Corpus design and analytical robustness

From the outset, the corpus design comprised ten distinct yearly corpora, each consisting of all *MultiLingual* journal issues published in a single calendar year. Texts were collected from publicly accessible archives and converted into a unified, machine-readable format to ensure consistency of analysis. Non-discursive material (such as advertisements and layout-related content) was removed without affecting the semantic integrity of the texts. An overview of authorship indicates that the corpus is dominated by industry practitioners, with contributions from academics and journalists providing complementary perspectives. This composition ensures that the corpus primarily reflects practitioner-oriented discourse while retaining a degree of conceptual diversity.

As shown in Table 1, the size of individual subcorpora varies considerably, reflecting fluctuations in publication volume, issue length, and editorial strategy over time. Notably, the period from 2019 to 2021 is characterized by smaller subcorpora, while earlier years (2014–2017) and 2022–2023 exhibit substantially higher token counts. This variation was taken into account at all stages of analysis since the study focuses on recurrent definitional frames and collocational tendencies that emerge across multiple subcorpora, rather than on isolated year-specific phenomena. Despite uneven yearly sizes, the overall corpus, comprising approximately 2.4 million tokens across a ten-year span, provides a sufficiently large and diverse empirical basis for identifying stable patterns as well as meaningful shifts in the conceptualization of localization. The combination of corpus-based and corpus-driven methods mitigates the effects of size imbalance and supports the robustness and interpretability of the findings.

Table 1. ***Composition of the MultiLingual subcorpora (2014–2024).***

Year	Tokens	Words	Sentences
MultiLingual 2014	291,634	244,725	10,422
MultiLingual 2015	404,678	333,688	13,718
MultiLingual 2016	313,998	266,777	11,289
MultiLingual 2017	278,474	236,765	10,264
MultiLingual 2018	237,023	200,457	9,129

MultiLingual 2019	80,916	70,643	3,385
MultiLingual 2020	132,755	115,710	5,692
MultiLingual 2021	152,121	132,038	6,137
MultiLingual 2022	232,681	201,208	9,451
MultiLingual 2023	216,684	187,647	8,634
MultiLingual 2024	165,856	143,954	6,888
Total	≈2.4M	≈2.07M	95,009

4 Methodology

To identify explicit definitional statements, the analysis focuses on definition frames, understood as recurrent syntactic patterns used to introduce definitions in discourse (Pearson 1998; Vintar and Grčić Simeunović 2017). In corpus linguistics, such frames are widely used for extracting terminology definitions from specialized texts.

The primary retrieval pattern employed in this study is the canonical X is a/the Y construction, operationalized through the following Corpus Query Language (CQL) query: [lemma="localization"] [lemma="be"] [word="a|an|the"]. This pattern reliably captures explicit definitional intent, as opposed to evaluative, metaphorical, or purely descriptive mentions of the term. The query was implemented in Sketch Engine, which provides lemmatization, concordance retrieval, and contextual analysis tools suitable for large-scale corpus research (Kilgarriff 2014). All retrieved concordance lines were manually screened to exclude non-definitional uses, truncated constructions, and cases where *localization* did not function as the definiendum. This combination of automated retrieval and qualitative validation follows established best practices in corpus-based terminology research.

To address the relationship between *localization* and *translation* in professional practice, a methodology for detecting functional synonymy was applied. The analysis of functional synonymy in this study is grounded in the assumption that synonymy in professional discourse is usage-based rather than absolute. Following corpus-linguistic and terminological approaches (Temmerman 2000), terms are treated as functionally synonymous when they perform overlapping communicative and conceptual functions within comparable discursive contexts, even if they are not interchangeable in all situations. In specialized and industry discourse, functional synonymy often emerges through contextual convergence, where different terms are used to describe similar processes, practices, or roles depending on perspective, tradition, or stakeholder orientation (Fernández-Silva and Freixa 2012). This is particularly relevant for the

relationship between *localization* and *translation*, which may alternate in use while referring to overlapping professional activities.

Functional synonymy was investigated through a multi-step corpus-based procedure, combining quantitative distributional analysis with qualitative contextual validation. To assess functional similarity, the analysis examined whether the candidate terms occur in comparable environments. This step involved collocation analysis: identifying shared collocates (e.g. *process*, *workflow*, *quality*, *service*, *strategy*) within a defined span.

A key diagnostic for functional synonymy was the comparison of definitional frames associated with each term. Definitional contexts were extracted using structured CQL queries. The analysis examined whether *localization* and *translation* are defined using similar hypernyms (e.g. *process*, *service*, *practice*), *described* through overlapping semantic components (e.g. adaptation, user focus, quality assurance), and embedded in parallel explanatory structures. Convergence at the definitional level was interpreted as strong evidence of functional synonymy.

5 Conceptualizations of localization in industry discourse: process-, and service-oriented perspectives

To identify explicit definitional contexts of *localization* in industry discourse, this study employs a highly productive CQL pattern: [lemma="localization"] [lemma="be"] [word="a|an|the"]. This query retrieves instances in which *localization* occurs in a canonical definitional structure (e.g. “localization is a/the …”). The analysis is based on 11 explicit definitional statements of *localization* identified in the *MultiLingual* corpus (2014–2024). The analysis reveals three definitional frames: process-oriented, service-oriented, and hybrid process–service conceptualizations.

The dominant definitional frame conceptualizes *localization* as a process, accounting for 7 out of 11 definitions (approximately 64%). In these cases, *localization* is explicitly described as a structured sequence of actions involving translation, adaptation, and market-specific adjustment. Typical examples include definitions such as *“Product localization is the process of translating product branding and customer-facing content, such as user manuals and product packaging, into target languages to meet local customer requirements and taste”* (2015) and *“Localization is the process of customizing the app’s content in order to serve different users with different preferences and habits in different markets”* (2019). Other definitions emphasize adaptation to linguistic and national contexts, for instance *“Localization is the process of adapting a product or solution for use in a particular language and country”* (2019), or explicitly frame *localization* as a specialized translation activity, as in *“Localization is a term defining a specific translation process”* (2021). Additional examples highlight the complexity of localization, describing it as a *“long and continuous process”* that requires testing at multiple stages, particularly in

domain-specific contexts such as game localization, or as a *"more complex and multistaged process that incorporates cultural aspects of the target market"*. This conceptualization reflects the shift towards agile development workflows, where *localization* is integrated continuously into the product development cycle.

Across all process-oriented definitions, three core components recur: *translation* of textual material, product adaptation (technical, cultural, and functional), and alignment with market expectations.

A smaller but analytically significant group of definitions – 2 out of 11 cases (approximately 18%) – conceptualizes *localization* as a value-creating service and business function. These definitions shift the focus away from procedural execution toward strategic contribution and organizational impact: *"Yes, localization is a service, but it's not a cost center. It's a partnership, and overall company success is directly tied to localization success"* (2022). Here, *localization* is framed as a strategic partnership whose value is measured in terms of business outcomes rather than operational efficiency. Similarly, *MultiLingual* (2020) links *localization* to user experience and personalization, noting that understanding users and customizing their experiences is central when *"localization is a goal"*. In these cases, *localization* is positioned as an integrated service that supports user-centered design and market performance.

6 Functional synonymy between translation and localization

To assess functional proximity between *translation* and *localization* in industry discourse, "noun–noun" collocation patterns were analyzed using immediate adjacency queries ([lemma="translation" & tag="N.*"] [tag="N.*"] and [lemma="localization" & tag="N.*"] [tag="N.*"]). The analysis is based on GDEX-sorted concordance data extracted from the *MultiLingual* corpus (2014–2024). Only unique noun lemmas occurring directly after the head nouns were counted; plural forms were lemmatized, and proper names (e.g., event titles) were excluded. The analysis identified the following numbers of unique noun collocates: "translation + noun": 52 unique noun lemmas; "*localization* + noun": 39 unique noun lemmas. This asymmetry reflects the broader discursive entrenchment of translation in industry texts. However, functional similarity is evaluated not through raw frequency, but through shared collocational structure.

Bondarenko, Kateryna and Bondarenko, Oleksandr. 2026. How language industry defines localization: a corpus-based study. In: L10N Journal. 1(5), pp. 24-38.

Figure 1. ***Collocates of*** **translation** ***and*** **localization** ***in*** **MultiLingual.**

16.12.25, 15:35 Concordance

SKETCH ENGINE

CQL [lemma="translation" & tag="N.*"] [tag="N.*"] • 7,828
3,227.37 per million tokens • 0.32%

Sort GDEX

	Details	sentence
1	doc#23	<s>Do you know your **translation volumes** and costs?</s>
2	doc#36	<s>The stakeholder who requests a **translation service** is a requester.</s>
3	doc#16	<s>In previous decades **translation technology** was concentrated on productivity gains in the area of user documentation.</s>
4	doc#31	<s>The main requirement is the **translation result** of high quality between the local language and interlingua.</s>
5	doc#36	<s>Some consider it to be at the core of **translation philosophy** and client satisfaction.</s>
6	doc#6	<s>Old functionality and translations had already been validated in earlier **translation projects** .</s>
7	doc#14	<s>A third major advantage of being highly specialized as a **translation company** is that many clients prefer working with specialists.</s>
8	doc#3	<s>Since then, many **translation tools** as well as other types of applications have implemented support for the format.</s>
9	doc#21	<s>I started my career working in different fields and capacities, before finally shifting back into the **translation realm** .</s>
10	doc#40	<s>Where do **translation service** providers add value to MT projects?</s>

16.12.25, 15:00 Concordance

SKETCH ENGINE

CQL [tag="JJ.*"][lemma="translation"] • 3,406
1,404.24 per million tokens • 0.14%

Sort GDEX

	Details	sentence
1	doc#13	<s>Expand your horizons and join our **5th translation** industry conference.</s>
2	doc#37	<s>Many people underestimate the extent to which **accurate translation** in financial services relies on expert industry knowledge.</s>
3	doc#3	<s>Many clients request **content translation** with quick turnaround without asking how it gets done.</s>
4	doc#16	<s>Another indicator of the market volume is the size of **top translation** companies headquartered there.</s>
5	doc#30	<s>WPML includes advanced features for translation management and an interface for professional **content translation** .</s>
6	doc#31	<s>For training documents – just like any other **pharmaceutical translation** – accuracy is essential.</s>
7	doc#32	<s>Who produced the **human translations** being compared?</s>
8	doc#36	<s> **Most translations** are done by average translators working with very tight deadlines and minimal context.</s>
9	doc#39	<s>A higher inverse edit distance signifies **better translation** quality.</s>
10	doc#42	<s>These related outcomes are in high demand in **medical translation** requests.</s>
11	doc#9	<s>Automation also turned this **quarterly translation** project into a more streamlined process.</s>

A comparison of the two noun sets reveals 21 shared noun lemmas that collocate with both *translation* and *localization*: *process, workflow, project, service, industry, system, management, team, quality, technology, platform, solution, tool, company, provider, cost, support, model, environment, strategy, operation*. These shared nouns constitute the conceptual core through which both activities are framed in professional discourse. The degree of overlap was calculated as follows: Overlap relative to *localization*: 21 / 39 = 53.8%; Overlap relative to *translation*: 21 / 52 = 40.4%. The results show that more than half of the noun collocational profile of *localization* overlaps with that of *translation*, while a substantial proportion of *translation* collocates are likewise shared.

The shared noun collocates cluster around several recurrent semantic domains: process and workflow (*process, workflow, project, operation*), framing both activities as structured, repeatable procedures; quality and evaluation (*quality, cost*), indicating comparable assessment criteria; technology and infrastructure (*system, platform, tool, technology, environment*), embedding both concepts in the same technological ecosystem; service and industry (*service, provider, company, industry, support*),

positioning both as commercial language services; management and strategy (*management, team, model, strategy*), reflecting organizational governance and coordination.

The noun–noun collocation analysis provides strong corpus-based evidence for functional synonymy between *translation* and *localization* in industry discourse. While *localization* shows a stronger tendency toward managerial and strategic framing, it remains anchored in the same operational, technological, and service-oriented conceptual space as *translation*. The high degree of overlap – particularly the 53.8% shared collocational profile of *localization* – suggests that, in contemporary industry discourse, *localization* is conceptualized as a *translation*-centered activity framed in organizational and market-oriented terms. This tendency may also reflect broader changes in industry discourse, whereby language service providers increasingly emphasize localization management, workflow coordination, quality assurance, and productivity, while the linguistic aspects of *localization* are probably more frequently discussed by freelance professionals.

To examine the degree of functional proximity between *translation* and *localization*, adjectival modification patterns were also compared. Adjectives occurring in attributive position were retrieved via Sketch Engine CQL queries and analyzed under the assumption that shared adjectival environments signal comparable functional roles, in line with the distributional hypothesis.

Figure 2. ***Adjectival evidence of functional proximity between*** **translation** ***and*** **localization.**

16.12.25, 14:59 Concordance

SKETCH ENGINE

CQL [tag="JJ.*"][lemma="localization"] • 1,464
603.59 per million tokens • 0.06%

Sort GDEX

	Details	sentence
1	doc#34	<s>The analysis results in a document containing **relevant localization** reviews as well as their translation and rating.</s>
2	doc#39	<s>Many **established localization** companies face talent shortages and struggle with high workloads and tight or unachievable deadlines.</s>
3	doc#42	<s>The company has reached those numbers after conducting **professional localization** .</s>
4	doc#16	<s>The simple fact is that most **continuous localization** apps are offered by companies at a price.</s>
5	doc#18	<s>Videolocalize.com is a cloud based online platform designed for **video localization** .</s>
6	doc#19	<s>External hires or outsourcing in those new marketplaces will become key for **crucial localization** purposes.</s>
7	doc#21	<s>Cultural differences and varieties in games lead translators and localizers on a **new localization** journey.</s>
8	doc#36	<s>We would also love to know a bit more about your **internal localization** mechanics.</s>
9	doc#40	<s>The core competency from **traditional localization** is less likely to be useful with these new challenges.</s>
10	doc#9	<s>Metrics can also reveal opportunities for a more efficient or **effective localization** process.</s>

16.12.25, 15:00 Concordance

SKETCH ENGINE

CQL [tag="JJ.*"][lemma="translation"] • 3,406
1,404.24 per million tokens • 0.14%

Sort GDEX

	Details	sentence
1	doc#13	<s>Expand your horizons and join our **5th translation** industry conference.</s>
2	doc#37	<s>Many people underestimate the extent to which **accurate translation** in financial services relies on expert industry knowledge.</s>
3	doc#3	<s>Many clients request **content translation** with quick turnaround without asking how it gets done.</s>
4	doc#16	<s>Another indicator of the market volume is the size of **top translation** companies headquartered there.</s>
5	doc#30	<s>WPML includes advanced features for translation management and an interface for professional **content translation** .</s>
6	doc#31	<s>For training documents – just like any other **pharmaceutical translation** – accuracy is essential.</s>
7	doc#32	<s>Who produced the **human translations** being compared?</s>
8	doc#36	<s> **Most translations** are done by average translators working with very tight deadlines and minimal context.</s>
9	doc#39	<s>A higher inverse edit distance signifies **better translation** quality.</s>
10	doc#42	<s>These related outcomes are in high demand in **medical translation** requests.</s>
11	doc#9	<s>Automation also turned this **quarterly translation** project into a more streamlined process.</s>

The analysis is based on two datasets of adjective–noun combinations extracted from the *MultiLingual* subcorpora. 68 unique adjective lemmas were identified as regularly modifying *translation*. These include, among others, *accurate, automated, audiovisual, computer-assisted, customized, domain-specific, effective, general, good, human, legal, literal, machine, manual, neural, professional, reliable, scientific, statistical, technical, universal.* 41 unique adjective lemmas were identified for *localization*, including *agile, automatic, component-based, continuous, corporate, deep, full, global, incremental, mature, medical, multilingual, professional, proper, pseudo-, scientific, strategic, systematic, technical, true.* Compared to *translation*, the adjectival profile of *localization* is narrower and shows a stronger orientation toward organizational, strategic, and process-management concerns, possibly also reflecting the editorial focus of the corpus and the perspective of language service providers.

Comparison of the two adjectival inventories reveals 18 shared adjective lemmas, which modify both *translation* and *localization*: *accurate, automated / automatic, audiovisual, continuous, effective, full, global, human, medical, professional, strategic, systematic, technical, true, multilingual, proper.* Nearly half of the adjectival profile of *localization* is shared with *translation*, while the shared adjectives constitute approximately one quarter of the broader adjectival space of *translation*.

The shared adjectival space clusters around quality-oriented adjectives: *professional, accurate, reliable, effective* which indicate that both concepts are evaluated against the same quality standards. Shared items such as *automated, continuous, systematic* frame both *translation* and *localization* as organized, repeatable workflows, reinforcing the view that *localization* is frequently conceptualized as a *translation*-centered process. Adjectives such as *human, technical, audiovisual, medical* show that *localization* inherits the same technological and domain distinctions traditionally associated with *translation*.

Despite substantial overlap, each term exhibits preferred adjectival domains: *translation* more frequently attracts linguistic-precision adjectives (e.g. *literal, faithful, idiomatic*), reflecting its traditional anchoring in linguistic theory and textual fidelity; *localization* more often co-occurs with organizational and technical adjectives (e.g. *corporate, mature, cross-functional, data-driven*), highlighting its framing as a broader operational and managerial construct. These differences point to functional specialization rather than conceptual opposition. The adjectival evidence provides strong corpus-based support for partial and asymmetric functional synonymy between *translation* and *localization*. While *translation* displays a broader and more heterogeneous adjectival profile, *localization* systematically draws on the same core evaluative, operational, and technical vocabulary. The high overlap relative to *localization* (43.9%) demonstrates that *localization* is largely framed through adjectival categories traditionally associated with *translation*, with differentiation emerging primarily at the level of scope, organizational framing, and strategic orientation, rather than core function.

In specialized and industry discourse, functional synonymy rarely arises from strict semantic identity. Instead, it emerges through contextual convergence, whereby different terms are mobilized to describe largely overlapping processes, practices, or professional roles, depending on disciplinary tradition, technological framing, or stakeholder perspective. This perspective is particularly productive when examining the relationship between *translation* and *localization*. Industry discourse frequently deploys the two terms interchangeably or in tightly coordinated constructions. Corpus evidence from professional publications demonstrates a strong tendency to conjoin the terms in coordinated structures such as *translation and localization*, especially in references to services, workflows, technologies, and professional identities. In such

contexts, the pairing does not signal conceptual opposition but rather reinforces a shared functional domain.

As a first diagnostic for functional synonymy, coordinated constructions were extracted. Coordination indicates that two terms are treated as conceptually compatible and operationally adjacent rather than mutually exclusive. CQL query: [lemma="translation"] [word="and|&"] [lemma="localisation|localization"] retrieved instances where the two terms are explicitly paired, typically in references to services, companies, workflows, tools, or industry events. The high frequency and stability of this pattern across the corpus confirm that *translation* and *localization* are routinely framed as part of a shared professional domain rather than as competing concepts.

The CQL query [lemma="translation"] [word="and|&"] [lemma="localization"] [tag="NN.*"] retrieving coordinated constructions of the form *translation and localization + noun* yielded 327 hits in the *MultiLingual* corpus (2,425,503 tokens), corresponding to 134.82 occurrences per million tokens and accounting for 0.01348% of the entire corpus. Although quantitatively marginal at the corpus level, the pattern is sufficiently recurrent to constitute a stable and analyzable discourse feature, enabling reliable observation of how *translation* and *localization* are jointly conceptualized in professional industry texts.

Figure 3. *Co-occurrence of* translation *and* localization *in* MultiLingual



18.12.25, 14:53 Concordance

SKETCH ENGINE

CQL [lemma="translation"] [word="and|&"] [lemma="localis… • 327

134.82 per million tokens • 0.013%

	Details	sentence
1	doc#0	<s>VSI Group www.vsi.tv TTC Language Services rebrands TTC Language Services Ltd., a **translation and localization company** , has changed its name and is now doing business as TTC wetranslate.</s>
2	doc#0	<s>River Linguistics, Inc. http://riverlinguistics.com TTC wetranslate www.ttcwetranslate.com suma redesigns website SpeakLatam moves office suma, a provider of **translation and localization services** in English, Latin American Spanish and Brazilian Portuguese, has redesigned its website as part of its first year anniversary celebration. suma www.sumalatam.com ExperTrans opens office in Myanmar ExperTrans, a provider of translation, interpreting, transcription, voiceover, subtitling and dubbing services, has opened an office in Yangon, Myanmar.</s>
3	doc#0	<s>Linguacle www.linguacle.com People Recent industry hires ■■ Netwire, a provider of language services, has hired José Messias de Oliveira and Bruno Nascimento to the technical depart- ment, Camila Souza to the quality assurance department and Daniela Guimarães de Paula to the marketing department. ■■ Localization Care, a **translation and localization agency** specializing in the languages of the CEE and FIGS regions, has hired Marta Fabin, a human resources management and marketing specialist. ■■ Diskusija UAB, a language service provider, has hired Ieva Barauskaite' as its sales manager and Indre' Panaviene' and Rasa Stalnionyte' as part of its project management team.</s>
4	doc#0	<s>The company is also celebrating its 20th anniversary. ■■ SpeakLatam, a **translation and localization service** provider with expertise in Latin American Spanish, has hired Luciana Frontoni to its project management team. ■■ Version internationale, a language service company focused on technical and medical translation, has hired Hélène Bajon as business development manager. ■■ Cloudwords, Inc., an online translation management platform, has hired Joseph Denzel as customer success manager. ■■ Across Systems GmbH, a manufacturer of corporate translation management systems, has hired Vanessa Hirthe as a supervisor for its business unit for language service providers and associated international network activities. ■■ Saudisoft Co. Ltd
5	doc#0	<s>Plunet GmbH www.plunet.com XTM International www.xtm-intl.com GFT Communicate GmbH www.gft-communicate.de Acrolinx GmbH www.acrolinx.com Alchemy Software Development Ltd. www.alchemysoftware.ie TextMaster partners with Transifex ACP Traductera, a.s., a **translation and localization company** , has added to its offering of computer-assisted translation tools.</s>
6	doc#0	<s>Globalization and Localization Association www.gala-global.org/conference The **Translation and Localization Conference** March 28-29, 2014, Warsaw, Poland.</s>
7	doc#0	<s>Nevertheless, the correlation between the automated MT quality metrics and Anna Rubina is responsible for ABBYY Language Services activities in Europe, promoting language technology solutions in the **translation and localization industry** .</s>
8	doc#0	<s>An open XML standard for the exchange of translation memory data created by computer-aided **translation and localization tools** . translation management system (TMS).</s>
9	doc#0	<s>A leading **translation and localization provider** , the company specializes in software localization and technical translation (IT, telecommunications, medical, automotive, engineering, marketing, financial, EU).</s>

https://app.sketchengine.eu/#concordance?corpname=user%2Fkaterinabond%2Fmultilingual_2014_2024&tab=advanced&queryselector=cql&attrs=word&viewmode=sen&attr_allpos=kw&linenumbers=1&refs_up=0&shorten_refs=1&g…

The analysis of nouns occurring immediately after the coordinated construction *translation* and *localization* yielded a total of 118 noun tokens. The most frequent nouns are *service(s)* (41 tokens; 34.7%), followed by organization-related nouns such as *company, firm, and agency* (29 tokens; 24.6%), and industry (18 tokens; 15.3%). The noun provider occurs 12 times (10.2%), further contributing to the organizational profile of the construction.

Process-related nouns, including *projects*, *process*, and *workflow*, account for 13 tokens (11.0%). Nouns referring to human or internal organizational units (*team*, *department*) are comparatively infrequent, with 5 occurrences (4.2%). All remaining nouns show very low individual frequencies (two tokens or fewer).

7 Conclusion

This study set out to examine how *localization* is defined and how its relationship with *translation* is conceptualized in professional industry discourse, using a corpus-based analysis of *MultiLingual* magazine (2014–2024). The findings demonstrate that, despite long-standing theoretical debates and prescriptive distinctions in academic literature, industry discourse exhibits a high degree of conceptual overlap and functional proximity between the two terms.

The analysis of explicit definitional frames shows that *localization* is most frequently conceptualized as a process, typically involving *translation*, adaptation, and alignment with market-specific requirements. At the same time, a smaller but consistent set of definitions frames *localization* as a service-oriented and strategic business function, indicating a shift from purely operational descriptions toward value-oriented perspectives. These definitional patterns confirm that *localization* is not conceptualized independently of *translation*, but rather consistently incorporates it as a core component.

Corpus-based evidence from noun–noun collocation patterns further supports this observation. More than half of the noun collocational profile of *localization* overlaps with that of *translation*, with shared collocates clustering around processes, services, industry, technology, management, and quality. This overlap indicates that both activities are embedded in the same operational, technological, and organizational environment. Differences in collocational preferences point to variation in scope and framing rather than to categorical separation.

Adjectival modification patterns reveal a similar tendency. While *translation* displays a broader and more heterogeneous adjectival profile, *localization* systematically draws on the same core evaluative and operational adjectives, particularly those related to quality, workflow organization, and technical domains. The observed asymmetry suggests functional specialization rather than conceptual opposition.

Coordinated constructions of the form *translation and localization + noun* provide an additional diagnostic for examining how professional actors conceptualize the relationship between the two activities. Although quantitatively marginal within the corpus as a whole, this pattern is recurrent enough to reveal systematic tendencies in noun selection. The nouns following the coordination predominantly refer to services, organizations, and market actors, while process-related and human-role nouns occur less frequently. This distribution suggests that, in professional discourse, the pairing *translation and localization* is most often used to characterize service offerings, organizational profiles, and industry structures rather than to signal functional differentiation.

This tendency can be related to the role of language service companies as the primary producers of the discourse represented in *MultiLingual*. Such companies are typically responsible for linguistic stages of *localization* – most notably translation, adaptation, quality assurance, and project coordination – while being less directly involved in downstream technical implementation or product engineering. As a result, *localization* is frequently framed in relation to *translation* as part of a shared service offering, rather than as a fully autonomous end-to-end process.

Taken together, the results support the conclusion that, in industry discourse, *translation* and *localization* function as partially and asymmetrically synonymous terms. *Localization* is typically framed as a translation-centered practice with an expanded organizational and strategic scope, rather than as an independent ontological category. The study demonstrates the value of corpus-based, usage-oriented approaches for clarifying contested terminological relationships and underscores the importance of grounding theoretical distinctions in empirical evidence from professional practice.

Bondarenko, Kateryna and Bondarenko, Oleksandr. 2026. How language industry defines localization: a corpus-based study. In: L10N Journal. 1(5), pp. 24-38.

Bibliography

Achkasov, Andrei V. 2016. If the Mountain Won't Come... Translation Studies Meets Localization. In: Journal of Siberian Federal University. Humanities & Social Sciences. 9(3): pp. 568-578. https://doi.org/10.17516/1997-1370-2016-9-3-568-578.

Chesterman, Andrew. 2009. The Name and Nature of Translator Studies. In: Hermes – Journal of Language and Communication in Business. 22(42): pp. 13-22.

Cockrell, Jordan. 2024. What Is Localization and Why Is It Important for Your Business? https://www.rws.com/blog/what-is-localization-why-it-is-important/. Accessed on: 22 July 2026.

Dam, Helle V. and Zethsen, Karen Korning. 2024. Professionals' Views on the Concepts of Translation: The Challenge of Categorisation. In: The Translator. 30(3): pp. 388-406.

DePalma, Donald. 2021. Augmenting Human Translator Performance. https://insights.csa-research.com/aim/en/reports/305013266. Accessed on: 22 July 2026.

Esselink, Bert. 1998. A Practical Guide to Software Localization. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Fernández-Silva, Sabela, Freixa, Judit and Cabré, M. Teresa. 2012. A Cognitive Approach to Synonymy in Translation. In: Brdar, Mario, Raffaelli, Ida and Žic Fuchs, Milena (eds.). 2012. Cognitive Linguistics between Universality and Variation. Newcastle upon Tyne: Cambridge Scholars Publishing. pp. 189-212.

Floros, Georgios and Charalampidou, Parthena. 2019. Website Localization: Asymmetries and Terminological Challenges. In: The Journal of Internationalization and Localization. 6(2): pp. 108-130.

Hatim, Basil. 2009. Translating text in context. In: The Routledge companion to translation studies. New York/London: Routledge. p. 36-53.

Jiménez-Crespo, Miguel A. 2016. What Is (Not) Web Localization in Translation Studies: A Prototype Approach. In: The Journal of Internationalization and Localization. 3(1): pp. 38-60.

Jiménez-Crespo, Miguel A. 2024. Localization in Translation. Abingdon and New York: Routledge.

Kabát, Marián and Koscelníková, Mária. 2022. Localization and Its Place in Translation Studies. In: L10N Journal. 1(1): pp. 4-26.

Kassawat, Madiha. 2021. Translation and Localization: Frequent Arguments through Asymmetric Binoculars. In: transLogos Translation Studies Journal. 4(2): pp. 27-44.

Kilgarriff, Adam et al. 2014. The Sketch Engine: Ten Years On. In: Lexicography. 1(1): pp. 7-36.

Pearson, Jennifer. 1998. Retrieval of Formal and Semi-Formal Defining Expositives. In: Terms in Context. Amsterdam/Philadelphia: John Benjamins Publishing Company. pp. 135-167.

Pym, Anthony. 2004. Localization from the Perspective of Translation Studies: Overlaps in the Digital Divide.https://www.researchgate.net/publication/250363090_Localization_from_the_Perspective_of_Translation_Studies_Overlaps_in_the_Digital_Divide. Accessed on: 22 July 2022.

Temmerman, Rita. 2000. Towards New Ways of Terminology Description: The Sociocognitive Approach. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Vintar, Špela and Grčić Simeunović, Larisa. 2017. Definition Frames as Language-Dependent Models of Knowledge Transfer. In: Fachsprache – Journal of Professional and Scientific Communication. 39(1-2): pp. 43-58.

Zethsen, Karen Korning. 2007. Beyond Translation Proper – Extending the Field of Translation Studies. In: TTR. 20(1): pp. 281-308.

Vraila, Stavroula Paraskevi. 2026. Political Translation in the AI Era: Journalists as Dual Agents and Systemic Vulnerabilities in Cross-Linguistic Political Communication. In: L10N Journal. 1(5), pp. 39-61.

Political Translation in the AI Era: Journalists as Dual Agents and Systemic Vulnerabilities in Cross-Linguistic Political Communication

Stavroula Paraskevi Vraila

Ionian University

svraila@ionio.gr

Abstract

This study examines how German political statements change when rendered for Greek audiences and what AI systems return for the same material. A corpus of sixteen documented cases from 2010 to 2025, thirteen in which the Greek departs from the German and three in which it holds, yields a taxonomy of five patterns: context isolation, systematic tone modification, narrative-driven translation, weaponization under polarization, and semantic manipulation. Three of the cases were submitted to GPT-4o and GPT-5.3 Instant under five prompt conditions, three runs per cell, and once each to DeepL and Google Translate, giving 96 coded outputs. Under the two conditions that ask only for a translation, seventeen of eighteen language model outputs in each carry none of the five patterns in the author's coding, as do five of the six machine translation outputs. An instruction to translate for a Greek news website returns adaptations in fifteen of eighteen outputs, against one of eighteen under a bare instruction, and no headlines.

Keywords: political translation, journalistic translation, neural machine translation, large language models, prompt design, cross-cultural communication

1 Introduction

The translation of political discourse is an ideological practice, and translational decisions are shaped by institutional constraints and by the norms of the context in which the text is received (Schäffner 2012). In journalism, the translator is seldom a separate figure. Journalists who translate act as what Bielsa and Bassnett (2009) call double agents, and the texts they produce reframe political statements for a new readership.

AI systems now take part in that work. In a survey of Greek journalists, 51% regarded AI as useful for translation and 32% reported using it (Triantafyllou et al. 2025).

The article compares the patterns documented in human journalistic translation with what AI systems return when the task put to them changes, and asks three questions. Which recurrent patterns characterize problematic translation of political statements in journalism? Do AI systems reproduce these patterns, and under which instructions? Which institutional, temporal and technological conditions make political mistranslation systemic?

2 Political Translation, News Translation and the Platform Environment

Venuti (1995) argued that the ideal of fluent, transparent translation conceals the translator's interpretive work, a premise recently revisited by Freeth and Treviño (2024) to account for contemporary sociotechnical and institutional settings, while Baker (2006) describes translation as renarration, through which a text foregrounds particular actors and evaluative frames. Schäffner (2004, 2012) ties such choices in political texts to the constraints that govern what is translated and how political actors are framed. Lexical choices that look neutral carry evaluative weight that accumulates across a text (Munday 2012).

News translation research, a recognizable subfield for some twenty-five years (Valdeón 2025), sits between journalism studies and translation studies, which have not conceptualized translation in the same terms (Valdeón 2023); Kalantari (2022) locates journalistic translation where the two meet. This article draws on both, taking its linguistic categories from one and its account of editorial practice from the other.

Machine translation had been shown to reproduce gender bias before language models entered the newsroom (Prates et al. 2020). The closest published work is Conway et al. (2025), who argue that AI-based machine translation has changed the conditions under which journalists could be assumed to compose news texts and not translate them, and who measure lexical distance between parallel corpora of AI-generated news translation. The present study shares the premise and differs in method: a distance measure registers how far wording diverges, whereas a controlled design can tie a given shift to the instruction that produced it.

Automatic translation also operates after publication. On social media, the translation offered to a reader depends on a language-identification step that can misfire, and at least one platform gives the reader no way to contest the translation it displays (Paullada and Haroutunian 2026). The journalist is then one translating agent among several between the source and the reader's screen.

Newsrooms that test headlines select among variants by click-through rate (Hagar and Diakopoulos 2019), and a headline can depart from the source while the body beneath

it stays closer. This article uses quality in the narrower sense of fidelity to the source statement and treats engagement as a separate criterion that need not coincide with it.

3 Methodology

3.1 Research design, corpus and analytical framework

The study combines a corpus of documented cases in which German political statements were rendered for Greek audiences, analyzed qualitatively, with a set of AI translation outputs produced under controlled conditions and coded. Framing analysis (Entman 1993) supplies the account of how a rendering positions actors and assigns responsibility, and the comparison of source and rendering follows House's (2015) procedure for setting a translation against its original along lexical, syntactic and pragmatic dimensions.

The corpus holds sixteen cases from Greek coverage of German political speech between 2010 and 2025: thirteen in which the Greek departs from the German, dated between 2011 and 2025, and three in which it holds. The window opens in 2010, with the onset of the debt crisis. Cases were selected on four criteria: a documented diplomatic response or public controversy; an available German source; Greek press coverage carrying the disputed rendering; and a difference between source and rendering that can be described linguistically. Table A1 in the appendix gives each case with its German segment, its Greek rendering, the outlet that carried it and the patterns coded for it. The German wording of fifteen cases was confirmed verbatim against the recording, the official statement or a German press report quoting it. For one of the three faithful cases only a secondary account of the German was available.

The number of outlets indicates how widely a rendering circulated: five of the thirteen are documented in three or more, two in two, and six in one. Where only one outlet is documented, the case shows that a rendering was published and circulated; how far it reached in Greek coverage is left open. One case is a wire translation reproduced by four outlets, which carry one rendering between them.

Three cases carry the experimental work: an unscripted remark at a press conference in 2015, a broadcast interview in 2022, and an institutional policy statement in 2025. They were chosen before the experiments for that spread of discourse types and periods, and because one of the three carries a formal diplomatic response. They are not offered as typical of the corpus, and the experimental results are not generalized to it.

3.2 The translation experiments

3.2.1 Design and prompt conditions

Each of the three German source texts was submitted to two large language models under five prompt conditions, three times per condition, and to two neural machine translation systems. This yields 96 outputs: 90 from the language models (three cases, five conditions, two models, three runs) and six from the NMT systems (three cases, two systems; Vraila 2026).

Prompt wording (Table 1) was fixed before execution and held identical across cases, systems and runs.

Table 1. ***Prompt conditions.***

ID	Label	Prompt as issued
C1	Bare translation	Translate the following text into Greek.
C2	Professional political translator	Translate as a professional translator of political texts.
C3	Newsroom	Translate the following statement for a Greek news website.
C4	Headline task	Write a headline in Greek for a Greek news website based on the following statement.
C5	Journalistic adaptation	Adapt the following statement in Greek for a Greek news audience.

Prompts were issued in English, which is recorded as a parameter of the design, since the language of the instruction may itself condition what the model returns. The source text followed the prompt after a single newline, unquoted, with speaker, date and outlet withheld.

C1 alone asks the language models to do what the NMT systems do, and every comparison between the two architectures is made against it; C2 to C5 vary the task.

3.2.2 Systems, access and execution parameters

Access through each provider's web interface left temperature, top-p, maximum token count, seed and the system prompt at the provider default in every run; no API was used. Web-interface access was chosen for ecological validity, mirroring authentic newsroom practice where journalists interact via default consumer UIs rather than API endpoints. The identifiers in Table 2 are the labels shown in the model selector at execution; such

a label names a product. It does not identify the version or the build that served a given request.

Table 2. ***Systems, access and execution parameters.***

System	Access route	Identifier as recorded	Session dates
GPT-4o	Web interface	GPT-4o (interface label)	13–15 March 2025
GPT-5.3 Instant	Web interface	GPT-5.3 Instant (interface label)	7–8 March 2026
DeepL, free web translator	Web interface	DeepL free web translator	13 March 2025
Google Translate	Web interface	Google Translate web	13 March 2025
GPT-4o	Web interface	GPT-4o (interface label)	13–15 March 2025

The two rounds of language model runs, roughly twelve months apart, used different models, and the change of model is confounded with everything else that altered between March 2025 and March 2026. The second round is therefore reported as a descriptive time point.

The runs were executed by round. For GPT-4o, the first runs were made on 13 March 2025, the second on 14 March and the third on 15 March, with one third run made on 13 March. For GPT-5.3 Instant, the first and second runs were made on 7 March 2026 and the third on 8 March. Each run was issued in a fresh session.

3.2.3 Stability of the machine translation outputs

Each NMT system was run once per case in the main matrix. Two uncoded passes on 8 September 2026 tested what a single run represents: one repeated the submissions within a single sitting, the other set the outputs against those of March 2025.

3.2.4 Coding procedure

The coding scheme is the taxonomy of five patterns, derived from the documented corpus before the experiments were run. Each output received one task type from four values, a binary judgment on each of the five patterns, which may co-occur, and a binary judgment on five modal and register features. A translation renders the propositional content in full. A condensed translation reduces the content without adding to it, and the source syntax remains recoverable. A headline is a nominal or elliptical form from which the source syntax cannot be recovered, and an output meeting that test is classed

as a headline whatever else has shifted. An adaptation asserts something the source does not assert, or drops something it does, so that truth conditions differ. A shift of mood, person or evaluative lexis on its own does not make an output an adaptation; such shifts are recorded in the pattern and modal variables. The first coder also recorded the text span that triggered each code; the second recorded codes alone. The codebook carries definitions, exclusions and five dated amendments; Amendment 5 records a criterion already applied and changes no code (Vraila 2026).

Both coders coded all 96 outputs independently, under different conditions. The second coder is a professional translator of journalistic and political texts working in the German to Greek pair, and had the German source texts alongside the Greek outputs. That access is required by the instrument: three of the five patterns turn on what the output does to material present in the source, as does the weaponization test, which asks whether the rendering adds hostility the German does not carry. The second coder worked blind to the condition, the system and the manuscript with its hypotheses, and outputs were presented under anonymized item numbers in an order shuffled with a fixed seed [random.Random(4242)]. The first coder is the author and coded in the master log, where the condition and the system of every output are visible. The agreement figures therefore compare a blind reading with an informed one.

Calibration followed the first independent round instead of preceding it. Amendments 1 and 2 were made during the first round, as questions arose, and applied to it. Amendments 3 and 4 came out of a calibration session, held after the first round, on the two variables whose agreement fell furthest below that of the rest of the set: Amendment 3 restricted weaponization to hostility added by the rendering, and Amendment 4 set the boundary between translation and adaptation. Task type was then recoded independently by both coders against the amended definition, weaponization was settled by consensus, and every other column was left as first coded, apart from six later changes by the first coder set out in the note to Table 3.

3.2.5 Intercoder agreement

Cohen's κ, computed in spreadsheet formulas, is reported in Table 3 per variable and never pooled, since pooling would let the mechanical variables mask weakness in the interpretive ones.

Table 3. ***Intercoder agreement by variable (Cohen's κ, n = 96).***

Variable	κ	Round	Disagreements
Task type	0.856	2	9
Context isolation	0.907	2	3

Tone modification	0.915	2	4
Narrative-driven translation	0.793	2	10
Weaponization	0.683	1	13
Semantic manipulation	0.915	2	4
Subjunctive preserved	1.000	2	0
Impersonal construction preserved	1.000	2	0
Ironic implicature preserved	0.938	2	3
Legal term rendered with established equivalent	1.000	2	0
Evaluative lexis added	0.957	2	2

Note. Only task type was recoded after calibration; every other variable is compared on its first-round codes. The values for tone modification, narrative-driven translation and semantic manipulation differ from those first computed because three spreadsheet cells overwritten with static values were restored to their formulas. The narrative-driven value also reflects one code the first coder later changed to satisfy the rule that weaponization entails narrative-driven translation, which lowered it from 0.813 to 0.793. The value for evaluative lexis reflects five codes the first coder changed from 1 to 0 after the agreement figures were first computed, on re-examination of outputs in which no added evaluative item could be identified; before the change it stood at 0.957, with two disagreements.

Weaponization is reported at its uncalibrated first-round value ($\kappa = 0.683$), below the threshold of 0.70 set for the study. Disagreements stemmed from distinguishing marked lexical sharpening from outright hostile intent; Amendment 3 resolved this by strictly requiring newly introduced hostility, after which the coders settled remaining discrepancies item by item. Task type, the other variable taken to calibration, stood at 0.762 before recoding.

4 The three cases

The linguistic analysis of the three cases develops an account first given in Vraila and Livas (2025), who examined them as instances of shifting emotional charge in Greek media renderings of German political speech. That analysis is what this article draws from the earlier paper, which has no experimental component. For the 2015 and 2022 cases the German passages used here are fuller than those quoted there, and for the 2022 and 2025 cases the conclusions reached here differ from the earlier ones. The

experimental design, the 96 coded outputs, the codebook and its amendments, and the intercoder agreement procedure are new here.

4.1 Case 1: Schäuble on Varoufakis (2015)

At a press conference on 10 March 2015, answering a question about his meeting with Varoufakis, Schäuble said: "Also, dass er nun plötzlich naiv in Sachen Kommunikation wäre, habe ich ihm gesagt, das ist mir ganz neu. Aber man lernt ja nie aus." (Well, that he should now suddenly be naive in matters of communication, I told him, is quite new to me. But one never stops learning.) The passage is continuous and was submitted to the systems as it stands.

Mega TV broadcast the remark the same day as a subtitled excerpt, cut from the surrounding exchange: «το να είναι τόσο ανόητα αφελής, δεν το περίμενα» (for him to be this foolishly naive was something I had not expected). The Greek ambassador in Berlin delivered a démarche that evening, reported on 11 and 12 March, and German officials answered by referring to what the minister had actually said and denying any intent to offend.

The subtitle removes the reporting frame «habe ich ihm gesagt» (I told him), which placed the judgment in an exchange that had already taken place. It states in the indicative what the German, through the subjunctive «wäre» (would be), puts forward without asserting, and it drops «in Sachen Kommunikation» (in matters of communication), which confined the judgment to one domain. Its degree modifier, «τόσο ανόητα» (so foolishly), has no counterpart in the German, where the predicate is qualified by domain alone. The patterns coded for the subtitle are context isolation, tone modification and semantic manipulation.

Table 4 records the coded outcomes for the six language model outputs in each condition. Under C1, five of the six language model outputs are full translations carrying none of the five patterns. The sixth is an adaptation, «Του είπα ότι είναι εντελώς αφελής στην επικοινωνία, δεν το περίμενα» (I told him he is utterly naive at communication; I had not anticipated it), which delivers the hypothetical as fact and turns a remark about what was new to the speaker into one about what he had anticipated. Under C2, the one condensed rendering drops the closing clause and gives «ganz neu» (quite new) as «με ξενίζει εντελώς» (strikes me as wholly strange). With both changes, the remark stops reporting a claim and starts recording an attitude, and that change to the utterance is why it is coded for context isolation, together with semantic manipulation. The hypothetical, which this rendering carries in a nominalized clause, and the irony that remains are coded as preserved in their own columns.

Under C3, tone modification is coded in five outputs, each time through a degree modifier: «τόσο αφελής», «τόση αφέλεια», «τόσο ανόητα αφελής» (this naive, this much naivety, this foolishly naive).

Of the six headlines, two assign the quality to Varoufakis without comment, two present the remark as an insult, as in «Αφελής επικοινωνιακά ο Βαρουφάκης» - Νέα προσβολή Σόιμπλε ("Varoufakis naive at communication": Schäuble insults him again), and one puts in Schäuble's mouth direct speech the source does not contain. The sixth sets an evaluation of Schäuble against his of Varoufakis: «Η «αφέλεια» του Βαρουφάκη και ο κυνισμός του Σόιμπλε» (Varoufakis's "naivety" and Schäuble's cynicism). Ironic implicature survives in none of the six headlines and in twenty-three of the twenty-four outputs produced under the other four conditions.

Table 4. ***Coded outcomes for Case 1 by condition (n = 6 language model outputs per condition).***

Condition	Task type	Context isolation	Tone modification	Narrative-driven	Weaponization	Semantic manipulation
C1 Bare translation	5 translations, 1 adaptation	0	1	1	1	1
C2 Professional political translator	5 translations, 1 condensed	1	0	0	0	1
C3 Newsroom	3 translations, 3 adaptations	0	5	3	2	3
C4 Headline task	6 headlines	6	6	6	6	6
C5 Journalistic adaptation	6 adaptations	0	4	4	2	6

Note. Under C1, the four coded patterns occurred co-extensively within a single adaptation output.

Of the machine translation outputs, Google Translate returned a translation carrying none of the five patterns and DeepL an adaptation carrying semantic manipulation; ironic implicature was coded as preserved in both.

The wording of the broadcast subtitles recurs in the experimental set. Three outputs use «ανόητα» (foolishly) or its adjectival form, two under C3 and one under C5, and three render the second clause with «περίμενα» (I expected), as the broadcast does, one each under C1, C3 and C5; the output under C5 does both. Where the systems drew this wording from is beyond what the design can show.

4.2 Case 2: Schäuble on sweaters (2022)

The input for this case joins two passages of a Bild TV interview broadcast on 10 October 2022, and the join is unmarked: "Dann zieht man halt einen Pullover an. Oder vielleicht

noch einen zweiten Pullover. Darüber muss man nicht jammern, sondern man muss erkennen: Vieles ist nicht selbstverständlich. Deshalb sollte man tatsächlich immer auch ein paar Kerzen, Streichhölzer und auch eine Taschenlampe zuhause haben." (Then one just puts on a sweater. Or perhaps a second sweater as well. There is no need to complain about it; one has to recognize that much is not to be taken for granted. That is why one really should always keep a few candles, matches and a torch at home.) The sentence to which *Deshalb* (that is why) refers, "Er (der Strom) kann auch mal ausfallen" (It, the electricity, can fail from time to time), lies outside the string, and none of the thirty-two outputs for the case supplies it.

The passage turns on a contrast. *Darüber muss man nicht jammern, sondern man muss erkennen* (there is no need to complain about it; rather, one has to recognize) denies any need to complain and puts the weight on recognition. Its subject, the generic *man* (one), assigns the advice to no one in particular while keeping the speaker inside its scope.

Table 5 records the coded outcomes by condition. Under C1 and C2, all twelve language model outputs are full translations. Each keeps the recognition clause and the generalizing construction, most often as a first-person plural, and each renders the German's denial of need as such, with «δεν χρειάζεται» (there is no need) or «δεν υπάρχει λόγος» (there is no reason): «Δεν χρειάζεται να γκρινιάζουμε γι' αυτό, αλλά να συνειδητοποιούμε ότι πολλά πράγματα δεν είναι αυτονόητα» (We have no need to complain about it; we need to realize that many things cannot be taken for granted).

Under C3, the generalizing construction survives in none of the six outputs, and the Greek addresses the reader: «Δεν χρειάζεται να γκρινιάζετε για αυτό, αλλά να καταλάβετε ότι πολλά πράγματα έπαψαν να είναι αυτονόητα» (You have no need to complain about it; you need to understand that many things are no longer a given). One output delivers the clause as a prohibitive imperative, «Μη γκρινιάζετε» (Do not complain).

Each of the six headlines omits material whose removal changes the force of what remains and adds evaluative wording with no counterpart in the source. GPT-4o returned, for instance, «Σόιμπλε: Να φορέσετε και δεύτερο πουλόβερ και να μη γκρινιάζετε» (Schäuble: Put on a second sweater too, and stop complaining). The three GPT-5.3 Instant headlines place the evaluation in the frame around the quoted words, as in «Αν κρυώνετε, να φοράτε πουλόβερ»: Η κυνική συνταγή Σόιμπλε για τον χειμώνα ("If you are cold, wear a sweater": Schäuble's cynical recipe for winter).

Table 5. ***Coded outcomes for Case 2 by condition (n = 6 language model outputs per condition).***

Condition	Task type	Context isolation	Tone modification	Narrative-driven	Weaponization	Semantic manipulation
C1 Bare translation	6 translations	0	0	0	0	0
C2 Professional political translator	6 translations	0	0	0	0	0
C3 Newsroom	6 adaptations	0	6	6	0	6
C4 Headline task	6 headlines	6	6	6	5	6
C5 Journalistic adaptation	6 adaptations	0	6	6	3	6

DeepL and Google Translate each returned a full translation carrying none of the five patterns.

ERT, Greek public television, published its report on 11 October 2022, and its body stays closer to the German than its headline does. The body keeps the recognition clause, the first-person plural for most of the passage and, because it draws on the whole interview, the causal link made by *Deshalb*. It departs from the German at two points. Its first sentence gives *man* as a third person plural, «Τότε απλά ας φορέσουν ένα πουλόβερ» (Then let them just put on a sweater), and «Δεν πρέπει να γκρινιάζουμε για αυτό» (We must not complain about it) renders *muss man nicht jammern,* which denies a need, as a prohibition.

The headline reads «Συμβουλές Σόιμπλε για τον χειμώνα: Μην γκρινιάζετε, φορέστε πουλόβερ, πάρτε κεριά και σπίρτα» (Schäuble's tips for the winter: Stop complaining, put on sweaters, pick up candles and matches). It turns the body's first-person plural into a second-person imperative and drops the recognition that the German sets against complaint, *sondern man muss erkennen*. An imperative of this kind selects an addressee, and a remark about a condition the speaker shares becomes an instruction to those who bear it. The case also carries semantic manipulation from a separate exchange at the end of the interview. Asked whether he could still laugh, Schäuble answered «Ja klar, lachen befreit und entspannt» (Of course, laughing frees you and relaxes you), which the report gives as «Το γέλιο είναι δωρεάν και χαλαρωτικό» (Laughter is free and relaxing), reading *befreit* as free of charge.

For this case, address in the second-person plural appears in the language model outputs under C3, C4 and C5, and in none produced under C1 or C2. What separates the six headlines from the C3 outputs is the loss of the recognition clause, which survives in every C3 output and in no headline; the same loss separates the ERT headline from

its body. The parallel is one of operation, and nothing in the material links the ERT report to such systems.

4.3 Case 3: Merz on entry bans (2025)

The Merz statement turns on a distinction of legal status. The two consecutive sentences submitted to the systems read: "Es wird ein faktisches Einreiseverbot in die Bundesrepublik Deutschland für alle geben, die nicht über gültige Einreisedokumente verfügen oder die von der europäischen Freizügigkeit Gebrauch machen. Das gilt ausdrücklich auch für Personen mit Schutzanspruch." (A de facto ban on entry into the Federal Republic of Germany will apply to everyone who lacks valid entry documents or who exercises European freedom of movement. This applies expressly to persons with a claim to protection as well.) *Oder* (or) holds apart two classes of person: those without valid entry documents, and those exercising European freedom of movement, an entitlement under EU law for which *Freizügigkeit* is the term of art. Whether an output keeps that second class, and keeps it as an entitlement, is the discriminating test in this case, coded on every output as the legal-term variable.

Table 6 records the coded outcomes by condition, together with the legal-term variable. Under C1 and C2, all twelve language model outputs are full translations, and each keeps both the disjunction and the term, most often with the entitlement made explicit: «ή κάνουν χρήση του δικαιώματος ελεύθερης κυκλοφορίας εντός της Ευρωπαϊκής Ένωσης» (or exercise the right of free movement within the European Union).

Under C3, each of the six outputs adds evaluative wording with no counterpart in the source, and five keep the term, though not always the syntax around it. GPT-4o returned «Θα ισχύσει στην πράξη ντε φάκτο απαγόρευση εισόδου στη Γερμανία για όλους τους παράτυπους μετανάστες που δεν διαθέτουν έγκυρα έγγραφα εισόδου ή κάνουν χρήση της ευρωπαϊκής ελεύθερης κυκλοφορίας» (In practice a de facto ban on entering Germany will apply to all irregular migrants who lack valid entry documents or use European free movement). A single noun phrase now governs both arms of the disjunction, and the people in the second arm are irregular migrants before the sentence reaches their right. Two outputs replace the disjunction with a conjunction, and two render the exercise of the right as its «κατάχρηση» (abuse); one does both and also substitutes a general noun, «ελεύθερη μετακίνηση» (free movement), for the term of art. An output is credited with the term where «ελεύθερη κυκλοφορία» (free circulation) carries a marker of the entitlement, and a merger of the two classes around it is recorded as semantic manipulation.

One GPT-4o headline reads «Μερτς: "Πρακτική απαγόρευση εισόδου" ακόμη και για όσους δικαιούνται προστασία» (Merz: "Practical entry ban" even for those entitled to protection). One GPT-5.3 Instant headline keeps that clause and adds «"Μπλόκο"» ("Blockade") and «Νέοι αυστηροί περιορισμοί στα σύνορα» (New strict restrictions at the border).

Table 6. ***Coded outcomes for Case 3 by condition (n = 6 language model outputs per condition).***

Condition	Task type	Context isolation	Tone modification	Narrative-driven	Weaponization	Semantic manipulation	Legal term rendered
C1 Bare translation	6 translations	0	0	0	0	0	6
C2 Professional political translator	6 translations	0	0	0	0	0	6
C3 Newsroom	6 adaptations	0	6	6	4	6	5
C4 Headline task	6 headlines	6	6	6	6	6	0
C5 Journalistic adaptation	6 adaptations	0	6	6	4	6	2

Merz made the statement in Berlin on 23 January 2025, in a five-point press statement given the day after the attack in Aschaffenburg. It reached Greek readers as one wire translation by ΑΠΕ-ΜΠΕ, the Athens-Macedonian News Agency, and the four outlets that carried it reproduce the quoted passages word for word: ΣΚΑΪ, Protagon and Athina 9.84 with an agency credit, Liberal without one.

In the wire, the two classes of the German become one. The body gives the measure as «"De facto απαγόρευση εισόδου" σε παράνομους μετανάστες, ακόμη και εάν βρίσκονται σε καθεστώς προστασίας» ("De facto entry ban" on illegal migrants, even where they hold protection status). The people the German places in the second arm of the disjunction, who exercise a right, are absorbed into the category of illegality. ΣΚΑΪ's headline names the same group «παράτυπους μετανάστες» (irregular migrants). Later in the report the wire fills the second arm with different content: «Όποιος δεν μπορεί να παρουσιάσει έγγραφα ή έχει ήδη εισέλθει στην Ευρωπαϊκή Ένωση από άλλη χώρα» (Anyone who cannot produce documents or has already entered the European Union from another country). Prior entry through another country takes the place of the entitlement the German names.

Twenty-one of the thirty-two outputs for this case render *Freizügigkeit* with an established Greek equivalent: all twelve language model outputs under C1 and C2, both machine translation outputs, five under C3, two under C5 and none of the six headlines. The wire, a translation of the body of a press statement, loses the term. The labels «παράτυποι μετανάστες» (irregular migrants) and «παράνομοι μετανάστες»

(illegal migrants), counted in any case form, appear in seven of the thirty language model outputs for this case: three under C3, two under C4 and two under C5, and none under C1 or C2.

5 A taxonomy of translational distortion

Five patterns recur across the documented cases of distortion. Each is defined by the operation performed on the source rather than by its effect on a reader, and the categories may co-occur. The definitions are those applied in coding the experimental outputs, and each pattern is illustrated from a case outside the three used in the experiments.

5.1 Context isolation

Material present in the source is removed, and its removal alters what the statement does. The test is the effect of the omission on the utterance, whatever other markers of the source survive it: where the statement still does what the German did, the pattern is not coded.

Merkel, speaking at Meschede on 17 May 2011, said: *Wir können nicht eine Währung haben und der eine kriegt ganz viel Urlaub und der andere ganz wenig* (We cannot have one currency in which one side gets a great deal of holiday and the other very little). Greek coverage in Τα Νέα, Πρώτο Θέμα and in.gr rendered the remark as «Κόψτε τις διακοπές και τις συντάξεις» (Cut the holidays and the pensions) and «Οι Έλληνες κάθονται με τα λεφτά μας» (The Greeks are sitting around on our money). The premise of a shared currency is gone, and with it the comparison of two parties within that currency. The Greek reader receives an instruction addressed to a national audience, where the German stated a condition under which a monetary union can work.

5.2 Tone modification

Evaluative intensity diverges from the source through lexical choice while the propositional content stands. Sharpening and flattening both count. Every case in the corpus also carries semantic manipulation, and the example shows the two patterns in one rendering.

Schäuble, in a Deutschlandfunk interview of 1 February 2012, said: *Wir können nur Hilfe leisten, wenn es nicht ein Fass ohne Boden ist* (We can only give help if it is not a bottomless barrel). news247.gr carried «Σε ένα βαρέλι δίχως πάτο, κάθε βοήθεια είναι μάταιη» (In a barrel with no bottom, all help is futile). The image survives intact. The German

attaches a condition to assistance, and the Greek turns it into a general claim that assistance is futile, which is the propositional shift. The evaluative weight sits in «μάταιη» (futile), which has no counterpart in the German.

5.3 Narrative-driven translation

The output aligns with a recurrent framing template in the receiving discourse that cannot be derived from the source. The coding requires documentary evidence that the template is in circulation.

Kauder said on 15 November 2011: *Jetzt auf einmal wird in Europa Deutsch gesprochen* (Now, all of a sudden, German is being spoken in Europe). Greek coverage in Το Βήμα, Ελευθεροτυπία, in.gr and newsit.gr carried both «Η Ευρώπη μιλάει τώρα γερμανικά» (Europe now speaks German) and «Η Ευρώπη υποτάσσεται στη Γερμανία» (Europe submits to Germany). The first stays with the German claim about which language now sets the terms of European argument. The second recasts it as the subordination of Europe to Germany, a reading the source does not supply, aligned with an antagonistic frame already established in Greek political discourse of the period (Stavrakakis and Katsambekis 2014; Katsambekis 2016).

5.4 Weaponization

The statement is reconstructed as attributing hostile intent or contempt, with the hostility introduced or amplified by the rendering. Hostility the German already carries, conveyed as it stands, is coded zero. Naming a person, or ascribing a negative quality to one, is neither necessary nor sufficient. Because weaponization relies on embedding the statement within an antagonistic receiving frame, it methodologically entails narrative-driven translation.

In an interview with Handelsblatt in 2022, Lindner said: *Nach der Pandemie sind die Schuldenquoten so hoch, dass diese Vorgabe bestimmte Länder objektiv überfordern würde* (Debt ratios after the pandemic are high enough that this requirement would, objectively, ask too much of certain countries). A signed opinion column on reporter.gr appeared under a headline set in quotation marks and attributed to him: «Εμπρός... πίσω στη λιτότητα» (Onward... back to austerity). The German weighs, in the conditional, what a fiscal requirement would do to countries it leaves unnamed. The Greek gives him a program of austerity, a word absent from the interview, and the quotation marks present it as his own. That attribution is where the hostility enters. The Greek text is commentary rather than a newsroom translation, which sets the case apart from the others in the corpus; it is also analyzed in Vraila and Livas (2025).

5.5 Semantic manipulation

Propositional content is added or removed, so that the truth conditions of the output differ from those of the source.

Friedrich, in *Der Spiegel* 9/2012 of 25 February 2012, said: *Ich rede nicht davon, Griechenland rauszuschmeißen, sondern Anreize für einen Austritt zu schaffen* (I am not talking about throwing Greece out, but about creating incentives for it to leave). news247.gr and Το Βήμα ran «Την Ελλάδα εκτός ευρώ θέλει ο Γερμανός υπουργός Εσωτερικών» (The German interior minister wants Greece out of the euro). The German separates expulsion from the creation of incentives and places the speaker on one side of that separation. The Greek attributes to him the position his sentence sets aside.

5.6 Incidence across the corpus

The classifications in Table A1, given in the appendix, were made by the author alone; the reliability figures apply to the experimental outputs.

Counts are given over the thirteen documented departures, since the three faithful cases carry no pattern and would depress every figure if counted.

Semantic manipulation is coded in all thirteen cases and does not discriminate within the corpus; the other four patterns record what accompanied it. Narrative-driven translation is coded in eleven cases and weaponization in seven. Tone modification is coded in five, context isolation in four. Individual cases carry between one and four patterns. Weaponization is coded in four of the seven departures from 2011 and 2012 and in two of the five from 2022 onward, all five of which carry narrative-driven translation.

5.7 Renderings that hold

Three cases in the corpus are faithful. Merkel's remark on multiculturalism of 16 October 2010, Gabriel's on Tsipras of 5 July 2015, and Merkel's *Wir schaffen das* (We can do this) of 31 August 2015 each reach Greek with their lexical and pragmatic force intact, the idiom in the Gabriel case included. Each was carried by an outlet that also carries departures in the corpus. The two from 2015 fall within the period the thirteen span, and the 2010 case precedes it.

6 Results

6.1 What the conditions can separate

Three accounts of where a departure comes from are available, and each predicts a different distribution of departures across the five conditions. They were formulated after the runs of March 2025 and March 2026 and serve as alternative readings of the results; none was set down in advance as a prediction.

On the first account, the dispositions belong to the model, and Greek journalistic convention shapes what it returns whatever it is asked. Incidence would then be similar under every condition, C1 included.

On the second, the repertoire is available to the model but dormant, and the wording of the instruction decides whether it is drawn on. This account expects incidence near zero under C1 and C2 and a rise once the instruction names a Greek news outlet.

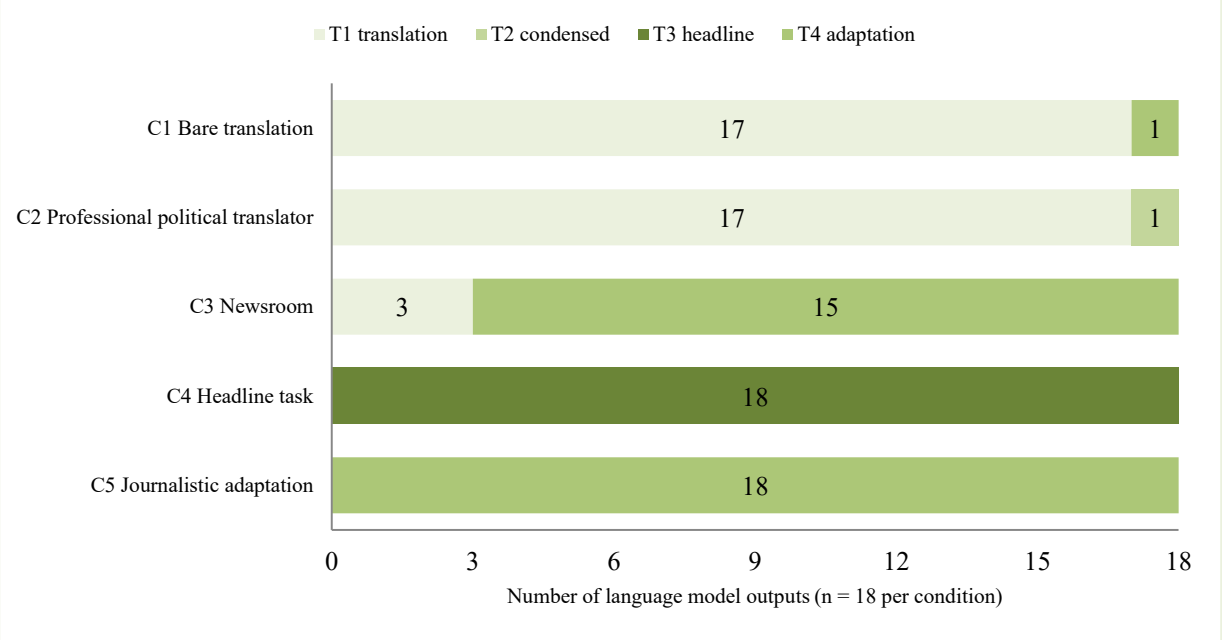
The third account holds that an instruction mentioning a news website is read as a request for a news text. In its strong form, C3 would return headlines, as C4 does. In its weaker form, C3 would return adaptations of the kind requested by name under C5.

6.2 Task type

The counts reported are the first coder's. The blind coder's are given where the two differ on a comparison the three accounts depend on.

Figure 1 classifies the language model outputs by the task they perform. Every output takes one of the four task types, so each bar sums to eighteen.

Figure 1. ***Task type by condition, language model outputs (n = 18 per condition).***



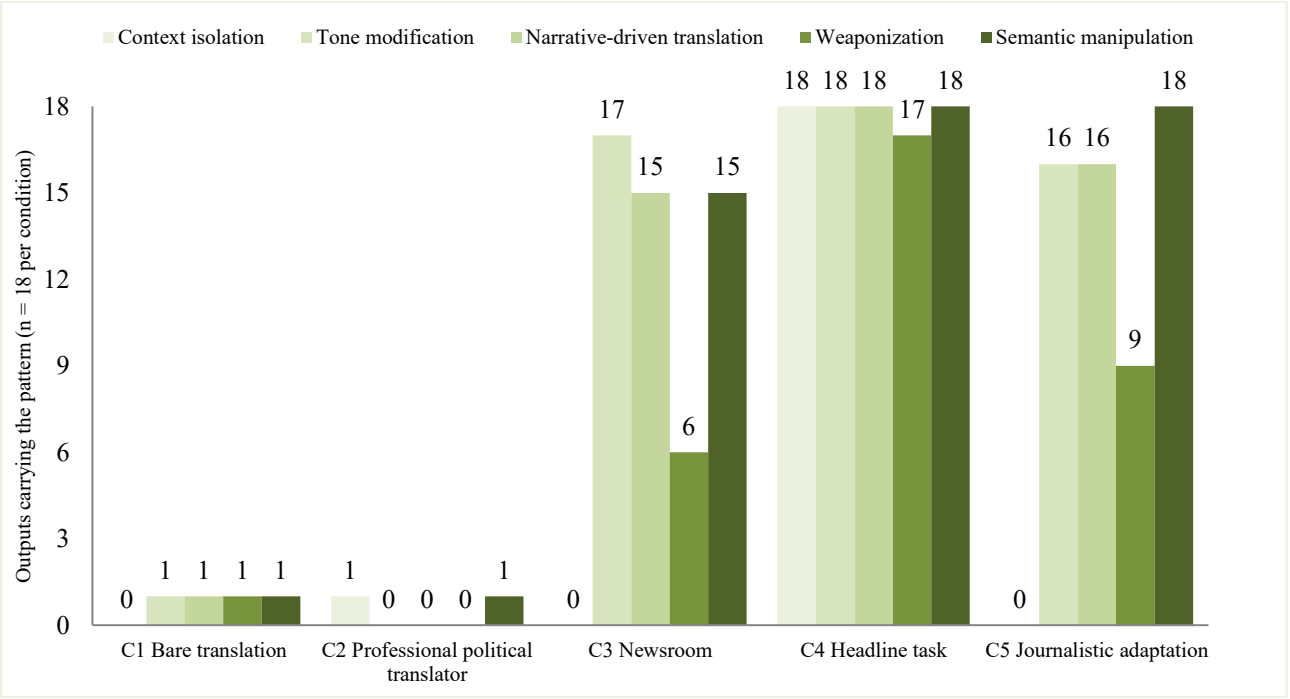
C3 and C4 divide completely: all eighteen outputs under C4 are headlines and none under C3 is. The blind coder counts seventeen headlines under C4 and none under C3. This excludes the strong form of the third account. The weaker form survives. C3 returns fifteen adaptations against one under C1, by either coder's count, and C5 returns eighteen by the first coder's count and fourteen by the blind coder's, who classified two of the other four as translations and two as condensed renderings.

Of the six NMT outputs, read against C1, five are full translations and one is an adaptation.

6.3 Pattern incidence

An output can carry more than one pattern, and each bar in Figure 2 is read against the eighteen outputs of its condition.

Figure 2. ***Pattern incidence by condition, language model outputs*** **(n = 18 per condition).**



Note. Hatching marks weaponization, the one variable reported below the agreement threshold of 0.70.

Under C1 and C2, seventeen of eighteen outputs are clean in each condition. The exception under C1 carries four of the five patterns and the exception under C2 carries two; both come from Case 1 and from GPT-4o. The blind coder finds sixteen clean outputs under C2, coding tone modification and semantic manipulation in a second Case 1 output from GPT-4o. Disaggregating outputs reveals consistent prompt sensitivity across both systems: GPT-4o and GPT-5.3 Instant alike preserved fidelity under C1–C2 and shifted toward adaptation under C3, justifying their aggregated reporting.

Weaponization is coded in six of the eighteen C3 outputs, against fifteen to seventeen for tone modification, narrative-driven translation and semantic manipulation.

The C4 and C5 bars record compliance with the instruction and serve as a benchmark for permitted deviation. Context isolation is coded in all eighteen headlines, a count that describes the genre more than the system.

Read as a benchmark, C5 falls closer to C3 than to C4. C4 reaches seventeen or eighteen on all five patterns. C5 records context isolation in none of its outputs by the first coder's count and in three by the blind coder's, and it stays within three outputs of C3 on every other variable. Here the second and third accounts are hardest to keep apart, and these data leave the choice between them open. The first account requires incidence under C1

and C2 at the level found under C3; the observed incidence there is one output in eighteen in each.

6.4 Variability across runs of one cell

Each of the thirty language model cells returned three different strings across its three runs. For each cell, pattern agreement is the proportion of the five patterns on which all three runs carry the same code. It averages 0.84 across the thirty cells. The coded profile thus varies less than the wording does, and least steadily in Case 1: six of its ten cells fall short of full agreement, against three of ten for Case 2 and two of ten for Case 3. The lowest value, 0.2, is returned by two cells. One of them is the C1 cell for Case 1 on GPT-4o, which holds the single patterned output under C1; its other two runs are clean.

6.5 Stability of the machine translation outputs

Within a single sitting, each of the six submissions returned the same string on resubmission. Set against the outputs collected in March 2025, five of the six differed. That comparison spans eighteen months and concerns change over time.

7 Conclusions

The same instructions, applied to the same passages, return full translations in seventeen of eighteen runs under C1 and under C2, and adaptations in fifteen of eighteen under C3. Of the technological conditions the third research question asks about, the instruction is the one this design isolates, and a newsroom that uses these systems makes an editorial choice when it words the request.

The patterns were derived from renderings by journalists and could be applied to machine outputs with agreement at or above the threshold on every variable except weaponization. In each of the three cases examined closely, the largest departure occurs where the text is excerpted, headlined or relayed, and the faithful renderings in the corpus come from outlets that also carry departures. Both observations point to the stage of production more than to the outlet, although three faithful cases can only suggest this. A wire extends the reach of a departure: the agency rendering of the Merz statement reached four outlets as one text.

The temporal evidence is thinner. The weaponized renderings of 2011 and 2012 are consistent with the antagonistic frame then in circulation (Stavrakakis and Katsambekis 2014), but a purposive corpus registers such distributions without measuring them. Machine translation systems also change, and a statement retranslated from the archive need not match an earlier machine rendering of it.

Two checks follow from the material, and neither requires a judgment about intent. A term lost in transfer, as *Freizügigkeit* was in the wire, is caught by comparing legal and institutional terms with their established equivalents, for which a translation obtained under a bare instruction can serve as reference. A statement rebuilt around a frame is caught by comparing modality, person and evaluative lexis with the source. For newsrooms, machine output is a draft whose instruction, system and date belong in the record, and requesting translation and headline as separate steps lets the headline be written against a full rendering. A check at a news agency reaches furthest, since a wire passes to every outlet that takes it.

The corpus is a purposive sample within one language pair, and the experiments cover three source texts. The reported counts are those of a coder who could see the condition of each output. The ten disagreements with the blind coder on narrative-driven translation all run toward the pattern; on tone modification, context isolation and semantic manipulation most run the other way. Weaponization, reported at its first-round value, has the lowest independent agreement of the coded variables, and its boundary with the neighboring categories is not cleanly drawn. Task type and semantic manipulation share a criterion, a shift in propositional content, and are not independent measurements. Language pairs without the asymmetry of the crisis years, and instructions that name an outlet's editorial standards, would show how far the finding reaches; how these tools enter editorial routine is a question for work with journalists.

7.1 Data availability

The data supporting this study are deposited in Zenodo (Vraila 2026, https://doi.org/10.5281/zenodo.22863810). The deposit holds the master log of the 96 coded outputs, with one row per output giving the source text, the prompt as issued, the system and its recorded identifier, the date, the unedited output and both coders' codes. It also holds the corpus of sixteen documented cases, the codebook with its dated amendments, the agreement spreadsheet with the formulas from which Cohen's κ was computed, and a README.

Vraila, Stavroula Paraskevi. 2026. Political Translation in the AI Era: Journalists as Dual Agents and Systemic Vulnerabilities in Cross-Linguistic Political Communication. In: L10N Journal. 1(5), pp. 39-61.

Bibliography

Baker, Mona. 2006. Translation and Conflict: A Narrative Account. London and New York: Routledge.

Bielsa, Esperança and Bassnett, Susan. 2009. Translation in Global News. London and New York: Routledge.

Conway, Kyle, Gramaccia, Julie Alice, Scholz, Nikita and Averbeck, Téana. 2025. Measuring Lexical Distance Between Parallel Corpora: The Case of AI-generated News Translation. In: Perspectives. Advance online publication. https://doi.org/10.1080/0907676X.2025.2590066

DeepL. 2025. DeepL Translator (March 2025). https://www.deepl.com/translator.

Entman, Robert M. 1993. Framing: Toward Clarification of a Fractured Paradigm. In: Journal of Communication. 43(4): pp. 51-58.

Freeth, Peter J. and Treviño, Rafael (eds.). 2024. Beyond the Translator's Invisibility: Critical Reflections and New Perspectives. Leuven: Leuven University Press. https://doi.org/10.11116/9789461665454.

Google. 2025. Google Translate (March 2025). https://translate.google.com/.

Hagar, Nick and Diakopoulos, Nicholas. 2019. Optimizing Content with A/B Headline Testing: Changing Newsroom Practices. In: Media and Communication. 7(1): pp. 117-127. https://doi.org/10.17645/mac.v7i1.1801.

House, Juliane. 2015. Translation Quality Assessment: Past and Present. London and New York: Routledge.

Kalantari, Esmaeil. 2022. Journalistic Translation: A Gate at Which Journalism Studies and Translation Studies Meet. In: Journalism. 23(7): pp. 1411-1429. https://doi.org/10.1177/14648849221074516.

Katsambekis, Giorgos. 2016. Radical Left Populism in Contemporary Greece: Syriza's Trajectory from Minoritarian Opposition to Power. In: Constellations. 23(3): pp. 391-403. https://doi.org/10.1111/1467-8675.12234.

Munday, Jeremy. 2012. Evaluation in Translation: Critical Points of Translator Decision-Making. London and New York: Routledge.

OpenAI. 2025. GPT-4o (March 2025). https://openai.com/.

OpenAI. 2026. GPT-5.3 Instant (March 2026). https://openai.com/.

Paullada, Amandalynne and Haroutunian, Levon. 2026. Language Justice and Machine Translation on Social Media. In: Modern Languages Open. 0(1): Article 7. https://doi.org/10.3828/mlo.v0i0.466.

Prates, Marcelo O. R., Avelar, Pedro H. and Lamb, Luis C. 2020. Assessing Gender Bias in Machine Translation: A Case Study with Google Translate. In: Neural Computing and Applications. 32: pp. 6363-6381.

Schäffner, Christina. 2004. Political Discourse Analysis from the Point of View of Translation Studies. In: Journal of Language and Politics. 3(1): pp. 117-150.

Schäffner, Christina. 2012. Unknown Agents in Translated Political Discourse. In: Target. 24(1): pp. 103-125.

Stavrakakis, Yannis and Katsambekis, Giorgos. 2014. Left-Wing Populism in the European Periphery: The Case of SYRIZA. In: Journal of Political Ideologies. 19(2): pp. 119-142. https://doi.org/10.1080/13569317.2014.909266.

Triantafyllou, Sotirios, Panagopoulos, Andreas M. and Kapos, Panagiotis. 2025. AI Pioneers and Stragglers in Greece: Challenges, Gaps, and Opportunities for Journalists and Media. In: Societies. 15(8): Article 209. https://doi.org/10.3390/soc15080209.

Vraila, Stavroula Paraskevi. 2026. Political Translation in the AI Era: Journalists as Dual Agents and Systemic Vulnerabilities in Cross-Linguistic Political Communication. In: L10N Journal. 1(5), pp. 39-61.

Valdeón, Roberto A. 2023. On the Cross-Disciplinary Conundrum: The Conceptualization of Translation in Translation and Journalism Studies. In: Translation Studies. 16(2): pp. 244-260. https://doi.org/10.1080/14781700.2022.2162573.

Valdeón, Roberto A. 2025. Journalistic Translation Research Twenty-Five Years On: Methods, Topics and Prospects. In: Perspectives. 33(4): pp. 635-653. https://doi.org/10.1080/0907676X.2025.2516364.

Venuti, Lawrence. 1995. The Translator's Invisibility: A History of Translation. London and New York: Routledge.

Vraila, Stavroula Paraskevi. 2026. Political translation in the AI era: Experimental Dataset and Coding Log (Version v1) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.22863810

Vraila, Stavroula Paraskevi and Livas, Sotiris. 2025. Translation, Emotion, and Ideological Reconstruction: Translating Political Discourse in Journalistic Contexts. In: Köksal, Onur, Dolmacı, Mustafa and Ramizoğlu, Ruşen (eds.). 2025. 5th International Congress of Language and Translation Studies: Full-Text Book. Konya: Çizgi Kitabevi. pp. 799-826.

Quaranta, Barbara. 2026. Molisano to English: a qualitative case study of AI-translated culture-specific items in migration. In: L10N Journal. 1(5), pp. 62-99.

Molisano to English: a qualitative case study of AI-translated culture-specific items in migration

Barbara Quaranta

Università degli Studi del Molise

barbara.quaranta@unimol.it

Abstract

In this paper, I give an account of how some cultural items used in Southern Italy (Molise) were processed and translated in English through an adaptive translation AI, Lara by Translated (2025). Following my translation of a paper by Norberto Lombardi, a migration scholar and editor from Molise, I will outline a tentative analysis of how terms that are culturally situated were translated by AI. To this end, Lombardi's essay on Frank Monaco, an Italian American photographer of Molise origin, was selected because of a range of translational challenges posed by dialectal expressions, Italianized terms, and theoretical concepts that are deeply ingrained in the local narratives on migration. Following my experiments from January to September 2025, I have analyzed AI renderings of CSIs to consider how and to what extent they could influence the construction of narratives that may result in hindering or fostering visibility.

Keywords: AI translation, Culture-Specific Items (CSIs), cultural visibility, Southern Italy, migration

1 AI translation: opportunities and challenges

The changes brought forward by artificial intelligence are fast approaching in language translation (Mohamed et al. 2024; Song 2025, Sun et al. 2025). Neural machine translation and deep learning have made it possible to reach a high level of output quality. Such changes have profoundly affected the notion of what translation is, as well as perspectives on when, how, and who can translate (Sun et al. 2025).

AI translation is nowadays generally considered reliable, with accuracy and quality reported as constantly improving (Mohamed et al. 2024; Lim et al. 2024). Advancements in how nuances of meaning are translated, including «idiomatic expressions, cultural subtleties, and emotional tones» (Mohamed et al. 2024, 25575) have also been claimed. While this fast progress in neural machine translation has

produced many positive changes, concern has grown regarding some aspects. Ethical issues are raised in AI development and usage, calling for respect of cultural specificities (UNESCO 2021) and attention to translation education and training (Li & Liu 2025; Desjardin 2024). Moreover, data protection has been mentioned as a potential issue that may put the users' information at risk. The quality of data incorporated into LLMs has also been questioned, highlighting the use of bycatch data "that harbours bias and incomplete views of most languages and cultures" (DePalma & Lommel 2025, 33). Underrepresented languages and cultures in LLMs may not reach the same quality levels as English ones due to a lack of material available for deep learning (DePalma & Lommel 2025). Texts containing uncommon terms may undergo major undesired shifts of meaning. In this case, the notion of invisibility introduced by Venuti (1995) may be applicable: when automatically translated into other languages, culturally situated meaning could go unnoticed and unquestioned by AI users and their readership. Consequently, standardized content could be reproduced and amplified through LLM training. The need for analyzing AI-translated texts has been highlighted to provide guidance in context-specific translation issues (Mohamed et al. 2024). In addition, computational lexical studies on CSIs in bilingual translations mention flaws and subjectivity in their models (Lim et al. 2024).

This paper aims at reporting experiments of translating a text that has multiple features of cultural embeddedness and language specificity from Italian into English through an adaptive translation AI with the aim of answering the research questions below:

- What strategies are used by Lara (Translated 2025) to translate migration-related culture-specific items (CSIs) from Italian into English?
- What aspects should be considered when using Lara (Translated 2025) to process such CSIs into English?
- What types of visibility does Lara (Translated 2025) support when processing CSIs?

I will now briefly describe the original text before defining culture and CSIs, as well as illustrating the translation strategies I have used to analyze the text. I will also present the tool I used for translation and the procedure for the experiments. Finally, I will analyze and discuss the results to advance tentative conclusions.

2 An essay in Italian on migration and photography

The essay, whose title in Italian is *"Il 'ritorno' alle radici di un fotografo globale. Frank Monaco (1917-2007)"*, is among the publications stemming from an international conference organized by the University of Molise's CUSIAC (University Centre for

Italian American and Canadian Studies) by Norberto Lombardi,[1] a Molisano migration scholar and editor of four book series on migration for the publisher Cosmo Iannone. The essay is a biographic account of the work of an internationally renowned photographer, Frank Monaco, who was born in the United States, spent a period of his early life in Cantalupo del Sannio, the village in Molise (Italy) where his parents were born, and then moved to the United Kingdom.

After a biographical account of Monaco's family, Lombardi describes the photographer's work, with a particular focus on the photographs he took at the beginning of his career, depicting the community in Cantalupo del Sannio, profoundly shaken by WWII and emigration. The photographs display women wearing traditional clothing and carrying out farming activities. Lombardi uses terms related to family relationships, daily routines in small villages, religious festivals and traditions, as well as regional clothing. He also uses some dialectal terms and quotes the words of Frank Monaco in his correspondence with his mother. Monaco used to mix both English and Italian in his letters to his mother, also quoted in the essay. Lombardi confirmed this in an interview in which he also acknowledged the fact that communication with Monaco was particularly challenging at times, because, when they met, Lombardi could not speak English, while the photographer could speak the Cantalupo del Sannio dialect he learnt from his parents combined with many English expressions. This was reportedly an effective way of communicating between the two.

After describing Monaco's last artistic period in India, Lombardi analyzes his style, inscribing it in a broader meridionalist stance.

3 Translating culture-specific items from Italian to English

Mailhac (1996, 133-134) defines cultural references in a text as «a cultural entity which, due to its distance from the target culture, is characterized by a sufficient degree of opacity for the target reader to constitute a problem». Aixelà brings this further when defining CSIs as

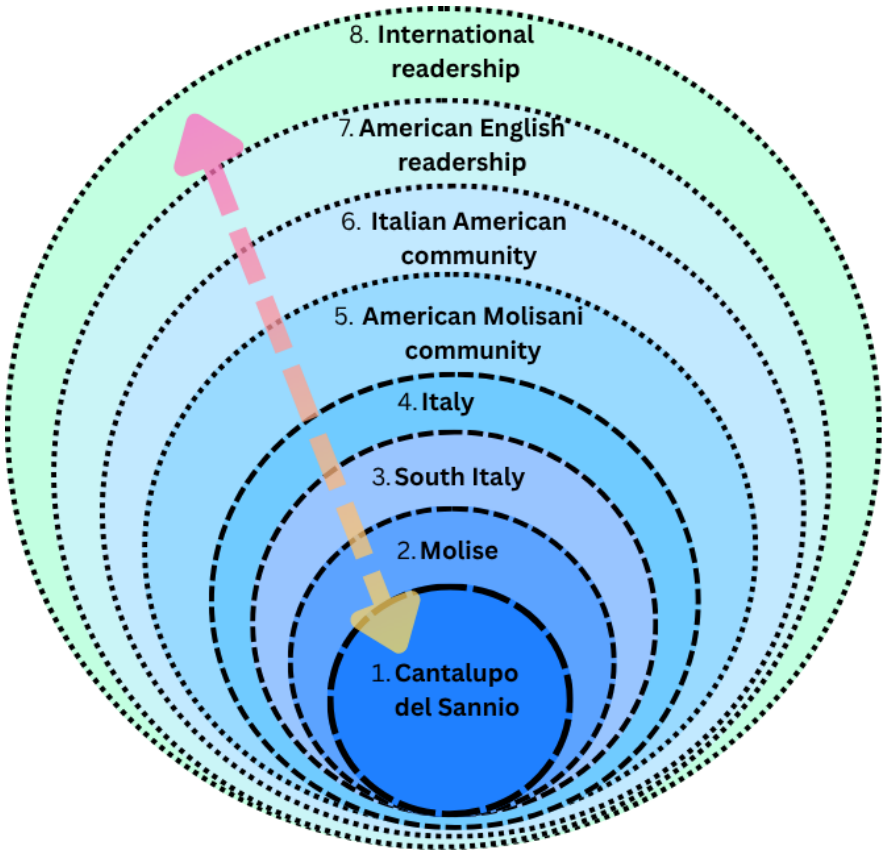
> *those textually actualized items whose function and connotations in a source text involve a translation problem in their transference to a target text, whenever this problem is a product of the nonexistence of the referred item or of its different intertextual status in the cultural system of the readers in the target language* (Aixelà 1996, 58).

[1] Lombardi was appointed a variety of local public and political roles among which member of the *Molisani nel Mondo* Council, Molise Region Councillor, member of the Southern-Italian Regions Committee for the Mezzogiorno special intervention, and Campobasso Town councillor. He was also a political advisor for the Italian Deputy-Minister of Italian Foreign Affairs (2006-2008) and a member of the National Committee for the promotion of the Italian language and culture in the world. He founded and coordinated the international Forum of Italians in the World and was a member of the General Council of Italians abroad (in Italian CGIE). He contributed to the founding of the Museum of the Italian Migration and promoted the Centre for the Study of Molisani in the World. During his studies and research, he became acquainted with a wide variety of renowned people whose origins were from Molise (Lombardi 2022).

Culture, however, cannot be solely identified with a country and national language (Bielsa 2022). Monceri (2006, 25) describes culture as a local communicative context in which individuals interact based on the information that they perceive as relevant and that may only partially correspond to what is perceived as such by others who are in the same spacetime continuum. To acquire a transcultural stance (Monceri 2019), which entails going beyond and moving across the borders of the usual notion of national culture or language, translation will be defined as a transcultural communicative method of (re)constructing meaning within a text by moving across spacetime continua that can be perceived as discontinuous by its readership.

In this paper, culture will consist of multiple spacetime continua, multiple neighboring concentric layers whose boundaries are not fixed, thus allowing for term exchange and recognition. Translation may foster or counter transcultural lexical mobility in an environment in which multiple layers correspond to as many spacetime continua as the translator is able to identify. Figure 1 shows how the bidirectional arrow of term transfer moves across the multiple layers that I have been able to identify for this case study.

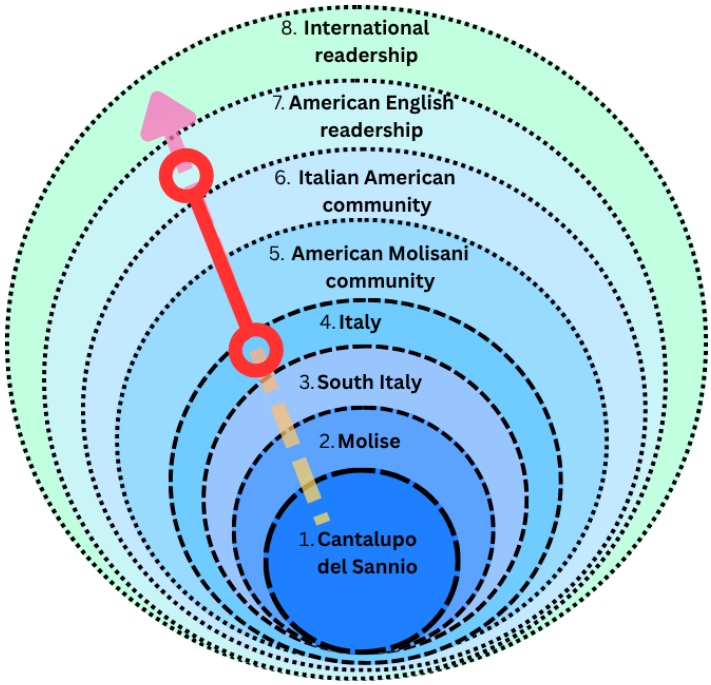
Figure 1.



Cantalupo del Sannio (first layer) can be viewed as the nucleus, the village in which Monaco's parents were born and the one where he decided to go back to. Within and outside the village, the regional terms of Molise (second layer) are also present. A third

cultural layer is the Southern Italian one, with its administrative peculiarities, as the events and the people included in Lombardi's essay come into play in Molise and in the nearby *Regioni* (administrative areas), including Campania and Lazio, and are inscribed in a meridionalist narrative. The fourth layer corresponds to Italy. Some migration-related Italian terms coming from all the preceding layers (1-3) were used untranslated also in English in the original text, when photographer Frank Monaco resorted to translanguaging to write letters to his mother. This aspect signals a community of Molisani abroad (fifth layer) and a broader extended Italian American community (sixth layer) that is also part of a larger American English readership (seventh layer), and of the international readership (eighth layer) of the essay finally translated into American English. As one can see from Figure 2, translating from Italian into American English may apparently be identified as a transfer that goes from layer 4 to 7 (red line), as the original text is in Italian and the translated essay is potentially accessible to the whole American English readership, including the international one. However, CSIs will more likely be known to communities that, despite usually being less visible, are the focus of the original text and are among the primary readership goal to reach for through translation. This is because the essay has been published by a New York-based publishing house dedicated to studies on the Italian American community. Thus, layers 1, 2, 3, 5 and 6 play a crucial role in establishing which terms should be identified and translated as CSIs.

Figure 2.

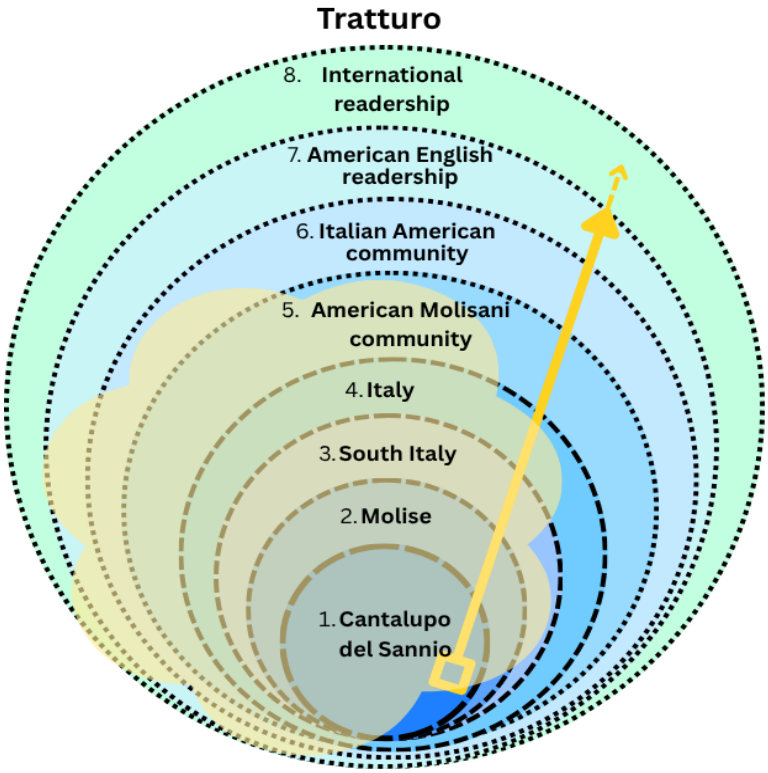


The number of CSIs a translator can perceive depends on their knowledge of the spacetime continua which the text touches upon. The more experience the translator has gained within each spacetime continuum, the more nuances they will be able to

grasp and, possibly, render. It would generally be up to the translator to perceive the degrees of opacity, acknowledge the existence of translational problems and consider adequate translation strategies depending on their audience's assumed amount of shared knowledge on the topic and on other factors such as text types, translation stakeholders, etc. (Aixelà 1997; Ivir 2003). The degree of a CSI's visibility thus ultimately rests on the translator's ability, attitude and willingness to perceive and convey difference while reconstructing meaning. Whenever translation is carried out through AI translation tools, the process of recognizing and rendering could be affected by the cultural bias potentially integrated in LLMs (DePalma & Lommel 2025, 33), and, in terms of Aixelà's framework (1997, 61), by varying degrees of manipulation based on term conservation and substitution.

As a result of this layering, translators should envisage the fact that term transference from one spacetime continuum to another (1-8) and the relevant adaptations might not be comprehensible by readers who are not familiar with all or some layers engaged in the process. For example, Figure 3 shows an analysis of the term *tratturo*, that is, a sheep and cattle track typical of Southern Italy, and its transfer from 1 to 8 (yellow arrow), as well as a prediction of the readership's shared knowledge related to the term (yellow cloud). A specific *tratturo* appears in the essay that may be known to most inhabitants in Cantalupo del Sannio. A high number of Molisani would also be familiar with the very same *tratturo*, while it may be less likely that people in layers 3 and 4 have an idea of the very same *tratturo* even though they may know other similar ones in their local spacetime continuum, other southern Italian Regioni. The transfer of the CSI *tratturo* is illustrated as originating in layer 1 and moving across the layers to get to the outer ones. It is important to stress the fact that being Italian does not automatically imply knowing and recognizing the term *tratturo*: only some Italians could be familiar with it, with southern Italians supposedly having more information. Moreover, the American community of Molisani may recognize the specific *tratturo* more easily than other Italians that could be part of layers 4 or 6.

Figure 3.



While many definitions of CSIs (Nida 1945; Ivir 2003) entail a stable relationship of equivalence and a measurable equivalent effect between the source and the target texts usually identified with a national language, for the purposes of this article, I will propose Aixelà's definition, as it appears to be in line with the process of translation outlined here. Prior experiments of computational linguistics dealing with CSI recognition based on quantitative criteria have highlighted their subjective nature (Lim et al. 2024). This is because CSIs are not predictable and universally recognizable, but rather they depend on the translator's subjective interpretation. It thus seems convenient to highlight their dependence on a series of variables. According to Aixelà,

> *a CSI does not exist of itself, but as the result of a conflict arising from any linguistically represented reference in a source text which, when transferred to the target language, poses a translation problem due to the nonexistence or to the different value (whether determined by ideology, usage, frequency, etc.) of the given item in the target language culture* (Aixelà 1997, 57).

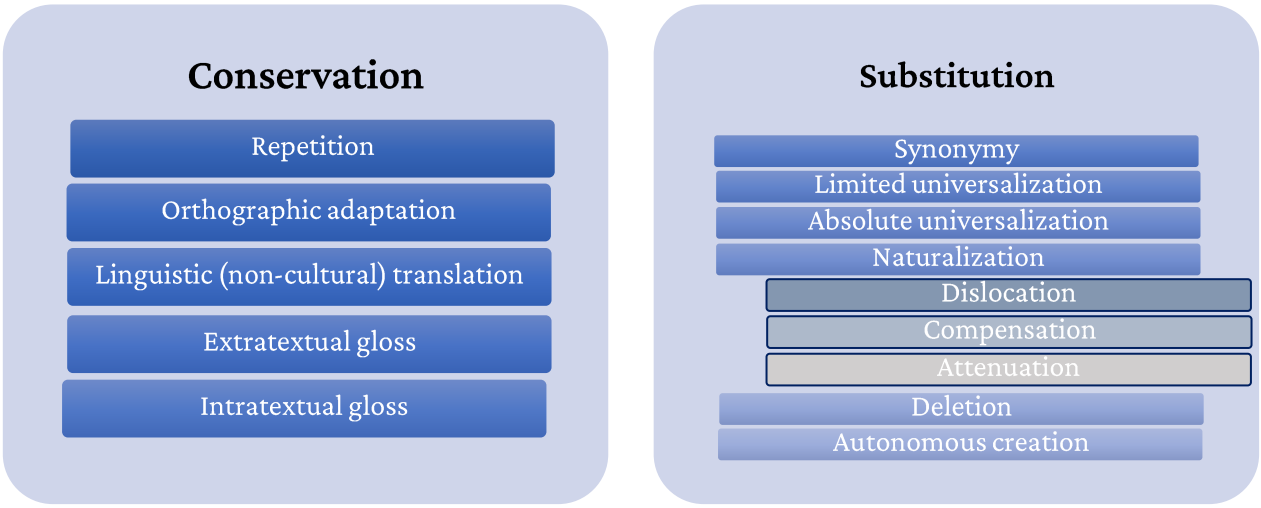
Multiple translations are therefore the result of interpretive acts performed by multiple translators based not only on spacetime variations, but also on the aspects that each translator perceives and deems valuable. This definition highlights how translation is fundamentally «an interpretive act that can be performed on any source text», also entailing that the source text can support «multiple and conflicting interpretations and therefore an equally heterogeneous succession of translations» (Venuti 2019, 175). This

case study is thus unavoidably entangled with the framework I have just sketched out, with my interpretation of the text and CSI retrieval and analysis, as I am a translator who was born in Campobasso (Molise, Italy) and has chosen to live there. I have also been exposed to all the layers that I have been able to identify above. I have selected the CSIs listed in this work, not others, drawing from my previous knowledge, and I have analyzed them based on the categories that I will illustrate below.

To ensure maximum recognition and a thorough categorization, I have integrated Aixelà's CSI classification with Newmark's typology (1988, 95), so that each CSI could be categorized more extensively. Aixelà (1996, 59) proposes a distinction between «proper nouns and common expressions (objects, institutions, habits and opinions restricted to each culture and that cannot be included in the field of proper names) ». Newmark (1988, 95) rediscusses Nida's (1945) categories and proposes the following: 1) Ecology, flora, fauna, winds, plains, hills (2) Material culture (artifacts): foods, clothes, houses and towns, transport, (3) Social culture (work and leisure), (4) Organizations, customs, activities, procedures, concepts, a) political and administrative, b) Religious, c) Artistic (5) Gestures and habits.

Regarding translation strategies, Aixelà (1997, 61-64) identifies the ones I have summarized in Figure 4 and illustrated in Appendix A.

Figure 4.



These strategies aim at analyzing CSIs' intercultural manipulation in terms of degrees of conservation and substitution with repetition being the most conservative strategy and autonomous creation being the most domesticating. Three secondary substitution strategies are compensation, dislocation and attenuation. CSI translation was previously analyzed only when performed by human translators and mainly involved literary texts (Zhu et al. 2022). In this case study, Aixelà's analytical strategies could be relevant in that AI translation contributes to creating texts that are to be read and understood by human readers. Mapping the way CSIs tend to be translated by AI

translation tools could help to establish the CSIs' visibility and ultimately to evaluate the presence of potential cultural bias in the LLMs that are employed and patterns of manipulation of terms that are uncommon or unknown to the specific AI.

4 Materials

Lara (Translated 2025) is an adaptive translation AI that combines machine translation with a large language model. It adapts to user-provided context, glossaries and translation memories without requiring model retraining.

At the time of the experiments, Lara (V1 and V2) offered three methods for translation: Translate Document, Translate Text and Interpreter. While the Interpreter option provided translation of instantly recorded voice messages, through the Translate Document option, users could upload files in 72 file formats and have them translated as whole documents, downloading the translated file directly in the original file format. It was also possible to translate through the Translate Text feature by copying and pasting text nto the platform interface. Through this feature, users could also get instant feedback on the quality of the output, as the Translate Text feature is also paired with an internal translation evaluation system to signal inconsistencies, other options based on context and styles and potential misunderstanding of the original.

Lara's users can choose up to three different styles for their target text: *faithful*, *fluid* and *creative*. The *faithful* style is described as leading to a precise and accurate translation that does not change the original structure and meaning; the *fluid* one is said to provide an easy-to-read translated text, thus emphasizing legibility. Thirdly, the *creative* translation option is reported as an inventive process which is inspired by the original text with the meaning being conveyed through realistic and evocative language. These features are available for both Translate Document and Translate Text methods, the translator can add a short description of the context of the translation to avoid ambiguities and decrease inconsistencies only in the Translate Text option.

The Learning mode is enabled by default. Through this feature, translation data are stored safely and used to enhance the model over time. To keep data safer, the Incognito mode can be switched on. In this case, the texts uploaded by users are not stored and used to refine the model.

On 4 June 2025, Translated publicly released a new version of the model (Lara V2) to which other features were added. Firstly, translation memories and glossaries could be created and uploaded both in the Translate Text and Translate Document options. The new version also introduced changes in the Translate Text method regarding context management and quality assessment. As a result, Lara's feedback on how the text has been translated is provided in the interface through pop-up messages, including advice

on which parts of the text should be checked or could be rendered differently. These indications also adapt to the specific context added by the translator.

5 Experiments

The experiments were carried out in Learning mode from 27 January to 12 September 2025, and translation of the source text in Italian underwent six phases.

In the first phase, the paper was translated through the Translate Document feature on 27 January 2025. The original Italian Microsoft Word doc. file was uploaded onto Lara (Translated 2025) and translated into American English using a Pro license. Two different English documents were downloaded. In the first one, the Fluid style had been selected, while for the second file the Creative option had been flagged. The Faithful style was not considered, as it is recommended for legal translations and manuals.

Upon initial observation, the text font had been changed, and footnotes had not been translated. This is why the CSIs in the source text's footnotes have been excluded from the Translate Document phases. Subsequently, I used both the documents to choose useful sentences and expressions based on the source text, its style and paragraphs. These two versions were a reference to produce the final text in American English. As translation progressed from 27 January until 13 June 2025, CSIs were identified as translation problems that arose from a discrepancy or inexistence of potential equivalents in the target language based on the two target texts downloaded in January. I thus identified the CSIs based on translation problems I perceived during translation, added them to a separate list where I classified them based on Aixelà's categorization (1996) and Newmark's typology. This procedure appeared to adequately respond to the varying nature of CSIs as described by Aixelà, because, as already mentioned, they do not exist on their own, independently from the context they belong to. On the contrary, they depend on the perception of the problem and on the translator's interpretation (Aixelà 1996, 57).

Each CSI was also classified based on the layers in which the CSI is expected to be better known and used by the most language users. By doing this, it is possible to visualize the degree of complexity when translating and, accordingly, consider the less visible layers that may affect translation choices.

Table 2 in Appendix B shows the complete list of the CSIs I have identified except for the ones that were part of the untranslated footnotes. In Table 2, for each CSI, I have included a definition in English, the context it was selected from in the Italian source text, the relevant layers and the classification (proper nouns and common expressions) based on Aixelà's distinction as well as Newmark's typology. I obtained a list of 23 CSIs among which there were 8 common expressions mainly classified as pertaining to Newmark's (3) social culture, (4) concepts and (5) gestures and habits. Out of the

remaining 15 proper nouns, 5 were loaded and most of them regarded ecology, material and social culture, and organizations. In Tables 3 and 4, Appendices C and D respectively, an overview of how proper nouns and common expressions were translated in Fluid and Creative styles is provided, with an analysis of the kind of manipulation strategies based on Aixelà's conservation and substitution categories.

In Phase 2, problematic sentences, paragraphs and expressions were checked and the translation finalized. This phase lasted from April to June 2025.

After the release of Lara V2 (Translated 2025), subsequent experiments were added. In Phase 3, the same source text was uploaded onto Lara and whole document-translation was performed again on 12 September 2025 from Italian into American English, resulting in two additional Microsoft Word .doc files. One was translated in the Fluid style, another in the Creative one. Both texts were created in Learning mode.

In this stage, the aim was to check whether there would be noticeable changes in the rendering of the same CSIs because of having performed initial translations on 27 January using the Learning mode. While in Incognito Mode the uploaded texts are not retained and are permanently deleted immediately after processing, in Learning mode, such texts and any accompanying contextual information are securely stored in an encrypted form and may contribute to improving the model's overall quality over time.

CSI renderings are reported in Tables 5 and 6, Appendices E and F. In the Fluid translated document, while there was recurring absolute universalization in Phase 1, in Phase 3 more conservative strategies appeared to have been used, such as repetition and synonymy. Similarly, in the Creative document, the preponderant absolute universalization strategy was substituted with synonymy; a limited number of terms seem to have been autonomously created. To check whether updates linked to having previously used the Learning mode could affect future CSI translation, I decided to add Phase 4 and investigate whether there would be any changes in the translation of the same CSIs through the Translate Text method.

During Phase 4, problematic sentences related to the presence of CSIs were translated through Translate Text on 12 September from Italian to American English in Learning mode. For terms *creanza*, *restanza*, *compaesani*, *Merica*, and *tratturo*, trials were run after applying the context *"academic paper on migration and photography"*. The whole paragraphs that contained the CSIs were copied and pasted onto Lara's user interface with a Pro account.

As per Figures 5 to 15 below, the translation of these terms through the Translate text method only partially confirms results in Tables 5 and 6, possibly pointing at a difference in text processing when using the Translate Document and Translate Text methods.

For the term *creanza*, synonyms are introduced in the Fluid style like *politeness* and *courtesy*, while in the Creative one, repetition is proposed once as well as deletion and autonomous creation.

Figure 5.

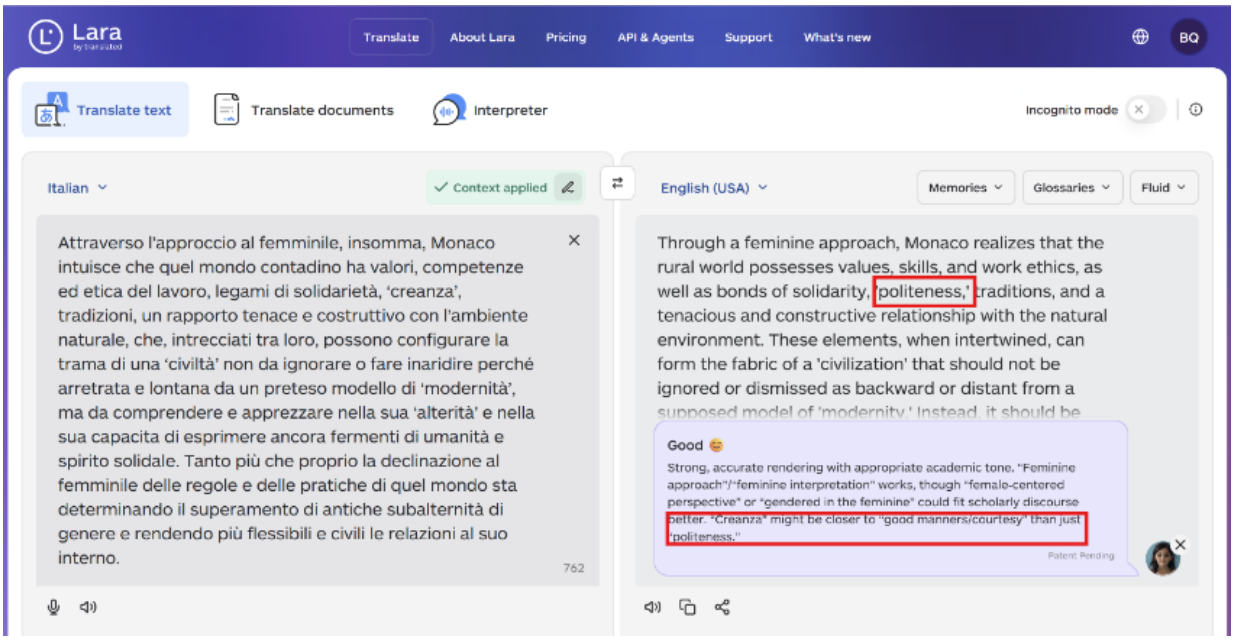


Figure 6.

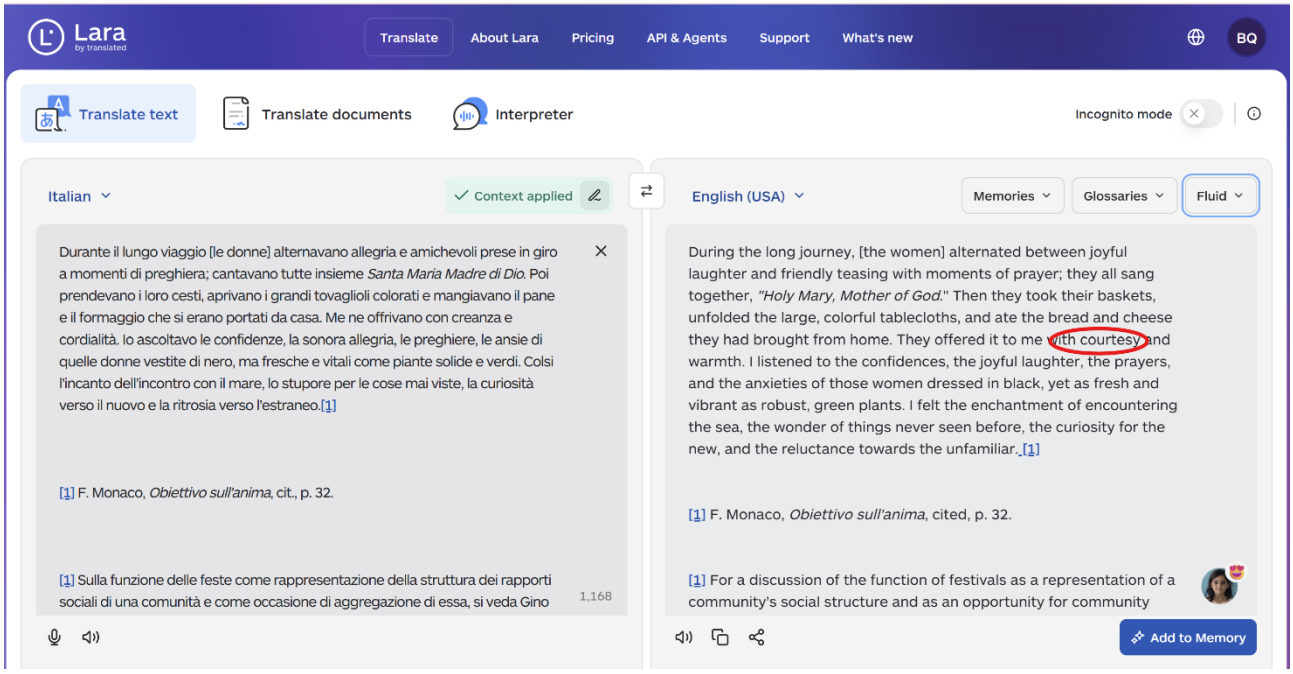


Figure 7.

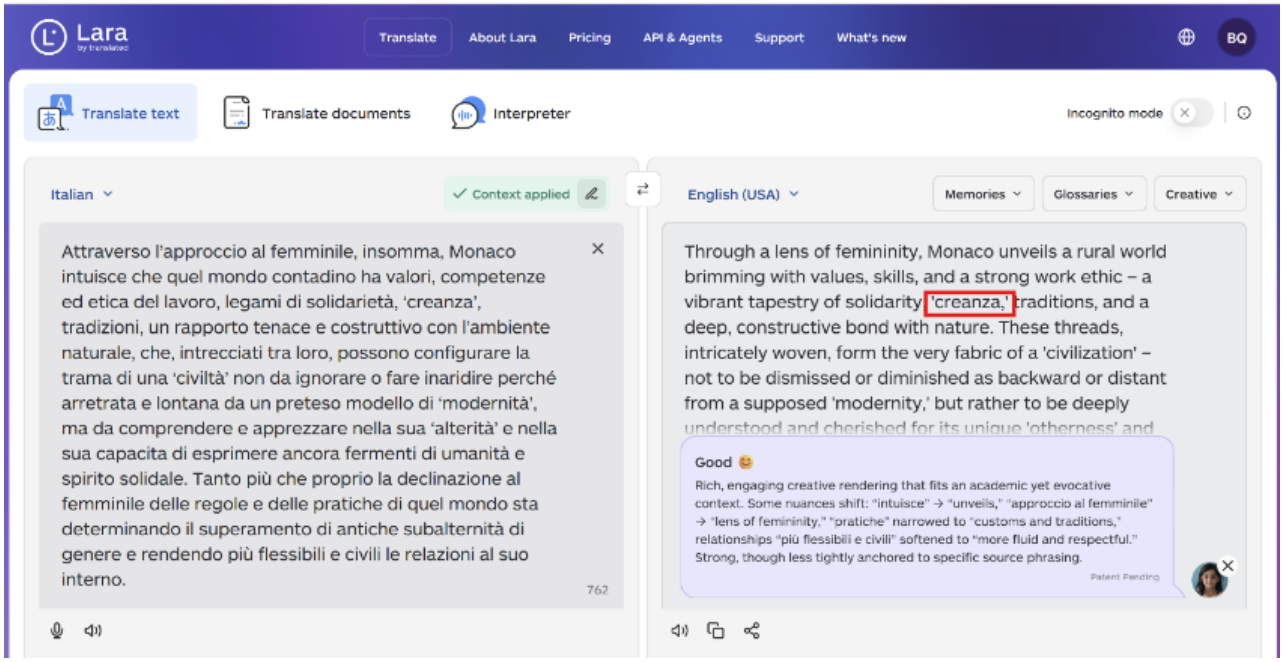
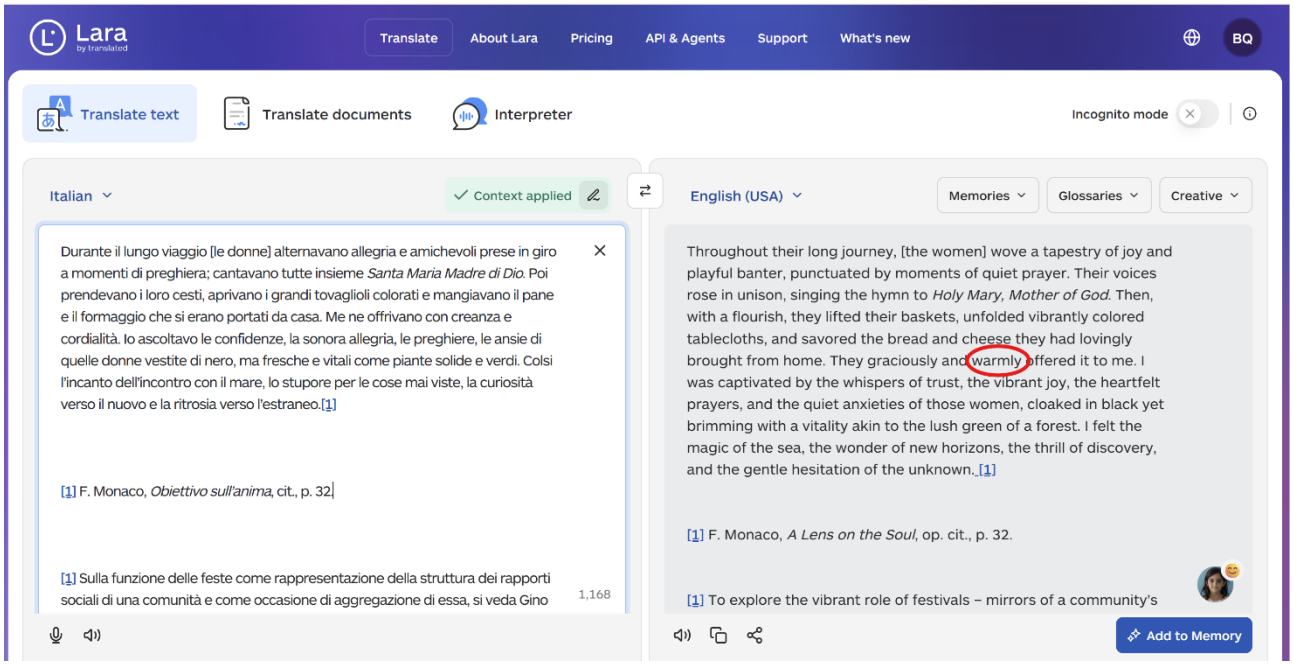


Figure 8.



The Fluid and Creative renderings of the term *compaesani* both provide the correct interpretation through synonymy.

Figure 9.

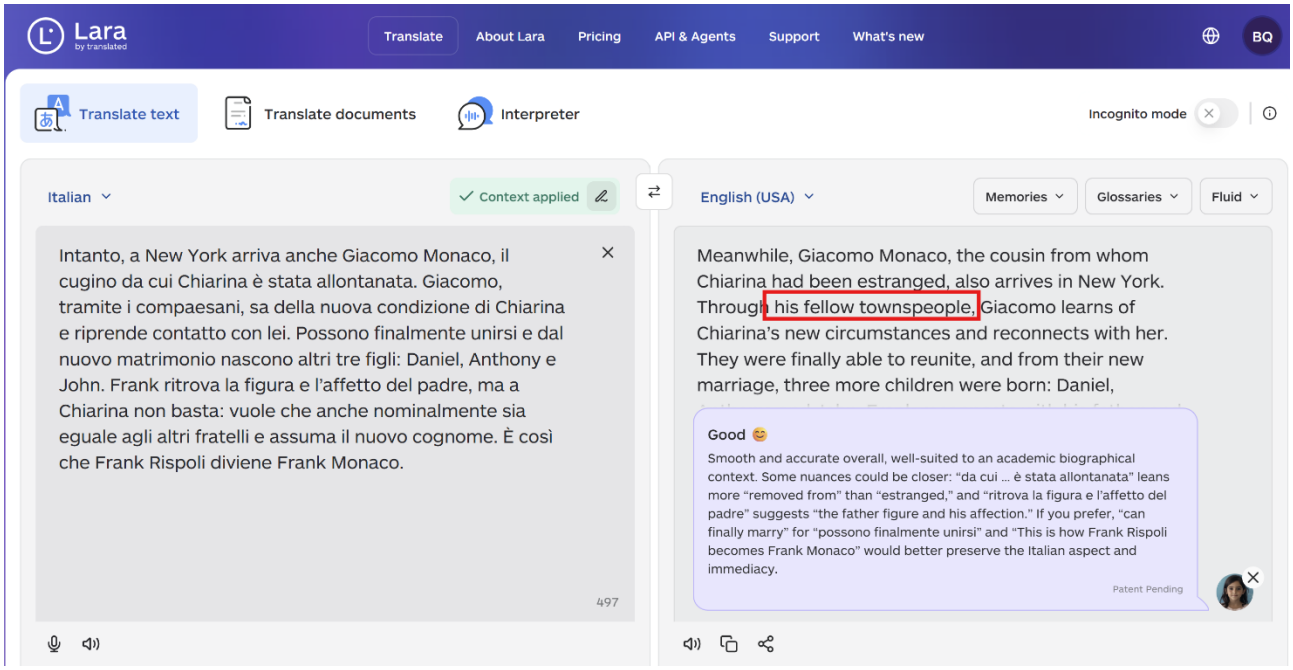
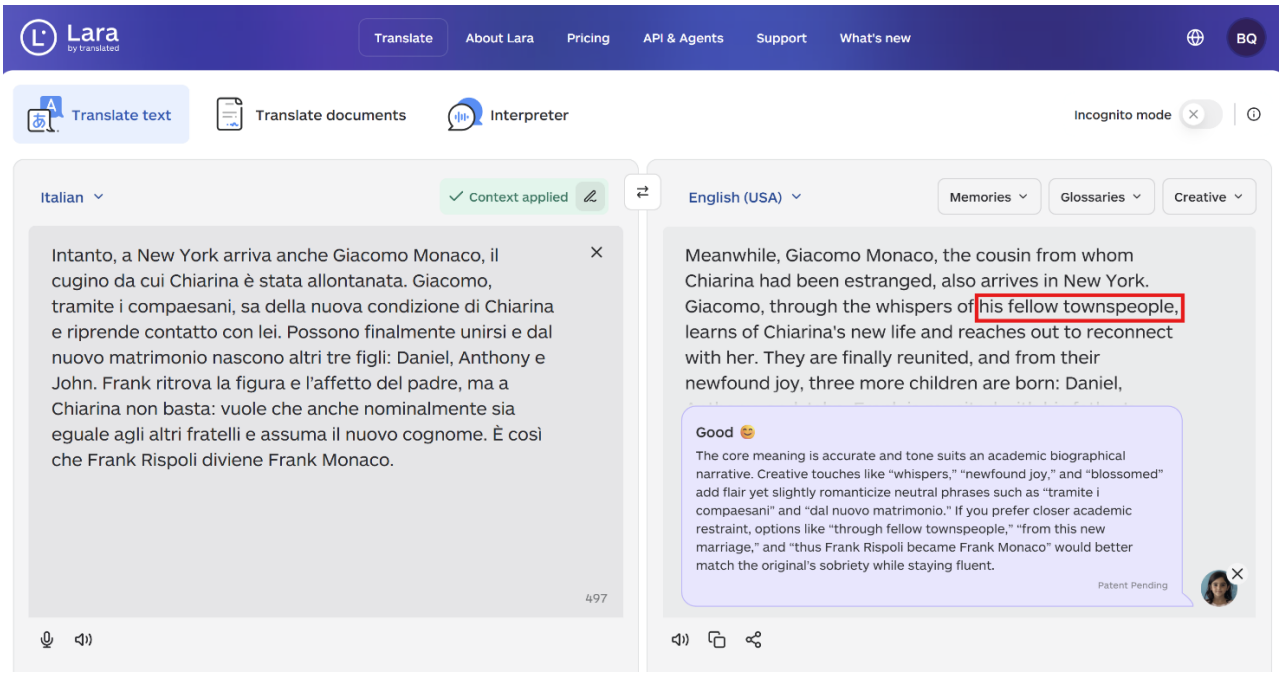


Figure 10.



The dialectal form *'Merica* was repeated in the Fluid style, while it was universalized in the Creative one.

Figure 11.

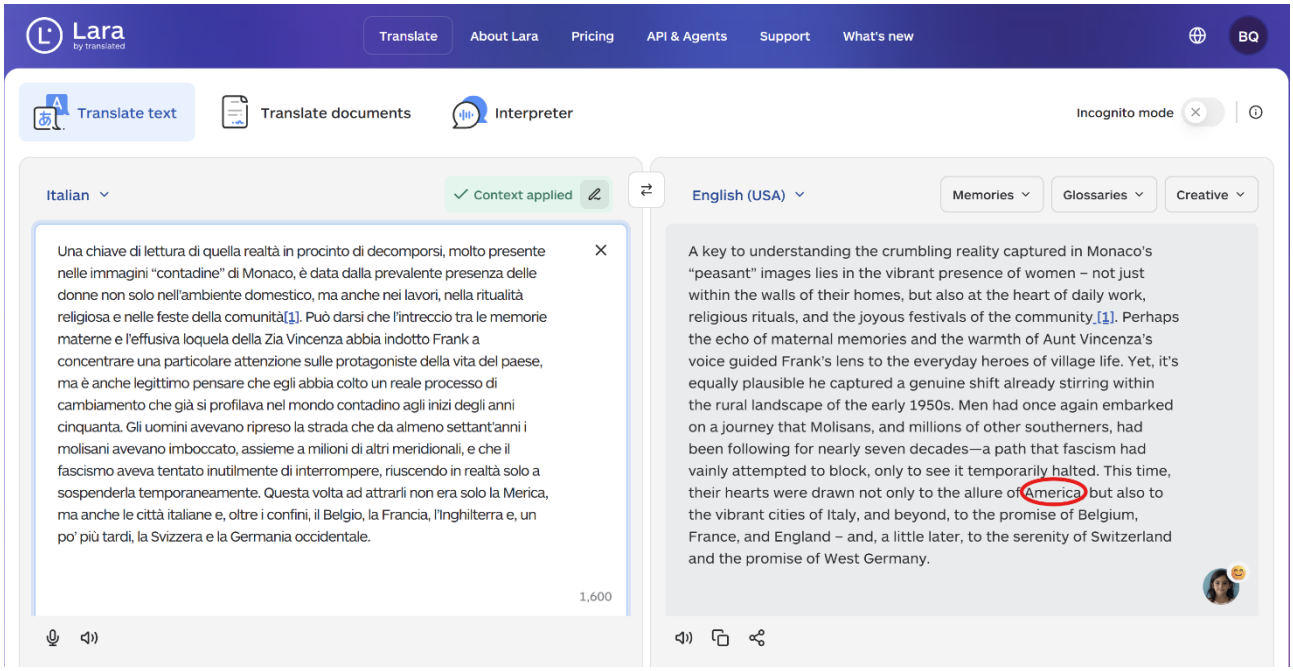
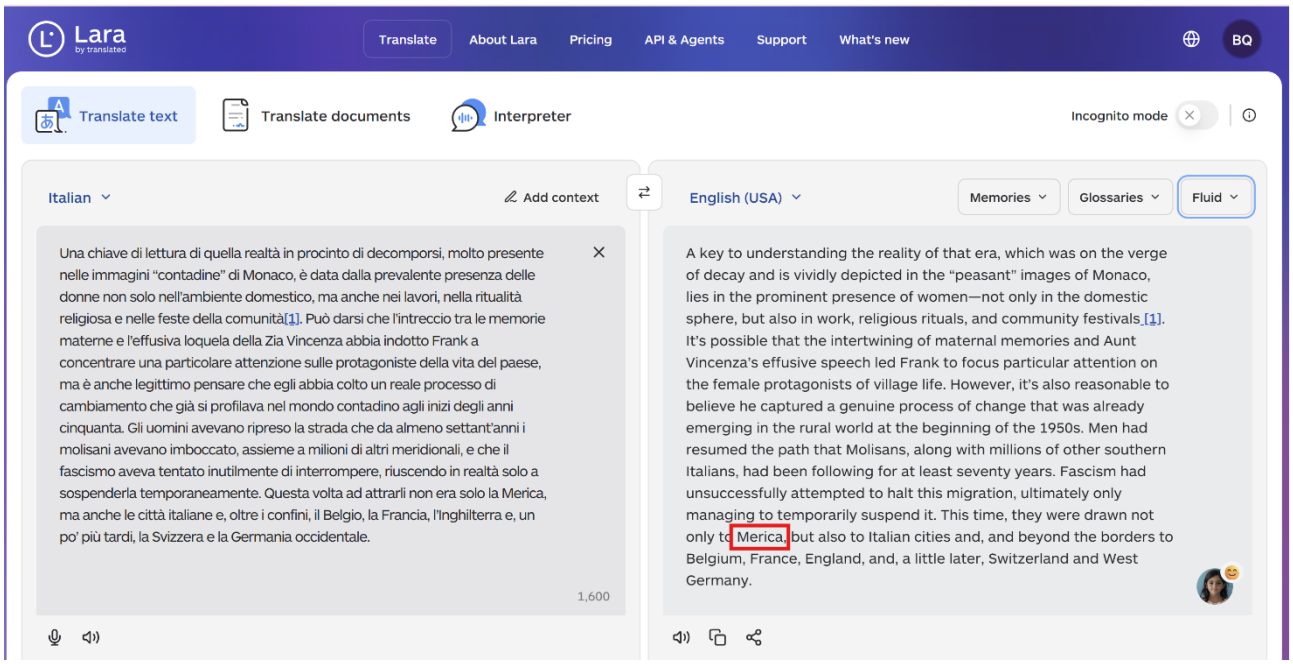
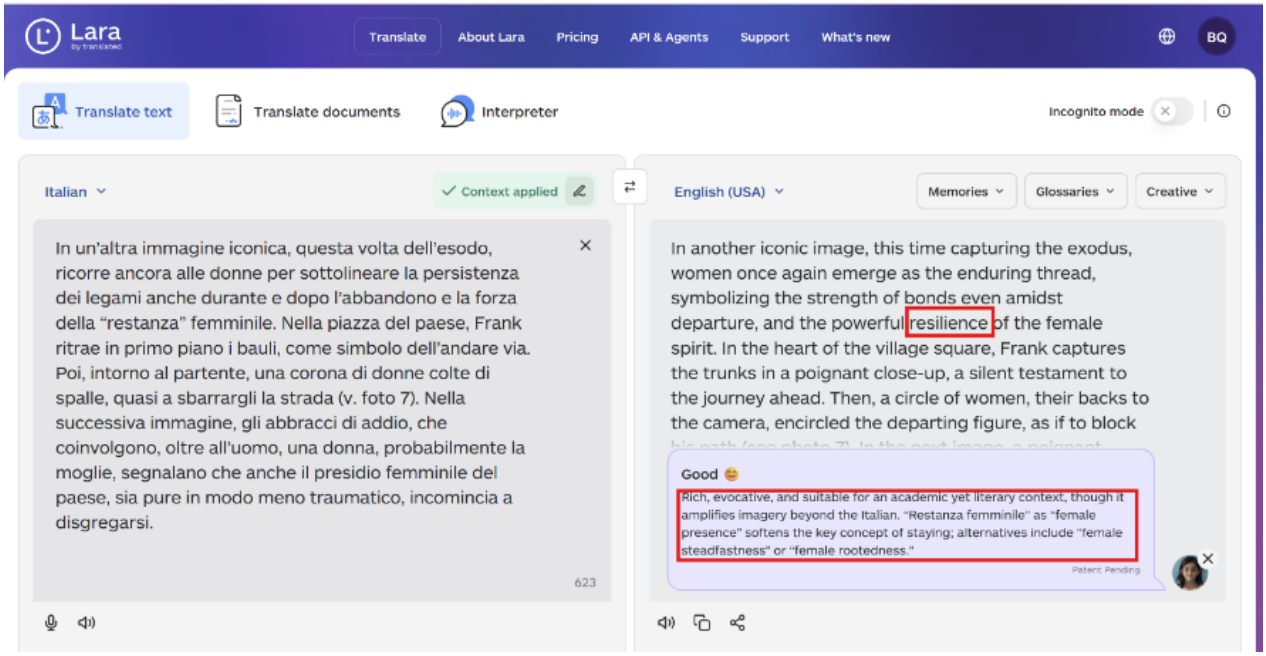


Figure 12.



The concept of *restanza* was translated using synonymy. What can be noticed here is that the explanations provided by Lara in the bottom left balloon describe it as being «natural and precise» and «rich, evocative and suitable» for an academic context also claiming to «amplify imagery beyond the Italian» and describing *restanza* as «contextually strong». As I will discuss below, this could be interpreted as a universalization strategy also through attenuation.

Figure 13.



Regarding the term *tratturo*, synonyms such as *cattle track* and *sheep track* were proposed respectively in the Fluid and the Creative styles.

Figure 14.

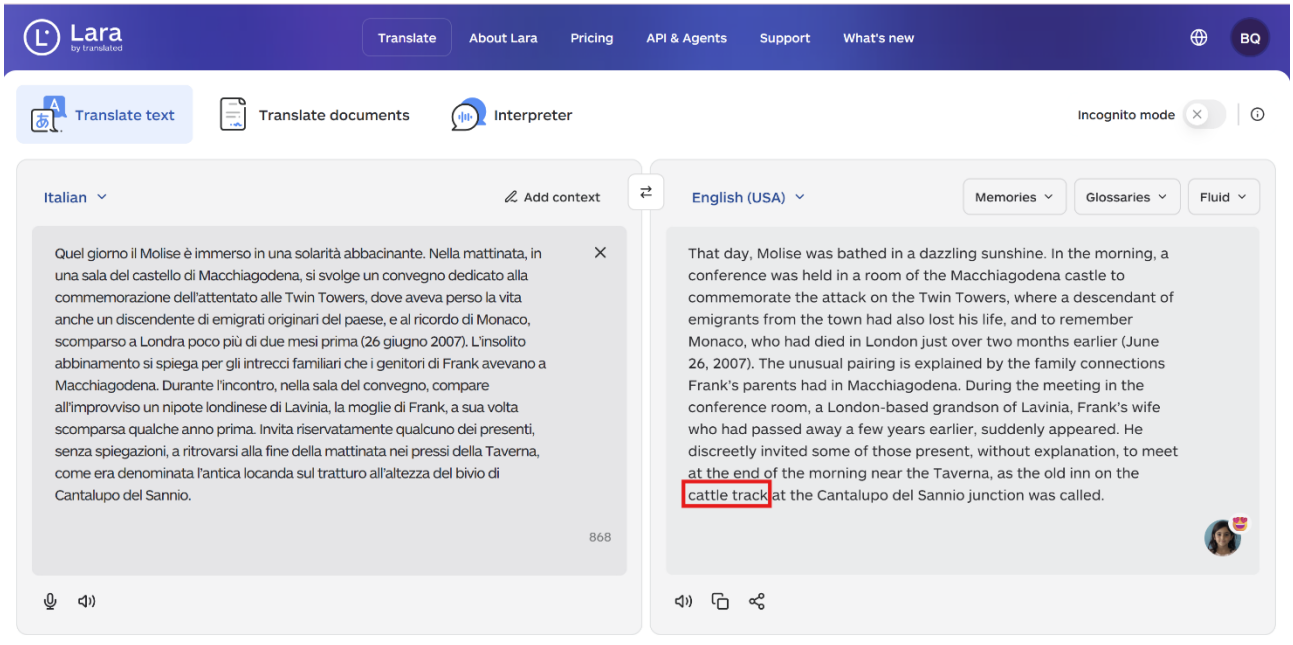
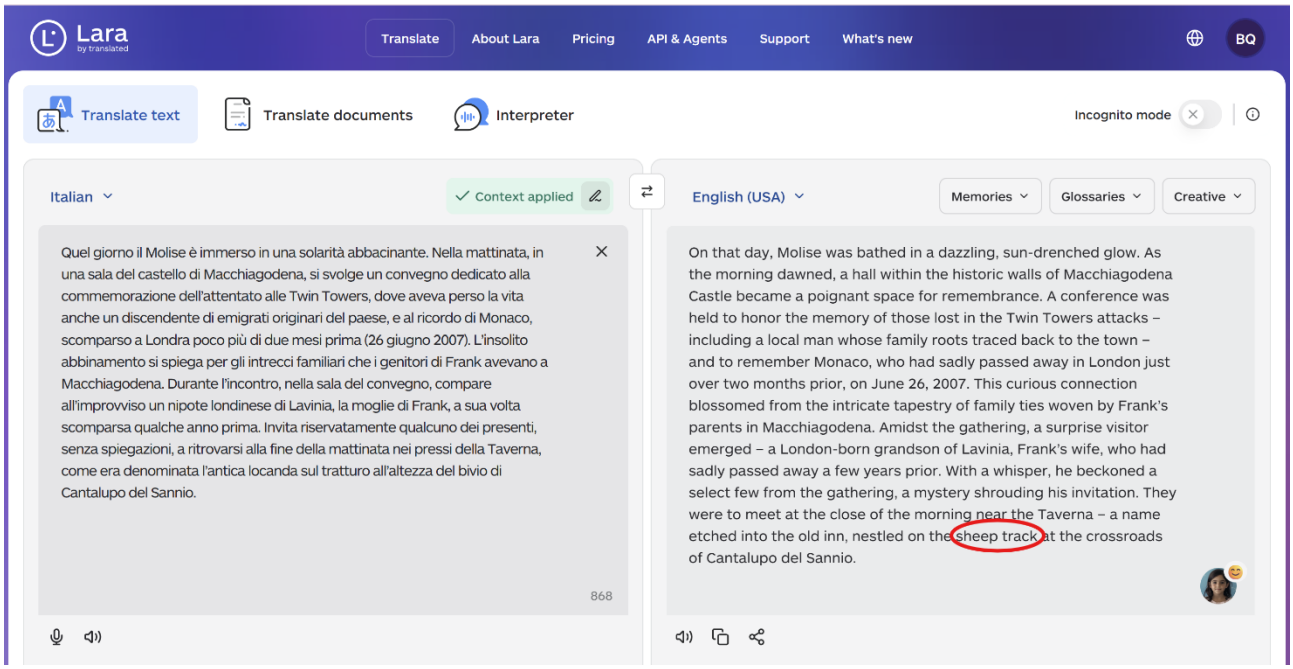


Figure 15.



Phases 5 and 6 envisaged revision by a native English translator and the editor, during either of which no edits about CSIs were required.

6 Discussion

The aim was to identify the «regularity of chosen options» as a «relevant and representative element, whereas exceptions will act as modifying factors» (Aixelà 1996, 60). The renderings investigated here could influence the way culture-specific items are conveyed and received by professional and non-professional translators as well as by their target readership. When consolidated as viable translation solutions, these may affect the way cultures and languages are perceived, since AI translation tools reproduce knowledge patterns including potential cultural bias (DePalma & Lommel 2025).

With the sole exception of the demonym *Molisans,* derived from *molisano,* which appears to have been created autonomously in English, proper nouns, especially the ones related to places, classified as ecology (1) and material culture (2), were generally repeated and left unvaried, moving toward conservation and ensuring visibility. Loaded proper nouns related to social culture, religious and political organizations seem to have undergone a limited degree of domestication due to their being linguistically, not culturally translated or partially translated. Southern Italian dialectal variations of the demonym *mericani* and of the proper noun *Merica* appear to have been universalized to their standard English version *Americans* and *America,* except for *Merica* in Phase 4, which was repeated as in the original text. Loaded kinship-bearing

proper noun phrases, such as *zio Vincenzo*, were subjected to linguistic translation, similarly to what happened with the title of the religious hymn. Despite having initially been repeated as it was, *Santa Maria, Madre di Dio* was later translated linguistically even though no evidence was found that such a hymn exists in English.

I will now discuss a limited number of common expressions whose translations varied widely across Phases 1, 3 and 4.

6.1 *Paesani and compaesani*

Since *paese (Paese)* in Italian may refer to both a village and Italy as a country, although with a different spelling, *paesani* and *compaesani* are likely to engender misunderstandings as to which *paese* we refer to. If the reference to *paese* and *compaesano* is misread as referring to Italy as a country and an Italian person, in the field of migration, the value of a community of *compaesani* – people who originate from the same village – and the bonds of solidarity and trust among them, which helped newly emigrated Italians navigate unknown foreign cities in the past century, would be replaced with a more generic national pride that could easily be dismissed by meridionalist studies on the Southern Question and the unification of Italy (Gabaccia 2003). In this case, while there was a clear preponderance of the first meaning through limited universalization in Phase 1, partly in Phase 3 and in Phase 4, this second meaning was correctly recognized.

6.2 *Creanza*

This Italian term, very common in Southern dialects including Molisano, seems not to have been recognized and processed by Lara in Phases 1 and 3. In its first occurrence, *creanza* was mentioned by Frank Monaco to describe the women of Molise's hospitable demeanor during a pilgrimage. In my classification, *creanza* has been classified under both (4) customs and (5) gestures and habits. Initially translated as *trust* and *trusting* invitation in the Fluid and Creative styles, respectively, it may have undergone a process of autonomous creation. It seems plausible here that a Translation error has occurred, in which «the translated content is inconsistent with the meaning expressed in the original language» (Sun et al. 2024, 8). In Phase 4, the Fluid version proposed the synonym *courtesy* as a translation, which adequately describes the women's good manners. In addition, the Creative style also propounds a more fitting paraphrase which, through dislocation, seems to be closer to the meaning of *creanza* (See Appendix F).

In the term's second occurrence, a more abstract *sense of belonging* was proposed in Phase 1 leading to substitution, possibly through autonomous creation. In Phase 3, the Fluid version proposed a repetition of the Italian term *'creanza'*, while in the Creative

style, *creanza* was deleted. Phase 4 in this case did not confirm the solutions provided through the Translate Document method, suggesting either synonymy in the Fluid style or repetition in the Creative style.

As may have already emerged from the discussion of the previous terms, although the terms *trust* and *sense of belonging* could apparently be in line with the positive depiction of women's rural communities by Lombardi, the concept of *creanza* points to values that are strongly connected with hospitability and food, as per its origins as a loanword from Spanish (*criar,* to nurture) in sixteenth century Southern Italy (Beccaria 1968).

6.3 *Restanza*

Restanza specifically refers to the condition of the people from the South of Italy who deliberately choose to remain in their place of origin. Unwilling to break their bonds with it, they keep a proactive and constructive attitude towards remaining and inhabiting it. A term first coined by anthropologist Vito Teti (Teti 2022) within studies carried out in Southern Italy, *restanza* is a theoretical concept. Translations of Teti's publications into English tend to leave the term untranslated to preserve its peculiarity and independence. Based on this information, I have proposed here that it underwent a process of absolute universalization in both Fluid and Creative versions in Phase 1. This is because the term *resistance* seems to be a more neutral universal term to substitute this concept with. Translations in Phases 3 and 4 suggest repetition in the Fluid style and absolute universalization in the Creative version. This is because the concept of *resilience,* which may also be considered as an adequate synonym giving more resonance beyond the Italian language, has acquired a general universal status. The term is applied to many contexts both in the sciences and the humanities, and may fail to convey the peculiar situation and feelings of the *restanti* (Salvo et al. 2025). Exactly because *resilience* is a more commonly used term in English outside of its disciplinary uses (mostly in physics and psychology), it could result in attenuating such context specificities and fail to acknowledge a theoretical concept that has already been accepted and used internationally in its original Italian even when translated.

6.4 *Tratturo*

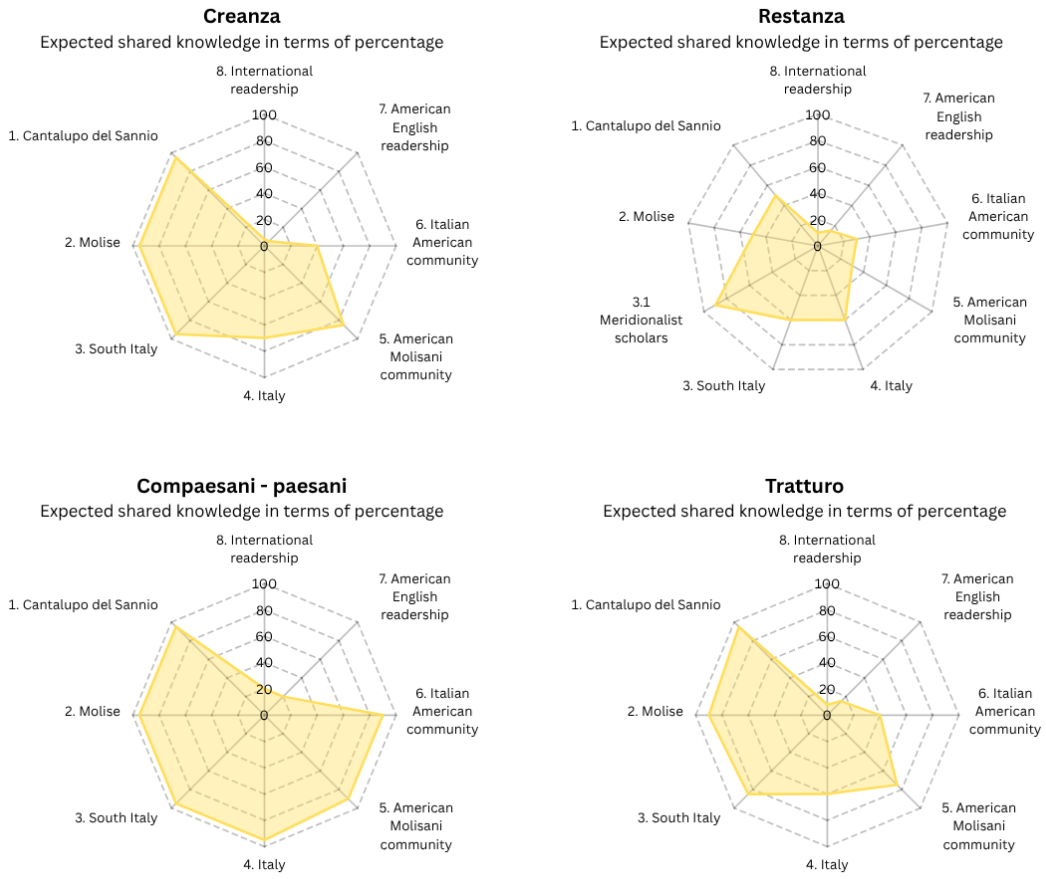
Translated with the term *sheep track* in both Fluid and Creative styles in Phases 1 and 3, this term has undergone absolute universalization. Phase 4 proposed *cattle track* in the Fluid style, while it retained *sheep track* in the Creative one. As explained in Table 1, *tratturo* is not only a sheep track, but it was also used for cattle and has been recognized as part of the UNESCO World Heritage in relation to transhumance. As Bindi (2022, 169) highlights, grassroots communities in Molise and Southern Italy nowadays aim at reviving depopulated territories and economies by revisiting pastoralism to promote

sustainable rural development and local identity. Universalization in this case may fail to convey such local attempts that could be relevant to the readership in the outer layers too, such as the American Molisani community and the Italian American one.

7 Summary of key findings

Figure 16 shows an overview of the expected shared knowledge across the different layers that I have identified. For the four terms discussed here, the layers in which they may be well known are yellow, while the ones where such terms could be transferred and/or are less known are empty.

Figure 16.



The analysis of the terms *paesani, creanza, restanza,* and *tratturo,* reveals the meridionalist reading of events narrated in Lombardi's essay about Molise and migration. Universalization of these terms and, sometimes, autonomous creation of similar or close ones in the case of *creanza* give back a blurred image in which the peculiarities that could and, in the author's intentions, should become known or be recognized by the Italian American community could go unnoticed. However, changes to CSI renderings during Phases 3 and 4 may signal adaptations. As shown in Figures 17

and 18, some of these initially recognized terms were translated more conservatively in Phase 3 both in the Fluid and Creative styles. This was the case of *creanza* in its first occurrence and *compaesani* and *paesani*. While the term *tratturo* did not undergo any changes throughout the experiments, phase 4 highlights how the second occurrence of *creanza* and the term *restanza* were still partially domesticated.

Figure 17.

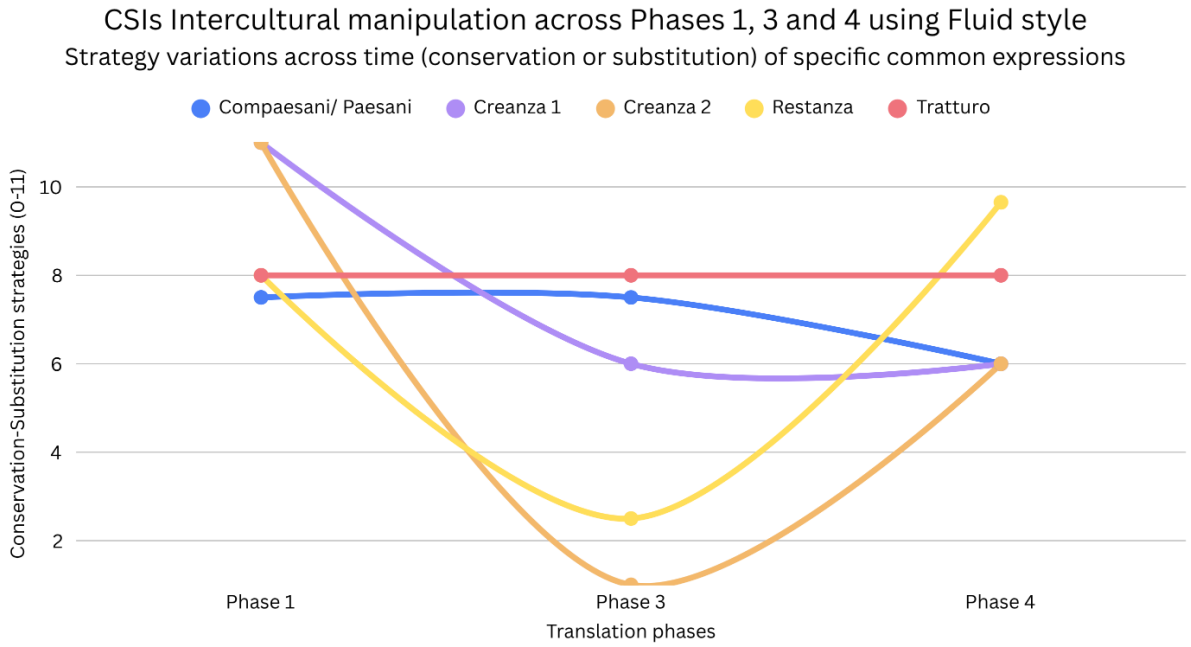
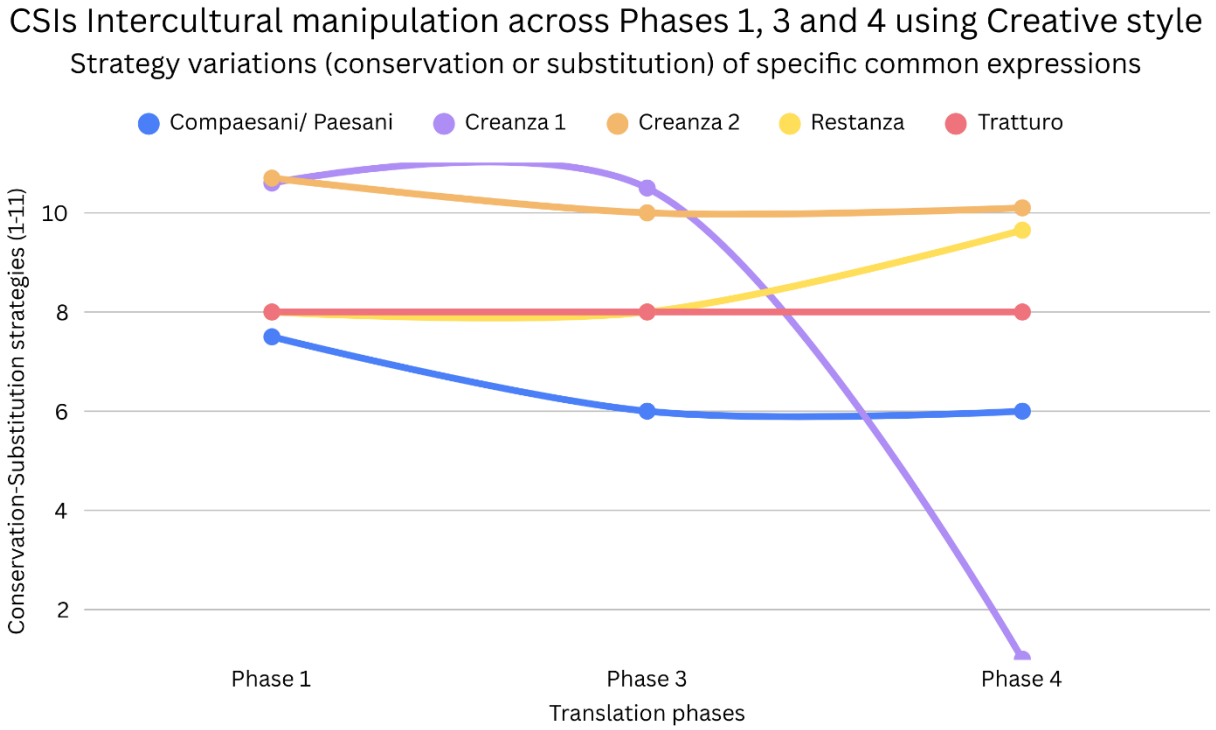


Figure 18.



My hypothesis is that because the source text was made available for internal evaluation because I processed it in Learning mode, more conservative strategies seem to have been employed in Phase 3 compared to terms proposed in Phase 1. Because Lara is an adaptive AI through which a limited selection of examples can be used for evaluation and system refinement, thanks to the support of specialized linguistic expertise the model's overall performance and behavior can be adjusted. This appears to have been the case in Phase 4, where even such terms as *creanza* and *restanza* have been processed differently.

8 Conclusions

AI translation tools have rapidly enhanced their output to such an extent that nuanced meanings and culturally loaded texts have been claimed to be effectively translated (Mohamed et al. 2024, Song 2025, Sun et al. 2025). However, there is a growing need for case studies to investigate these aspects and how they convey little-known and underrepresented languages and cultures (Mohamed et al. 2024).

Studies on computational culture-specific term recognition claim the possibility of identifying cultural keywords through quantitative comparison of word usage across languages (Lim et al. 2024). However, such studies conclude that there is a high level of subjectivity when classifications are proposed. Drawing from a qualitative stance, this

study has tried to highlight the fact that underrepresented cultures and languages in the form of CSIs may not be successfully processed through quantitative and statistical analysis only. CSI translation had been previously analyzed only when performed by human translators and mainly involved literary texts (Zhu et al. 2022). In the case of AI translation, these analytical tools may only partially be able to account for translation choices that may fall short of reasonable explanations or predictability because they involve machine processing. However, investigating the translation strategies of AI translation tools when processing CSIs by considering them as producing texts that are to be read and understood by human readers may help assess the quality of the solutions proposed to human professional and non-professional translators in terms of visibility. These solutions can be deemed acceptable or could be questioned by translators depending on their knowledge, social positioning, perceived value of the CSIs, time available for translations etc. Mapping the way CSIs tend to be translated may help establish the skills professional and non-professional translators need to possess when using AI translation tools, as well as the areas, fields and cultures that need to be covered when training LLMs to avoid cultural bias.

The aim of this case study was to analyze AI translation of a text rich in culture-specific items to outline possible recurring patterns in how they are recognized and translated. Apart from proper nouns related to ecology and social culture that generally tend to be repeated, little known terms referring to cultural values and practices translated into American English through Lara seem to have been, at least initially, universalized, also pointing to cultural invisibility (UNESCO 2021; DePalma & Lommel 2025). To confirm such limited results, similar studies in different fields, language pairs and using other AI translation tools will be necessary, also with the aim to establish virtuous protocols to help refine models.

Even if it may appear sensible to pre-identify CSIs and eventually build glossaries and translation memories for dialectal and regional terms, cultural flattening could always happen, due to their being textually actualized and interpreted, especially when dealing with new or uncommon concepts coming from inner cultural layers (see Fig.1). In case translators have not experienced such layers directly, researching and studying the history and customs of remote areas beforehand could help to outline a more complex, thus inclusive, multilayered translational environment within which relevant CSIs can be identified.

Results in phase 3 seem to suggest that, in case of translating texts rich in CSIs, the Translate Document method could work only after pre-instructing the model with similar texts, before processing the text to be translated. However, this will need further investigation. The Translate text method employed in phase 4 may suggest that another option when dealing with culturally embedded terms would be to promote a process of considering multiple options through the pop-up feedback. In addition, the use of

translation memories and glossaries has not been tested in this study and will need further exploration. This may also positively affect the way in which underrepresented cultures and languages are conveyed in the highest-resourced languages.

Whenever CSIs are processed and accepted unquestioningly, users may miss an opportunity to make valuable changes to the model, thus renouncing their agency and, consequently, a range of forms of visibility, including cultural and translator visibility. It might be with this in mind that Kaindl describes "systems of visibility that stipulate what is considered an acceptable and so socially recognized form of visibility" as not being explicitly regulated, but rather emerging "in the performative act of doing translation" (Kaindl 2024, 44). Keeping this in mind, one may suggest that communities of translators be established in which the translator's visibility may also contribute to renegotiating meanings that are culture specific. For example, the feature of pop-up feedback in Lara resembles what would happen in a real professional translational context in which translators, language specialists and other translation stakeholders negotiate and reconstruct meaning based on their viewpoints and, finally, agree on a shared solution. However, as of now, it does not seem possible to trace translation solutions back to the texts, corpora and individuals who have produced them to evaluate their reliability. An encouraging aspect, also highlighted above, is the fact that translation is gradually acquiring the status of a discussion among human-like entities. Even if it seems difficult for translators to be held accountable for their contributions to the system, providing digital visibility (Freeth 2024) to such translators and users also within AI professional and non-professional environments may contribute to establishing a more collaborative idea of translation.

Quaranta, Barbara. 2026. Molisano to English: a qualitative case study of AI-translated culture-specific items in migration. In: L10N Journal. 1(5), pp. 62-99.

Bibliography

Aixelá, Javier Franco. 1996. Culture-Specific Items in Translation. In: Álvarez, Román and Vidal, M. Carmen África (eds.). 1996. Translation, Power, Subversion. Clevedon, Philadelphia and Adelaide: Multilingual Matters, pp. 52–78.

Beccaria, Gian Luigi. 1968. Spagnolo e spagnoli in Italia: riflessi ispanici sulla lingua italiana del Cinque e del Seicento. Torino: Giappichelli.

Bielsa, Esperança. 2022. For a Translational Sociology: Illuminating Translation in Society, Theory and Research. In: European Journal of Social Theory. 25(3): pp. 403–421. https://doi.org/10.1177/13684310211002132.

Bindi, Letizia. 2022. Transhumance Is the New Black: Fragile Rangelands and Local Regeneration. In: Bindi, Letizia (ed.). 2022. Grazing Communities: Pastoralism on the Move and Biocultural Heritage Frictions. New York and Oxford: Berghahn Books, pp. 149–173. https://doi.org/10.1515/9781800736672-010.

DePalma, Donald A. and Lommel, Arle. 2025. Transforming Translation: The Evolution and Impact of AI on Language Transfer and Communication. In: Sun, Sanjun, Liu, Kanglong and Moratto, Riccardo (eds.). 2025. Translation Studies in the Age of Artificial Intelligence. London and New York: Routledge, pp. 18–41.

Desjardins, Renée. 2025. Reconfiguring the Translator's/Translation's Online and Digital (In)visibility. In: Freeth, Peter J. and Treviño, Rafael (eds.). 2025 [2024]. Beyond the Translator's Invisibility: Critical Reflections and New Perspectives. Leuven: Leuven University Press, pp. 49–72.

Freeth, Peter J. 2025. "Yes: I Translated It!": Visibility and the Performance of Translatorship in the Digital Paratextual Space. In: Freeth, Peter J. and Treviño, Rafael (eds.). 2025 [2024]. Beyond the Translator's Invisibility: Critical Reflections and New Perspectives. Leuven: Leuven University Press, pp. 147–172.

Gabaccia, Donna R. 2003. Italy's Many Diasporas. London and New York: Routledge.

Ivir, Vladimir. 2002. Translation of Culture and Culture of Translation. In: Studia Romanica et Anglica Zagrabiensia. 47–48: pp. 117–126. https://hrcak.srce.hr/21000.

Kaindl, Klaus. 2024. Visibilities of Translation – Visibilities of Translators: Reflections on the Theoretical Foundations of an Opaque Concept. Translated by Peter J. Freeth. In: Freeth, Peter J. and Treviño, Rafael (eds.). 2024. Beyond the Translator's Invisibility: Critical Reflections and New Perspectives. Leuven: Leuven University Press, pp. 31–48.

Li, Bo and Liu, Yuan. 2025. Ethical Issues in Translation Education in the AI Era. In: Sun, Sanjun, Liu, Kanglong and Moratto, Riccardo (eds.). 2025. Translation Studies in the Age of Artificial Intelligence. London and New York: Routledge, pp. 210–229.

Lim, Zheng Wei, Stuart, Harry, De Deyne, Simon, Regier, Terry, Vylomova, Ekaterina, Cohn, Trevor and Kemp, Charles. 2024. A Computational Approach to Identifying Cultural Keywords Across Languages. In: Cognitive Science. 48(1): e13402. https://doi.org/10.1111/cogs.13402.

Lombardi, Norberto. 2022. Altrove. Intellettuali molisani nella diaspora. Isernia: Cosmo Iannone.

Mailhac, Jean-Pierre. 1996. The Formulation of Translation Strategies for Cultural References. In: Hoffmann, Charlotte (ed.). 1996. Language, Culture and Communication in Contemporary Europe. Clevedon, Philadelphia and Adelaide: Multilingual Matters, pp. 132–151.

Mohamed, Yasir Abdelgadir, Akbar Khanan, Mohamed Bashir, Abdul Hakim H. M. Mohamed, Mousab A. E. Adiel and Muawia A. Elsadig. 2024. The Impact of Artificial Intelligence on Language Translation: A Review. In: IEEE Access 12: pp. 25553–25579. https://doi.org/10.1109/ACCESS.2024.3366802.

Quaranta, Barbara. 2026. Molisano to English: a qualitative case study of AI-translated culture-specific items in migration. In: L10N Journal. 1(5), pp. 62-99.

Monceri, Flavia. 2006. Interculturalità e comunicazione. Una prospettiva filosofica. Roma: Edizioni Lavoro.

Monceri, Flavia. 2019. Beyond Universality: Rethinking Transculturality and the Transcultural Self. In: Journal of Multicultural Discourses. 14(1): pp. 78–91. https://doi.org/10.1080/17447143.2019.1604715.

Newmark, Peter. 1988. A Textbook of Translation. London: Prentice-Hall International.

Nida, Eugene. 1945. Linguistics and Ethnology in Translation-Problems. In: WORD. 1(2): pp. 194–208. https://doi.org/10.1080/00437956.1945.11659254.

Salvo, Caterina, Tomnyuk, Viktoriia and Membretti, Andrea. 2025. Immobility Aspirations Among Young Adults in Italian Inner Areas: Towards a «Capability to Stay». In: Scienze Regionali – Italian Journal of Regional Science. 24(2): pp. 143–160. https://www.rivisteweb.it/doi/10.14650/108355.

Song, Z. 2025. Research on the Technological Development of AI Large Model and Its Application on Translation Practice in the GenAI Era. In: International Journal of Science and Technology, 9(3). http://www.itspoa.com/UploadFiles/2025-08/369/2025081309465766582.pdf.

Sun, Yujie, Sheng, Dongfang, Zhou, Zihan and Wu, Yifei. 2024. AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content. In: Humanities and Social Sciences Communications. 11: Article 1278. https://doi.org/10.1057/s41599-024-03811-x.

Sun, Sanjun, Liu, Kanglong and Moratto, Riccardo. 2025. Introduction: Navigating the Paradigm Shift – Translation Studies in the Age of AI. In: Sun, Sanjun, Liu, Kanglong and Moratto, Riccardo (eds.). 2025. Translation Studies in the Age of Artificial Intelligence. London and New York: Routledge, pp. 1–17.

Teti, Vito. 2022. La restanza. Torino: Einaudi.

Translated. 2025. Lara Translate, (V1 and V2). https://app.laratranslate.com/translate.

UNESCO. 2022. Recommendation on the Ethics of Artificial Intelligence. Paris: United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

Venuti, Lawrence. 1995. The Translator's Invisibility: A History of Translation. London and New York: Routledge.

Venuti, Lawrence. 2019. Contra Instrumentalism: A Translation Polemic. Lincoln: University of Nebraska Press.

Zhu, Hongxiang, Ang, Lay Hoon and Mansor, Nor Shahila. 2022. Genre-Based Translation Strategies on Cultural References: A Systematic Review. In: Journal of Language and Linguistic Studies. 18(1): pp. 898–912. https://doi.org/10.52462/jlls.227.

Appendix A

Table 1. Aixelà's (1996, 61-64) translation strategies for CSIs

Strategy	Category	Description
1. Repetition	Conservation	Keeps the original CSI unchanged. The translator reproduces the source-language item exactly to preserve its foreignness.
2. Orthographic Adaptation	Conservation	Adjusts spelling or writing system (e.g., transliteration or transcription) while preserving the original reference.
3. Linguistic (Non-cultural) Translation	Conservation	Translates the literal linguistic meaning without replacing the cultural reference. The CSI becomes easier to understand but still clearly belongs to the source culture.
4. Extratextual Gloss	Conservation	Adds explanatory information outside the main text, such as footnotes, endnotes, glossaries, or translator's notes.
5. Intratextual Gloss	Conservation	Explains the CSI within the body of the text itself, integrating clarification naturally into the translation.
6. Synonymy	Substitution	Replaces the CSI with a close synonym or a parallel expression when repetition would sound complicated or repetitive.
7. Limited Universalization	Substitution	Replaces the CSI with a more general item that still retains some cultural flavour, though less specific than the original.
8. Absolute Universalization	Substitution	Replaces the CSI with a completely neutral or universal expression, removing all source-culture specificity.
9. Naturalization	Substitution	Replaces the source CSI with an equivalent item from the target culture, making the text feel familiar to target readers.
10. Deletion	Substitution	Omits the CSI entirely when it is judged unnecessary or too difficult to translate effectively.
10.1 Dislocation	Substitution	Shifts the structural placement of a CSI to respect target language syntax/flow.
10.2 Compensation	Substitution	Balances textual or cultural loss by introducing different elements elsewhere in the text.
10.3 Attenuation	Substitution	Softens ideological, moral, or social intensity to fit target-culture values, while retaining the reference, it downgrades its vulgarity, sharpness, or shock value.
11. Autonomous Creation	Substitution	Introduces a new cultural reference or information that is not present in the source text to achieve a similar communicative or stylistic effect.

Appendix B

Table 2. List of CSIs extracted in Phase 1 with definition, indication of source layer and classification

	CSIs	Definition	CSIs in context (ITA)	Layers in which CSIs are better known	CSI classification
1	Arma dei Carabinieri / Corpo dei Carabinieri	The carabinieri corps is a section of the Italian Army.	L'America, nella famiglia di Frank, si materializza nel periodo della Grande emigrazione, quando il nonno materno, Francesco Compagnucci, decide di lasciare l'Arma dei carabinieri e di inserirsi nella grande corrente emigratoria che porta milioni di italiani e decine di migliaia di molisani a fare un'importante esperienza di lavoro e di vita negli Stati Uniti. Francesco ha l'idea di rappresentare la sua situazione e di chiedere aiuto nientemeno che al Comandante supremo del corpo dei carabinieri, il Re d'Italia, che a distanza di qualche tempo dispone in suo favore un soccorso economico.	1, 2, 3, 4, 5, 6	Proper noun, Organisation a) political and administrative
2	Biferno	A river in Molise	Dopo poco, il nipote londinese e alcuni parenti italiani guidano i convenuti verso un ponte su un torrente poco distante, il Rio, affluente del Biferno. Una giovane donna, con un sorriso, offre un fiore a ciascuno dei presenti. Subito dopo, il nipote spiega che Frank ha espresso il desiderio che le sue ceneri siano affidate al Biferno. Poi, in un silenzio ormai assoluto e partecipe, legge alcuni versi in inglese.	1, 2, 3	Proper noun, Ecology
3	Cantalupo del Sannio	The village where Frank Monaco's mother was	Invita riservatamente qualcuno dei presenti, senza spiegazioni, a ritrovarsi alla fine della mattinata nei pressi della Taverna, come era	1, 2, (3), 5	Proper noun, (2) material culture (houses and towns)

		born and where he returned in the mind-1950s.	denominata l'antica locanda sul tratturo all'altezza del bivio di Cantalupo del Sannio.		
4	Compaesani	People coming from the same village.	Intanto, a New York arriva anche Giacomo Monaco, il cugino da cui Chiarina è stata allontanata. Giacomo, tramite i compaesani, sa della nuova condizione di Chiarina e riprende contatto con lei. Possono finalmente unirsi e dal nuovo matrimonio nascono altri tre figli: Daniel, Anthony e John.	1, 2, 3, (4), 5, 6	Common expression, (3) social culture
5	Creanza	A form of courtesy to guests and members of a community, usually associated with good or bad manners. This is also often associated with offering food to the guest as it comes from Latin (creare) and Spanish (criar), in which language it acquired the meaning of feeding and nurturing.	Durante il lungo viaggio [le donne] alternavano allegria e amichevoli prese in giro a momenti di preghiera; cantavano tutte insieme Santa Maria Madre di Dio. Poi prendevano i loro cesti, aprivano i grandi tovaglioli colorati e mangiavano il pane e il formaggio che si erano portati da casa. Me ne offrivano con creanza e cordialità. Attraverso l'approccio al femminile, insomma, Monaco intuisce che quel mondo contadino ha valori, competenze ed etica del lavoro, legami di solidarietà, 'creanza', tradizioni, un rapporto tenace e costruttivo con l'ambiente naturale, che, intrecciati tra loro, possono configurare la trama di una 'civiltà' non da ignorare o fare inaridire perché arretrata e lontana da un preteso modello di 'modernità', ma da comprendere e apprezzare nella sua 'alterità' e nella sua capacita di esprimere ancora fermenti di umanità e spirito solidale.	1, 2, 3, (4), 5, 6	Common expression, (4) customs / (5) Gesture and habits
6	La Taverna (9 occurences)	The place where Frank Monaco's mother was	Francesco può così tornare in Italia, con la moglie Cristina e la figlia Bianca, nata in America, e acquistare l'antica taverna sul	1, 2, 3, 4	Proper noun, loaded, (2) Material culture

		born and where the photographer returned in the mid-1950s. It used to be a local restaurant osteria, and it was a restaurant until recently.	tratturo di Cantalupo. La Taverna, oltre a diventare l'abitazione di famiglia, diventa anche la nuova fonte di reddito per i Compagnucci, che nel frattempo si moltiplicano velocemente: nascono altre sei figlie, tra cui Chiarina, la madre di Frank, e infine Vincenzo, il maschio agognato.		
7	Macchiagodena	A village close to Cantalupo del Sannio	Nella mattinata, in una sala del castello di Macchiagodena, si svolge un convegno dedicato alla commemorazione dell'attentato alle Twin Towers, dove aveva perso la vita anche un discendente di emigrati originari del paese, e al ricordo di Monaco, scomparso a Londra poco più di due mesi prima (26 giugno 2007). L'insolito abbinamento si spiega per gli intrecci familiari che i genitori di Frank avevano a Macchiagodena	1, 2, 3, 5	Proper noun, (2) material culture (houses and towns)
8	Merica	A dialectal variation of the country America.	Gli uomini avevano ripreso la strada che da almeno settant'anni i molisani avevano imboccato, assieme a milioni di altri meridionali, e che il fascismo aveva tentato inutilmente di interrompere, riuscendo in realtà solo a sospenderla temporaneamente. Questa volta ad attrarli non era solo la Merica, ma anche le città italiane e, oltre i confini, il Belgio, la Francia, l'Inghilterra e, un po' più tardi, la Svizzera e la Germania occidentale.	1, 2, 3, (4), 5, 6	Proper noun, loaded, (4) concepts, a) political and administrative, / (3) social culture
9	Mericani	A dialectal variation of the adjective of nationality *Americani*, plural of *Americano*. Generally used	«In quei tempi – dirà – i miei parenti, come quasi tutti gli immigrati, non avevano altro pensiero che di diventare prima possibile "mericani"».	1, 2, 3, (4), 5, 6	Proper noun, loaded, (3) social culture + (4) concepts, (a) political and administrative

		to describe those Italian migrants who found a job and went on living in the USA. They tend to have some family relationships within a village or town in Italy and are called as such by those who have stayed. Using the dialectal expression rather than the Italian Americani			
10	Molisano, Molisani	An adjective or a noun referring to people or objects from the regione Molise		1, 2, 3, 4, 6	Proper noun, (4) concepts, (a) political and administrative
11	Opere di civiltà	Public ministry programme for the Mezzogiorno of Italy that aimed at building new infrastructure. Works of Civilization	". Questo passaggio avrebbe certamente comportato un miglioramento delle condizioni di vita e una diffusione delle cosiddette 'opere di civiltà', ma, nello stesso tempo, avrebbe anche eroso un patrimonio di tratti antropologici, di tradizioni e di modelli di socialità.	1, 2, 3, 4, 6	Common expression, (4) concept, a) political and administrative
12	Paesani	Fellow village people, people who live in the same village. Monaco uses the Italian word when writing to his	Ho provato a stare il più vicino possibile ai tuoi paesani, dovunque li incontrassi. Qualche volta sono rimasto ai margini, come straniero. Ci sono stati momenti, invece, in cui la loro terra e le loro tradizioni mi erano familiari, come la tua voce, le tue	1, 2, 3, 4, 5, 6	Common expression, (4) organizations, a) political administrative,/ (3) social culture

		mother. Lombardi reports this in Italian, quoting it from a collection of Monaco's photos. The original text is in English interspersed with some words in Italian.	risate, le tue preghiere. All'inizio in lontananza, poi abbastanza vicino da riuscire a sentire il contatto delle loro oneste mani. Ho provato a mostrare la loro dignità, la loro forza e la loro integrità		
13	Restanza	This theoretical concept was coined by Italian Meridionalist anthropologist Vito Teti. It refers to the	In un'altra immagine iconica, questa volta dell'esodo, ricorre ancora alle donne per sottolineare la persistenza dei legami anche durante e dopo l'abbandono e la forza della "restanza" femminile.	1, 2, 3, 4	Common expression (becoming a proper noun) (4) concepts, a) political and administrative
14	Rio	A stream near Cantalupo del Sannio, it comes from the Latin word rivus directly into the local dialect	Dopo poco, il nipote londinese e alcuni parenti italiani guidano i convenuti verso un ponte su un torrente poco distante, il Rio, affluente del Biferno. Una giovane donna, con un sorriso, offre un fiore a ciascuno dei presenti.	1, 2, 3, 4, 5	Proper noun, (1) Ecology loaded
15	Santa Maria Madre di Dio	The women of Molise described by Frank Monaco sang this religious hymn during a pilgrimage. No English hymn by this name has been retrieved.	Durante il lungo viaggio [le donne] alternavano allegria e amichevoli prese in giro a momenti di preghiera; cantavano tutte insieme Santa Maria Madre di Dio. Poi prendevano i loro cesti, aprivano i grandi tovaglioli colorati e mangiavano il pane e il formaggio che si erano portati da casa.	1, 2, 3, 4, 5, 6	Proper noun, loaded, (4) activities b) religious
16	Santa Maria del Molise	Village in Molise	Chiarina cresce accudendo il piccolo Vincenzo e dando una mano nella Taverna, finché non intreccia un rapporto di simpatia	1, 2, 3, 5	Proper noun, (2) material culture (houses and towns)

			con un cugino, Giacomo Monaco, di Santa Maria del Molise, un paese vicino.		
17	Straniero	Foreigner. Monaco uses the Italian word when writing to his mother. Lombardi reports this in Italian, quoting it from a collection of Monaco's photos. The original text is in English interspersed with some words in Italian.	Ho provato a stare il più vicino possibile ai tuoi paesani, dovunque li incontrassi. Qualche volta sono rimasto ai margini, come straniero. Ci sono stati momenti, invece, in cui la loro terra e le loro tradizioni mi erano familiari, come la tua voce, le tue risate, le tue preghiere.	1, 2, 3, 4, 5, 6	Common expression, (4) concept, a) political and administrative.
18	Tratturo	A track in rural areas used by shepherds to transfer cattle and sheep flocks from a pasture to another one based on climate and pasture conditions. This process is known as *transhumance. Both tratturo and transhumance have been recognized as UNESCO's World Heritage.*	Invita riservatamente qualcuno dei presenti, senza spiegazioni, a ritrovarsi alla fine della mattinata nei pressi della Taverna, come era denominata l'antica locanda sul tratturo all'altezza del bivio di Cantalupo del Sannio.	1, 2, 3, 4, 5, 6	Common expression, (2) Material culture, transport
19	Zio Vincenzo	Uncle Vincenzo in	Tanto più che, nonostante le persistenti difficoltà dei	1, 2, 3, (4), 5, 6	Proper noun,

		Italian. The use of *zio* in Southern Italy also applies to other people who are not properly an individual's uncle, but also to other people who are close to the person or the person's mother/father. It is also used out of respect of an older person on which to rely or lean, in the absence of one's mother or father.	collegamenti postbellici, ogni giorno un pullman parte da Roma e raggiunge Campobasso, passando per Isernia, dove zio Vincenzo Compagnucci, con il suo distributore di benzina, è conosciuto da tutti.		(3) social culture,
20	Zia Vincenza	Aunt Vincenza in Italian. (See above)	Dopo qualche giorno, lo zio li accompagna in macchina a Cantalupo, alla Taverna, tra le braccia di zia Vincenza, autoritaria e vulcanica, che li riceve con rumorosa affabilità.	1, 2, 3, (4), 5, 6	Proper noun, (3) social culture

Appendix C

Table 3. CSIs (proper nouns) – Phase 1 – Fluid and Creative styles 27 January 2025					
	CSIs - Proper nouns	**Lara's Translate Document – Fluid 1**	**Aixelà's CSI manipulation strategy**	**Lara's Translate Document – Creative 1**	**Aixelà's CSI manipulation strategy**
1	Arma dei Carabinieri	Carabinieri Corps	Linguistic (non cultural) translation	Carabinieri Corps	Linguistic (non cultural) translation
2	Biferno	Biferno	Repetition	Biferno	Repetition
3	Cantalupo del Sannio	Cantalupo del Sannio / Cantalupo	Repetition	Cantalupo del Sannio	Repetition
4	Corpo dei carabinieri	Carabinieri Corps	Linguistic (non cultural) translation	Carabinieri Corps	Linguistic (non cultural) translation
5	La Taverna	The Tavern, the Taverna, La Taverna	Linguistic (non cultural) translation, repetition	The Taverna, the tavern	Linguistic (non cultural) translation
6	Macchiagodena	Macchiagodena	Repetition	Macchiagodena	Repetition
7	Merica	America	Absolute universalization	America	Absolute universalization
8	Mericani	Americans	Absolute universalization	to blend seamlessly into the American fabric	Deletion + autonomous creation
9	Molisani	Molisans, Molise people	Autonomous creation,	Molisans, Molise people	Autonomous creation
10	Molisano	Molise	Linguistic (non cultural) translation	Molise	Linguistic (non cultural) translation
11	Rio	Rio	Repetition	Rio	Repetition
12	Santa Maria del Molise	Santa Maria del Molise	Repetition	Santa Maria del Molise	Repetition
13	Santa Maria Madre di Dio	Saint Mary, Mother of God	Linguistic (non cultural) translation	Santa Maria, Mother of God	Repetition + Linguistic (non cultural) translation
14	zia Vincenza	Aunt Vincenza	Linguistic (non cultural) translation + repetition	Aunt Vincenza	Linguistic (non cultural) translation + repetition

15	Zio Vincenzo,	Uncle Vincenzo	Linguistic (non cultural) translation + repetition	Uncle Vincenzo	Linguistic (non cultural) translation + repetition

Appendix D

Table 4. CSIs (common expressions) – Phase 1 - Fluid and Creative styles – 27 January 2025

	CSIs Common expressions	**Lara's Translate Document – Fluid 1**	**Aixelà's CSI manipulation strategy**	**Lara's Translate Document – Creative 1**	**Aixelà's CSI manipulation strategy**
1	Compaesani	fellow countrymen	Limited universalization	fellow countrymen	Limited universalization
2	Creanza	Trust	Autonomous creation (translation error)	They extended a warm and trusting invitation, offering me a taste of their hospitality.	Autonomous creation (compensation)
3	Creanza	a sense of belonging	Autonomous creation (translation error)	a sense of belonging	Autonomous creation (translation error)
4	Opere di civiltà	'works of civilization'.	Linguistic non cultural translation	'works of civilization'.	Linguistic non cultural translation
5	Paesani	fellow countrymen	Limited universalization	fellow countrymen	Limited universalization
6	Restanza	'resistance'	Absolute universalization	'resistance'	Absolute universalization
7	Straniero	Outsider	Synonym	a stranger in a foreign land	Autonomous creation
8	Tratturo	Sheep track	Absolute universalization	Sheep track	Absolute universalization

Appendix E

Table 5. CSIs (proper nouns) – Phase 3 - Fluid and Creative styles – 12 September 2025					
	CSIs - Proper nouns	**Lara's Translate Document – Fluid 2**	**Strategy**	**Lara's Translate Document – Creative 2**	**Strategy**
1	Arma dei Carabinieri	Carabinieri	Deletion + partial Repetition	Carabinieri	Deletion + partial Repetition
2	Biferno	Biferno	Repetition	Biferno	Repetition
3	Cantalupo del Sannio	Cantalupo del Sannio	Repetition	Cantalupo del Sannio	Repetition
4	Corpo dei carabinieri	Carabinieri Corps	Linguistic (non cultural) translation	Carabinieri	Deletion + partial Repetition
5	La Taverna	The Tavern / La Taverna	Linguistic (non cultural) translation, repetition	The Tavern / the Taverna	Linguistic (non cultural) translation + partial repetition
6	Macchiagodena	Macchiagodena	Repetition	Macchiagodena	Repetition
7	Merica	America	Absolute universalization	America	Absolute universalization
8	Mericani	American	Absolute universalization	American	Absolute universalization
9	Molisani	Molisans	Autonomous creation,	Molisans,	Autonomous creation,
10	Molisana	Molise women	Linguistic (non cultural) translation	Molise women	Linguistic (non cultural) translation
11	Rio	Rio	Repetition	Rio	Repetition
12	Santa Maria del Molise	Santa Maria del Molise	Repetition	Santa Maria del Molise	Repetition
13	Santa Maria Madre di Dio	Holy Mary Mother of God	Limited universalization	"Holy Mary, Mother of God."	Limited universalization
14	zia Vincenza	Aunt Vincenza	Linguistic (non cultural) translation + repetition	Aunt Vincenza	Linguistic (non cultural) translation + repetition
15	Zio Vincenzo,	Uncle Vincenzo	Linguistic (non cultural) translation + repetition	Uncle Vincenzo	Linguistic (non cultural) translation + repetition

Appendix F

Table 6. CSI (Common expressions) – Phase 3 - Fluid and Creative styles – 12 September 2025					
	CSIs Common expressions	**Lara's Translate Document – Fluid 2**	**Strategy**	**Lara's Translate Document – Creative 2**	**Strategy**
1	Compaesani	fellow countrymen	Limited universalization	fellow townspeople	Synonymy
2	Creanza	Courtesy	Synonymy	They graciously and warmly extended their offerings to me.	Compensation
3	Creanza	Creanza	Repetition	--	Deletion
4	Opere di civiltà	works of civilization	Linguistic (non cultural) translation	The blossoming of civilization	Deletion+autonomous creation
5	Paesani	Your fellow villagers	Synonymy	your villagers	Synonymy
6	Restanza	female "restanza" (resistance).	Repetition + extratextual gloss	female resilience	Absolute universalization
7	Straniero	A stranger	Synonymy	a quiet observer, a stranger in a new world	Synonymy + Autonomous creation
8	Tratturo	Sheep track	Absolute universalization	Sheep track	Absolute universalization

Paulínyová, Lucia. 2026. Šveda, Djovčoš, Perez (eds). 2026. Confronting Digital Dilemmas in Translator and Interpreter Training. In: L10N Journal 1(5), pp. 100-102.

Šveda, Pavol, Djovčoš, Martin, and Perez, Emília (eds.). 2026. Confronting Digital Dilemmas in Translator and Interpreter Training

Reviewed by Lucia Paulínyová

Comenius University in Bratislava

lucia.paulinyova@uniba.sk

Confronting Digital Dilemmas in Translator and Interpreter Training, edited by Pavol Šveda, Martin Djovčoš and Emília Perez, addresses a question that has become increasingly difficult for translator and interpreter educators to avoid: how should training respond to rapid technological change? The volume brings together contributions that approach this question from different angles, including technology, curriculum design, professional competences, ethics and accessibility. Rather than treating digitalization simply as either a threat to the profession or a source of inevitable progress, the contributors consider what technology means for the skills, practices and forms of education that remain distinctly human.

The book is organized into three sections, dealing with theoretical foundations and pedagogical innovation, ethical and social issues, and ways of adapting translator and interpreter education to technological change.

The opening discussion questions the assumption that technological development will necessarily reduce the role of human translators and interpreters. This concern with the continuing role of human expertise recurs throughout the volume. Anthony Pym, for example, takes a critical approach to claims about translation technologies and argues that their effects should be discussed on the basis of evidence rather than speculation. His contribution provides a useful starting point for the volume as a whole, particularly because it shifts attention away from predictions about technology and toward the evidence needed to assess them.

David Orrego-Carmona approaches the issue from another direction by examining digital divides. His contribution draws attention to unequal access to technological

resources and to the fact that digitalization cannot be reduced to questions of efficiency or productivity. As he aptly observes, "Understanding translation as a fundamentally human activity that requires empathy, cultural sensitivity, and deep engagement with both source and target contexts becomes crucial as automation continues shaping the industry." This observation captures one of the volume's central messages: technological competence remains important, but it does not replace the human and cultural dimensions of translation.

Karolina Levanaitė considers translator competence as something that needs to develop continuously rather than as a fixed collection of skills. Translator education, in her view, must prepare students not only for the demands of the present profession but also for changes that cannot yet be clearly defined. Her formulation that there is "no single recipe" for educating translators, and that this ultimately means "educating them for the unknown," captures an important difficulty facing programs that have to prepare students for an increasingly unpredictable professional environment.

The chapters in Part II focus on curriculum design and professional preparation. Paola Gentile's contribution is particularly valuable because it draws on graduates' perspectives and links them to developments in the profession. The responses point to an interest in broader interdisciplinary preparation, with law, economics and marketing mentioned alongside traditional language training. The contribution also draws attention to post-editing, remote work and technological competence, all of which have implications for the way curricula are structured.

Jonathan Downie addresses another increasingly prominent development: machine interpreting and AI-based language technologies. His discussion does not frame these developments simply in terms of replacement. Instead, it considers which human skills continue to matter and how technological change might alter rather than eliminate professional opportunities for interpreters.

João Brogueira's chapter on AI literacy is particularly valuable for its practical focus. He convincingly argues that understanding AI tools is becoming an essential part of a translator's professional profile and should therefore be integrated into translator training. One of the chapter's strengths is that it goes beyond technical skills and considers how AI literacy can help translators adapt to change and feel more confident in their work.

This discussion is complemented by Ivo Poláček and Lucia Tonková's chapter, which presents a case study of CAT tools in Slovak translator education. The authors show how such technologies can be integrated into training while developing skills directly relevant to professional practice. Together, these chapters provide some of the most practical insights in the volume for educators seeking to update their curricula.

The final part of the book turns to the social and pedagogical implications of digital transformation. Martin Djovčoš proposes a sociocentric approach to translator and

interpreter education that places student engagement, networking and links with the professional community at the center of training. His examples, including discussion series and podcasts, show how activities extending beyond conventional classroom teaching can give students opportunities to gain practical experience and establish professional connections. The chapter clearly shows the value of connecting university education more closely with professional practice.

Emília Perez and Alina Secară highlight the growing importance of media accessibility and show that some forms of accessibility, particularly audio description, remain difficult to automate. Their discussion is especially convincing in demonstrating why audio description still requires strong linguistic and creative skills. The chapter also offers a more optimistic perspective on the future of the profession by identifying areas in which human creativity continues to play a central role.

Barbara Heinisch provides a different example of digital innovation through the ReTrans project. The project combines podcasts, videocasts, quizzes, and interactive stories that introduce students to the challenges of interpreting in refugee transit zones. The interactive stories are of particular interest because they place students in realistic situations and require them to respond to problems rather than simply read about them. The project demonstrates how digital tools can support more active forms of learning.

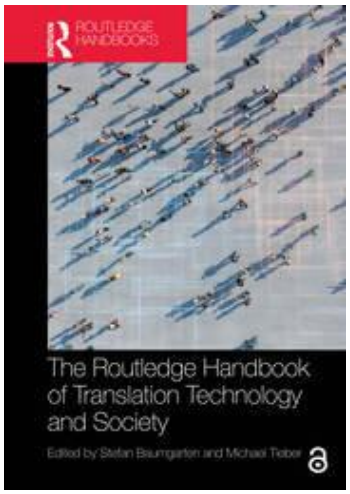
Pavol Šveda examines remote interpreter training by comparing students' experiences of on-site and remote interpreting through the i-nest and RSTIP platforms. The results suggest that students generally found on-site interpreting slightly easier, while remote sessions were associated with lower levels of stress. The study is limited by its small sample, but the findings nevertheless support the idea that remote and on-site formats need not be viewed as mutually exclusive. A combination of the two may offer a more flexible approach to interpreter training.

The closing contribution by Judith Platter, Katia Iacono, and Maria Bernadette Zwischenberger extends the discussion to cooperation between spoken-language and speech-to-text interpreters. Using mock conferences, the authors introduce students to different interpreting profiles and professional practices. The emphasis on cooperation between different fields works well as the final contribution to the volume.

Taken as a whole, *Confronting Digital Dilemmas in Translator and Interpreter Training* is a significant contribution to contemporary Translation and Interpreting Studies. The volume highlights the need to find a balance between technological development and the human skills and judgment that remain central to the profession. The volume captures an important moment in the development of the field and offers a balanced, thoughtful and forward-looking perspective on the future of translator and interpreter education.

Final Variable

The Final Variable section highlights noteworthy publications in translation and interpreting studies and related disciplines. Reflecting the theme of the Translation, Interpreting & Culture 2025 conference, *Translators, Interpreters, and Society,* this issue presents four works that explore translation and interpreting as socially situated practices. Together, they examine how technological infrastructures, public discourses, and institutional arrangements shape language professionals' work, multilingual access, and relations of power.



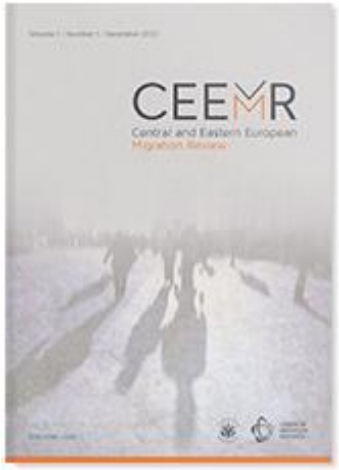
The Routledge Handbook of Translation Technology and Society (2025), edited by Stefan Baumgarten and Michael Tieber, examines translation technology in relation to contemporary social, political, and economic structures. Its chapters address the ethical, ideological, socioeconomic, and environmental implications of AI-driven translation, everyday machine translation, crowdsourcing, and platform-based work. They consider linguistic inequality, data colonialism, the commodification of language, labor conditions, and changing relations between human and machine agency. By situating translation technologies within the broader structures of digital capitalism, the handbook offers a critical perspective on their effects on professional practice and multilingual communication.

https://doi.org/10.4324/9781003271314



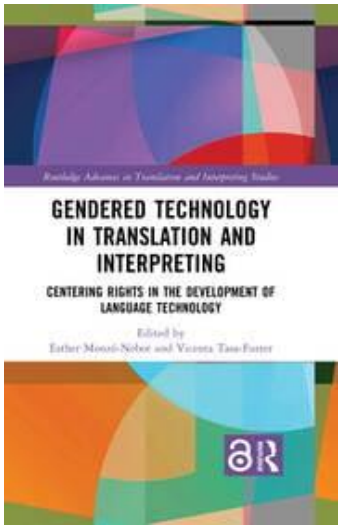
Concepts of Translators and Translation in Online Social Media: Construal and Contestation (2025) by Haidee Kotze investigates how translators and translation are conceptualized in online social media. Using a corpus-assisted critical discourse analysis of 2,639 English-language tweets concerning the 2021 debate over the selection of translators for Amanda Gorman's poetry, the article examines how these representations are constructed and contested. It also considers how links to traditional news media circulate within these debates. Kotze shows that references to translation can serve as discursive resources through which users express ideological positions and construct online identities, rather than merely commenting on translation practice itself.

https://doi.org/10.1080/14781700.2023.2282581



Who Accommodates Meaning? The Privileges of Being Interpreted and Housed from the Perspective of Ukrainians in Slovakia (2026) by Martina Wilsch, Miroslava Hlinčíková, Helena Tužinská, and Markéta Štefková examines how interpreting and institutional communication shape the experiences of people displaced from Ukraine to Slovakia. Based on qualitative research with displaced persons, ad hoc interpreters, and institutional actors, the study analyzes epistemic injustice in the contexts of housing and healthcare. The authors show how gaps in professional interpreting provision, institutional monolingualism, and inaccessible or uncertain information can marginalize refugees' knowledge and constrain their ability to understand entitlements, communicate needs, and act on their own behalf. The study thus situates language mediation within broader questions of refugee agency, equitable service provision, and social inclusion.

https://doi.org/10.54667/ceemr.2026.14



Gendered Technology in Translation and Interpreting: Centering Rights in the Development of Language Technology (2024), edited by Esther Monzó-Nebot and Vicenta Tasa-Fuster, brings translation and interpreting studies into dialogue with law and gender studies to examine how language technologies reproduce or challenge inequalities. The chapters consider machine translation, large language models, remote interpreting, and their effects on language professionals and communities. Particular attention is given to gender bias, legal regulation, and the protection of individual and collective rights, including concerns affecting marginalized groups and speakers of less-resourced languages. The volume advocates a more inclusive, rights-centered approach to developing and using language technologies.

https://doi.org/10.4324/9781003465508