# Pilot Study on Medical Translations in Lay Language: Post-Editing by Language Specialists, Domain Specialists or Both?

Ayla Rigouts Terryn[1], Lieve Macken[1], Els Lefever[1], Robert Vander Stichele[1], Koen Vanneste[2], and Joost Buysschaert[1]

[1]Language and Translation Technology Team, Ghent University, Belgium
[2]iscientia, Belgium
name.surname@ugent.be, ivs@iscientia

**Abstract**

Despite the rich history of research into medical translation, there is a notable lack of empirical studies on the best workflow for this task, especially in a modern translation setting involving post-editing of machine translation. This pilot study was conducted in preparation for a large translation project of medical guidelines for laypeople from Dutch into French. It is meant to shed light on how medical post-editing is best handled. How do medical specialists (doctors) versus language specialists (translators) perform on this task? How can their respective strengths lead to the highest quality translation? To gain more insight into these questions, errors in the machine translation output of medical guidelines were annotated and labelled. Based on these annotations, the product of doctors' and translators' post-editing could be analysed and classified into necessary changes (mistakes that were correctly solved), under-revisions (mistakes that were not corrected during post-editing), over-revisions (new errors introduced during post-editing) and hyper-revisions (preferential changes made by the post-editor). The results of this small-scale research illustrate the complexity of the task and reveal some surprising findings (e.g., doctors sometimes struggle with domain-specific terminology, and translators appear to be less efficient because they introduce many hyper-revisions).

## 1 Introduction

Medical translation has long been a subject of discussion in translation research. Nevertheless, whether medical translation is best performed by medical practitioners or by translators is a question that has received only limited attention, with a notable lack of empirical research (Jensen and Zethsen, 2012). Even less is known about this task in a modern translation setting, with easily accessible domain-specific information online and machine translation (MT).

In this paper, we present a pilot study that investigates the optimal strategy for the translation (post-editing) of medical texts intended for lay people. It is meant to shed light on how medical post-editing is best handled. How do medical specialists (doctors) versus language specialists (translators) perform on this task? How can their respective strengths lead to the highest quality translation? The research was conducted for, and in close cooperation with, the company iscientia (https://www.iscientia.com), to help them reach an informed decision prior to accepting a large translation project involving the Dutch into French translation of patient-oriented medical guidelines.

## 2 Background and Aim

There is an increasing trend to tailor health communication so as to make it patient-centred, which should contribute to patient empowerment (Holmstrôm and Rôing, 2010; Vâsquez *et al.*, 2017); other researchers, however, have also voiced criticism of the trend (Cornwall and Brock, 2005; Hansen and Zethsen, 2018; Salmon and Hall, 2004). If patients are to profit from access to medical information, a first prerequisite is understandable lay language. This will require the challenging task of intralingual translation between more technical originals and lay versions (Zethsen, 2018).

The Dutch-language medical guidelines that constituted the source texts in the present experiment (https://www.gezondheidenwetenschap.be/richtlijnen) are an example of such an

intralingual translation effort. They are patient-oriented medical guidelines based on technical ones that can be consulted in French and Dutch on the Belgian ebpracticenet website (https://www.ebpnet.be) and that are themselves in large part translations of the DUODECIM evidence-based medical guidelines (https://www.duodecim.fi/english/products/ebmg). A plan to have the Dutch-language patient-oriented guidelines translated into French raised the question whether machine translation could be used to speed up the process. As post-editing of the raw results would obviously be needed, a further question was whether this task had better be entrusted to translators or whether French-speaking doctors with a good passive knowledge of Dutch were better placed to do the job.

Research in this area is fairly limited. Jensen and Zethsen (2012) investigated medical translation by either pharmacists or translators, focusing on the suitability of the final product for laypersons by analysing nominalisations and neo-classical terms. In this study, the differences between the two groups were surprisingly small. However, the overall quality of the translations is not discussed. Older research by O'Neill (1998) discusses how "medically knowledgeable linguists" versus "linguistically knowledgeable medical professionals" tackle medical translation, concluding that both types of professionals are capable, but that a combination of both is ideal. The discussion focuses on the respective strengths and weaknesses of translators and medical professionals: the former may have trouble with the specialised medical terminology, while the latter may have weak writing skills. In other research (Davies, 1998), it was found that medical professionals focus most on terminology and assign less importance to syntax and grammar. However, the context of medical translation has changed significantly since 1998. It is now much easier to find up-to-date and high-quality terminology online (for those who know how to look for it) and neural machine translation quality (for some language pairs) has vastly improved. Therefore, both translators and medical professionals now have tools available to help them overcome their respective disadvantages.

In an attempt to make an informed decision about the best workflow for the translation of the medical guidelines for iscientia, this pilot study is meant to provide some first insights in how the task is handled by translators versus medical professionals and to see if their skills are complementary. Therefore, the provisional decision was made to go for a double post-editing + revision process involving translators as well as doctors. Moreover, the pilot study was considered a testing ground for practical guidelines, so that any potential errors or difficulties could be identified before tackling the larger assignment. Considering the limited size of this pilot study, and taking into account findings from previous research which emphasise that translation (and post-editing) strategies are highly personal (Daems and Macken, 2019), the results need to be carefully interpreted to avoid overgeneralisation. Nonetheless, the current study reveals some interesting trends and yields inspiration for further research.

## 3 Experimental Setup

For this pilot study four Dutch source documents were selected of varying length (569-1849 source words). The experiment involved two translators (T1 and T2) and two French-speaking doctors with knowledge of Dutch (D1 and D2). T1 is a native French speaker who was raised bilingually with Dutch as a second language. She has no degree in translation but several years of translation experience, though none in the medical domain. T2 is also a native French speaker who has a master's degree in translation with a specialisation in medical and scientific translation. She has almost 20 years' experience as a translator, of which over 10 years in the medical domain. However, she does not usually translate from Dutch. Neither translator had any previous experience post-editing. D1 is a French-speaking family doctor with a fair knowledge of Dutch, acquired at school. She had no previous experience with translation or post-editing. Her contact with other languages is limited to reading English medical texts and

occasional consultations in English or Dutch. D2 is a young family doctor in the Brussels area who is at ease with French, Dutch and English; she had no prior experience with translation or post-editing.

Google Translate (neural) was used for the initial translation. A two-step approach was adopted, with the translators and the doctors first post-editing the raw MT output and subsequently revising the target texts as post-edited by their colleagues. The translators performed post-editing within their familiar CAT-tool **(SDL** Trados Studio) using a translation memory and a termbase offering suggestions for patient-friendly terms. The doctors performed post-editing in a bilingual Word-document and had access to the termbase in a simple excel format. Revision was performed in bilingual Word by translators and doctors alike. The tasks were distributed as displayed in Table 1 over the two workflows: Workflow I (TD) consists of post-editing by a translator and revision by a doctor and Workflow II (DT) is the opposite, i.e. post-editing by a doctor and revision by a translator. Thus, all participants (T1-2 & D1-2) looked at all four texts (A to D), allowing comparisons of (1) the two workflows for the same text, (2) the two workflows for the same participants, (3) the product of post-editing/revision of the same text by the same participants.

Translators and doctors were given exactly the same instructions. They were told that their work had to result in a French version that was ready for publication. The instructions also said that correction work was to focus on content as well as on form. They were instructed to take patient-orientedness into account, so that the target text, like the source text, would be understandable to laypeople. Both parties were provided with the same list of typical content-related mistakes to look out for, as well as a list of possible formal mistakes.

All machine translation output was also independently annotated by the authors, assisted by a Dutch-speaking medical doctor with a good knowledge of French. A master translationwas made in which all annotated errors were solved with the least possible amount of edits.The machine translation error taxonomy as described by Tezcan, Hoste and Macken (2017) was used to divide errors into 11 different categories (see Table 2) to facilitate a fair and detailed evaluation. It was expected at the outset that the doctors would be better at solving content-related issues, while the translators were likely to have better expertise in correcting language and formatting errors. In addition to this error annotation, the amount of editing that took place was measured by means of Translation Edit Rate (TER), which takes into account all insertions, deletions, and substitutions of single words as well as shifts of word sequences. Combining all the information (error codes, master translation and TER scores), allowed us to identify necessary changes (annotated errors that were solved), underrevisions (annotated errors that were not solved), overrevisions (new errors that were introduced), and hyperrevisions (all edits that were not deemed necessary to solve annotated errors) (Rigouts Terryn et al., 2017). Any potential overrevisions made during the post-editing phase were counted as errors to be solved in the revision phase.

| | **Workflow I (TD)** | | **Workflow II (DT)** | |
|---|---|---|---|---|
| **text** | post-editing | revision | post-editing | revision |
| A | T1 | D2 | D1 | T2 |
| **B** | T2 | D1 | D | T1 |
| **C** | T1 | D1 | 2 | T2 |
| **D** | T2 | D2 | D | T1 |

Table 1: Overview of methodology for four texts (A-D) with Workflow I (post-editing by translator, revision by doctor) and Workflow II (post-editing by doctor, revision by translator)

| Style & Fluency Errors | | Mistranslation Errors | |
|---|---|---|---|
| ADD | additions | MTP | partial translation |
| DEL | deletions | MTT | terminological error |
| INC | inconsistencies | MTC | |
| GRM | grammatical errors | MTW | translation contradicts source text |
| | | MTG | word sense error |

Table 2: Error codes; see also (Tezcan *et al.*, 2017)

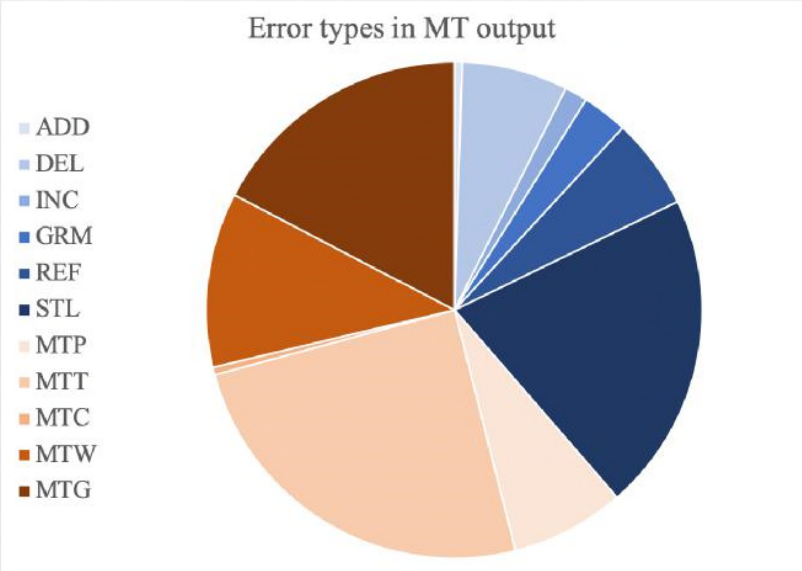


Figure 1: Pie chart of all error types in the MT output; mistranslations have a reddish-brown shade, style & fluency errors are a shade of blue

## 4 Analysis

### 4.1 Quality of MT

Overall, the quality of the MT output was satisfactory, though it varied per text. There were many segments in which no errors were found at all: 44%, 44%, 35% and 26% of all segments in texts A, B, C, and D respectively. As mentioned, a master translation was made in which all of these errors were corrected with the least possible amount of editing. The TER score of the master translations, compared to the original MT output, was similar for all texts, ranging from 11.846 (A) to 15.738 (D). In total, 202 errors were annotated in the four texts and these errors range over all error types. As can be seen in the pie chart (see Figure 1), there are more mistranslation errors (61%) than style & fluency errors (39%). Looking at the individual error types, mistranslations related to terminology (MTT), style errors (STL) and general mistranslations (MTG) are most common; this is the case for all four texts. Two error types appear only once in the entire MT output: additions (ADD) and mistranslations that are contradictions (MTC). In conclusion, the MT output is of decent quality and, in most cases, the translations are understandable, even when they are not 100% correct.

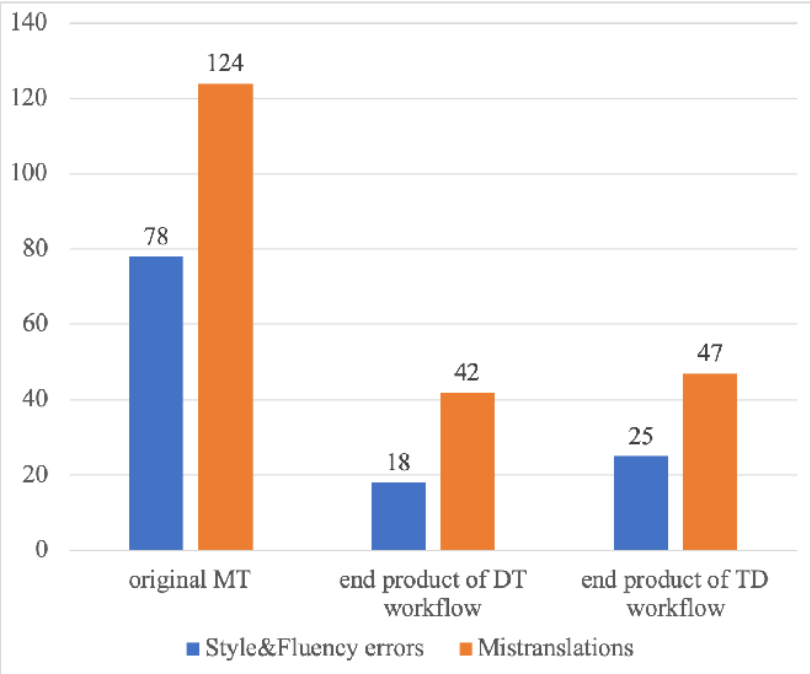


Figure 2: Number of errors in the original MT output, in the end product of the DT workflow, and in the end product of TD workflow

### 4.2 Comparing Procedures

When comparing the two workflows (DT versus TD), the first thing that stands out, is that the end product is very comparable. As shown in figure 2, the number of errors is slightly lower in the DT workflow, i.e. when doctors perform post-editing first, followed by a revision step by translators. In the end product of the DT workflow, 60 of the 202 original errors remain, versus 72 in the TD workflow. The distribution between the two groups of error types remains more or less stable. In the original MT output, 39% of all errors are style & fluency errors, versus 30% in the end product of the DT workflow and 35% in the TD workflow. Differences between individual error types are marginal, with no especially noteworthy exceptions.

**TER** scores do reveal more differences between the two workflows, with higher edit rates for the TD workflow than for the DT workflow **(TER** scores of 155 and 127 respectively). This means that the DT workflow not only resulted in the translation with the least errors, but was also more efficient (with fewer edits), than the TD workflow. Another expected, yet still interesting, finding when comparing **TER** scores for the different workflows, is that most of the edits occur in the first, post-editing, step and fewer edits are made during the second, revision, step. As shown in Figure 3, all participants have higher **TER** scores when post-editing than when revising. The combined total **TER** score of the post-editing phase is 218, versus 118 in the revision step. This difference is easily explained by the fact that the original MT contains many more errors than the already post-edited texts, and so, requires fewer edits. Another explanation could be, that participants are more likely to change a machine-generated text, than a text which has already been edited by another person.

The conclusions from the **TER** scores are also reflected in the number of annotated errors. Figure 4 shows how many errors remain in the MT, after post-editing, and after revision. The green lines show the total number of errors, the orange lines are only the mistranslation errors, and the blue lines are the style & fluency errors. Solid lines represent the TD workflow, while

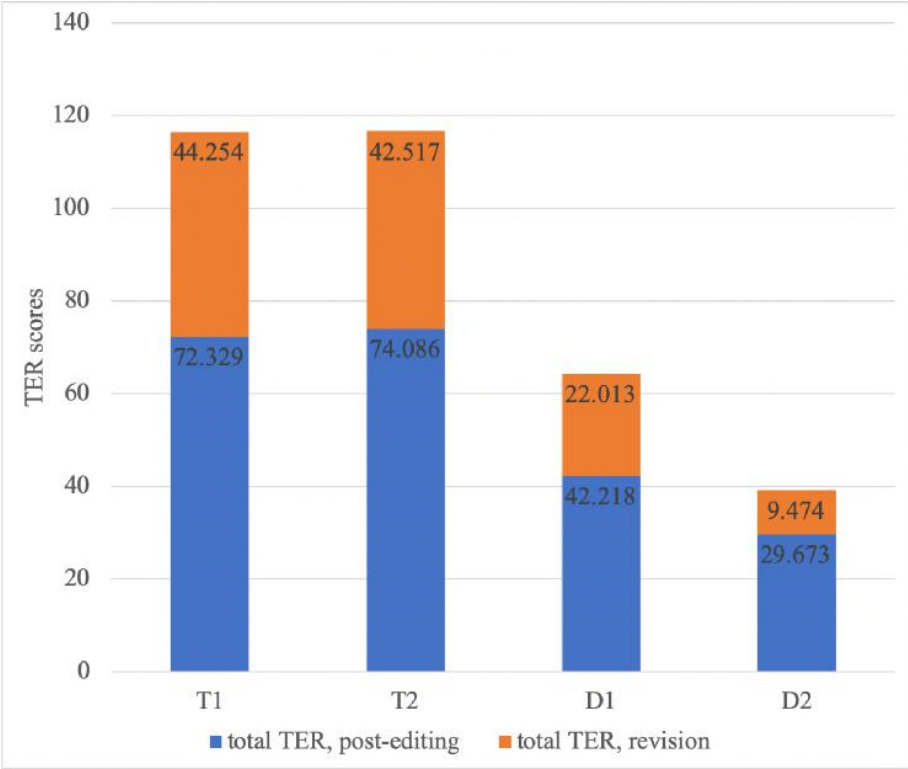


Figure 3: TER scores per person and per phase (post-editing or revision)

interrupted lines indicate the DT workflow. In the first step (MT), both workflows start off identical, with the same input. After post-editing, the total number of errors is still almost identical, with 112 remaining errors in the DT workflow and 13 remaining errors in the TD workflow. However, there is a slight difference in the types of errors: the doctors solved more mistranslation errors, the translators more style & fluency errors. In the final step, revision, fewer errors are solved. In the DT workflow, translators reduced the number of errors by 30, and in the TD workflow, doctors only brought down the errors by 17. This shows how the second step has much less impact, and how the main difference between the two workflows only occurs during this final step. Perhaps doctors are overly trusting of the first corrections of the translators, and too hesitant to make additional changes. Looking back at the TER scores in Figure 3 in more detail, there is a slight difference between doctors and translators that seems to support this theory, namely that, proportionally, doctors make fewer edits in the post-editing phase than translators. 36%-38% of the translators' TER scores originates during the post-editing step, compared to only 24%-34% of the doctors' TER scores.

While the error types already provide some idea of the kinds of errors that can be expected, they do not suffice when trying to measure the severity of the errors, which is why it is important to trace some of the most severe errors individually. The MT output only contains a few errors that make a translation completely unintelligible, such as translating "nails" with the French equivalent for the tool ("clou"), rather than the body part ("ongle"), or saying a swelling will increase instead of decrease. Such errors were mostly caught in both workflows. However, it was found that other serious errors in the final translations, errors which were very obvious or completely changed the meaning of the text, still remained after both workflows. For instance, in the final product of the DT workflow, a

translator changes the doctor's post-edited translation "que pouvez-vous faire vous-même" into "que pouvez-vous faire vous faire", incorrectly repeating two of the words.

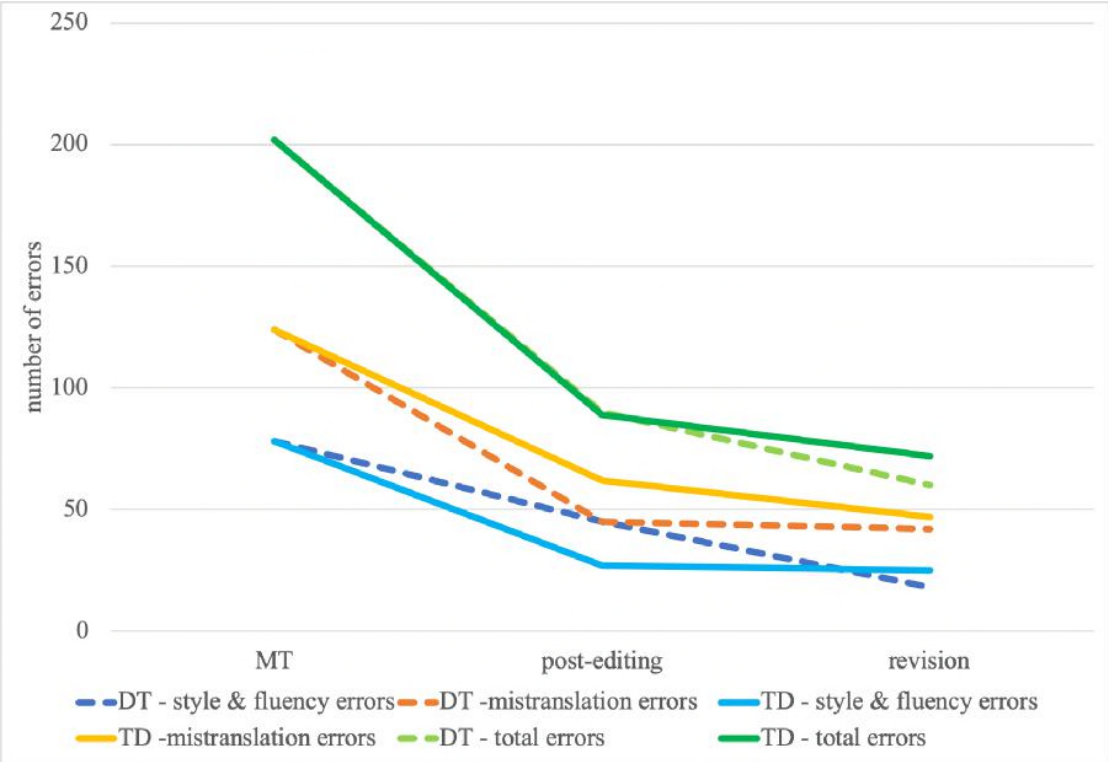


Figure 4: Number of errors after each step (MT, post-editing and revision), split per workflow and per error type

In the TD workflow, a serious grammatical error was introduced when the doctor wrote "la cuvette dans lequel", which should be "la cuvette dans laquelle". Also in the TD workflow, the source text mentions that pain will occur gradually during exercise and get worse faster in case of sudden exercise. The translation states that pain occurs gradually during exercise and gets worse as the exercise continues, but it does not mention the "sudden exercise" that is crucial in this case. This mistake is introduced by the translator and not corrected by the doctor. In another, admittedly more debatable example, the source text advises the use of nail polish in the case of *cosmetic irritation*, presumably referring to nails affected by discolouring or other aesthetically unattractive damage. While the formulation is somewhat doubtful, there is no actual error and the MT solution "irritation esthétique" is fairly acceptable. In the TD workflow, the MT output is not changed. In the DT workflow, the doctor changes it to recommending nail polish in the case of irritation by cosmetics, which makes little sense, as nail polish is itself a type of cosmetic. This is not corrected by the translator.

In conclusion, the DT workflow resulted in fewer errors in the final product, while at the same time being more efficient with a lower total TER score. However, the difference is not very pronounced, especially considering the substantial interpersonal differences, and the fact that the end products still contain serious errors in both workflows. The fact that new errors (over-revisions) are regularly introduced while post-editing and revising, may be a consequence of the participants' inexperience working with an existing translation, rather than translating from scratch.

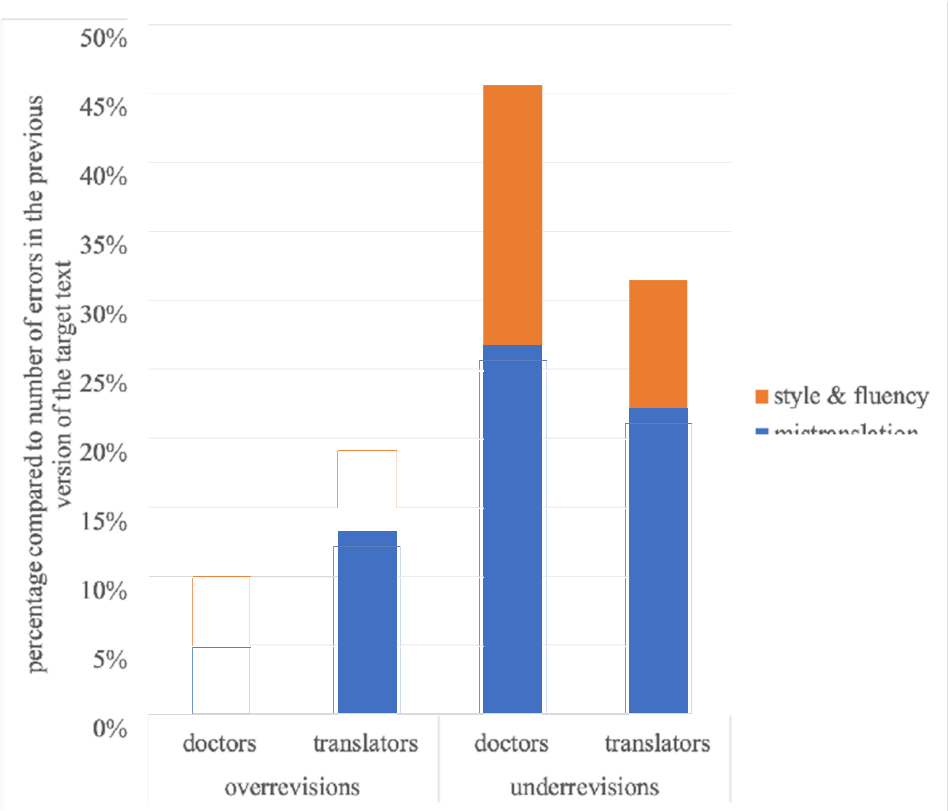


Figure 5: Percentage of over- and under-revisions, compared to number of errors the target text contained in the previous step

### 4.3 Comparing Doctors and Translators

Aside from the difference between workflows, the interpersonal differences, especially those between doctors and translators, were considered as well. The most obvious difference, which could already be seen in Figure 3, is that translators make many more edits than doctors. This is reflected in the combined total **TER** scores of both groups, namely 103 for the doctors, versus 233 for the translators. Even though the difference in **TER** scores is already striking, these scores do not even reflect the edit behaviour of translators strictly enough yet. The 'track changes' function showed that translators often completely delete the MT output and start to translate from scratch, even when their own translation then ends up being identical to the original MT. This would not show up in the **TER** scores, but it is nonetheless an important indication of the translators' inefficiency. It illustrates the translators' inexperience with post-editing, especially since, towards the end of texts, they are less likely to simply delete and rewrite the MT, as if, by then, the translators started to notice that the MT was not that bad. Both translators were very similar in this respect, as shown also by their almost identical **TER** scores in Figure 3. In the same Figure, it can be seen that the differences between the doctors are a bit more pronounced (D2 edits less).

A ranking based on error types shows inconclusive results for the question whether translators or doctors perform best. T2 performs best, followed by D1, T1, and finally D2. T2 solved 72% of the errors, D2 only 47%. However, T2 did make more over-revisions than D2. T1 solved 61% of all errors (similar to D1), but made 38 over-revisions, so T1 just made a lot of revisions, which were sometimes correct, but also often wrong.

Figure 5 shows how many over- and under-revisions were made by translators and doctors, relative to the number of errors present in the previous version of the text. Regarding error type groups, the translators fixed more style & fluency errors than the doctors. Out of all errors fixed by the translators, 47% (T2) and 51%(T1) are style & fluency errors, versus only 26% (D1) and 39% (D2) for the doctors. This might be explained by the fact that translators are more focused on the language and the text as a whole, and doctors might spot content-errors more easily, though further research is required to test this hypothesis. A similar pattern can be seen in the over-revisions: translators make relatively more mistranslation errors, doctors relatively more style & fluency errors. Overall, it appears that translators are more likely to edit a text than doctors. This results in more necessary changes and fewer remaining errors but, at the same time, leads to more over-revisions. Doctors are more likely to under-revise and have fewer over-revisions.

One of the hypotheses about the difference between doctors and translators for medical post-editing, was that the doctors' medical expertise would help them spot mistranslations. This seems already to be partially confirmed by the fact that they solve relatively more mistranslations than translators. It is also reflected in the numbers for the individual error codes, e.g. doctors solve more terminology (MTT) errors than translators (50% vs. 36%). For instance, one of the source texts contained an error, using "mond- en klauwzeer" (hoof-and-mouth-disease) as a synonym for "hand-voet-mondziekte" (hand, foot, and mouth disease), even though the former is only used for animals and the latter for humans. The MT, of course, translated the error without correcting it. In the DT workflow, the doctor spotted the error and removed the wrong synonym in the post-editing phase, but, in the revision phase, the translator re-introduced the error. This is a powerful example of how their medical expertise grants doctors an advantage over translators. It may even be a warning against blindly choosing the DT workflow, even though it resulted in fewest errors. However, there is also a surprising example of how even doctors do not always have sufficient expertise. It concerns the translation of "(...) 69 chromosomen in plaats van de normale 46 chromosomen", which translates into English as "(...) 69 chromosomes instead of the usual 46 chromosomes". In the post-editing phase, T1 makes the serious error of reversing digits, talking about 96 chromosomes instead of 69. D1 revised this translation and was expected to easily catch this error. While D1 did indeed spot the mistake, she changed it to 92 instead of the original 69. She mentioned this in her communication, stating that she did not know where the number 69 in the original came from and that she assumed 92 was meant, since that made sense as 46+46. However, 69 does make sense as 46+(46/2) and a quick online search proved that 69 chromosomes is indeed possible and what is meant in this context. This goes to show that, with a domain as large and specialised as medicine, even supposed domain-experts will not have knowledge of all aspects of the domain and cannot rely on their expertise alone, but will need appropriate research skills.

### 4.4 Patient-friendly Translations

While patient-friendliness of the translations was part of this pilot study, it was not the main focus of the research, in part because it is extremely difficult to measure objectively. Currently, the analysis of this aspect is only anecdotal, with examples of both translators and doctors using more specialised terminology in the target text than had been used in the source text. For instance, "fietsproef" (bicycle test) in the context of coronary artery disease, is originally translated as "test de cyclisme", but then D2 revises this to "test sur cycloergomètre", which is not wrong, but noticeably less understandable to laypeople. Translators make similar decisions, e.g., in the same text, T1 changes "maladies des artères coronaires" (literally: "disease of the coronary arteries") into "coronaropathies".

### 4.5 Notes on the Process and Practical Guidelines

Apart from the more theoretical findings of this pilot study, the experiment was also meant as a learning process to improve the practical aspects. For instance, it may be worth stressing patient-friendliness more and providing some examples in the guidelines, to encourage participants to pay more attention to this aspect. The main conclusions in this regard concern mechanical errors. These were not taken into account for any of the previous experiments, but were plentiful in all target texts. All texts contain tags, e.g. <1> and </1>, which are not always processed well by the MT. Participants did not receive specific instructions regarding these tags. This resulted in many tag-related errors, ranging from inconsistent placement around terms including or excluding the article, to more serious errors such as the removal of tags. For future tasks, a better strategy would be to completely remove tags from the source text before running the MT, and to either have very clear instructions for tag placement in the guidelines, or, perhaps even better, a separate step to manually place all tags. Other mechanical errors regard spacing (mostly double spaces), punctuation (not using the correct French punctuation, which is different from the Dutch one), and capitalisation (at the beginning of sentences, but also for the layout of a title). Some of these errors originated in the MT output, but they were also regularly introduced by the participants. Since it was discovered that the source text can be ambiguous, not well written, or sometimes even wrong, participants may benefit from instructions to contact the client in such cases and to make sure the target text is clear(er). These practical considerations all need to be taken into account in future research.

## 5 Conclusions and Future Work

The goal of this project is to research workflows for the translation of medical texts in a modern translation setting. The motivation behind this was the start of a large translation project of medical guidelines for lay people by the company iscientia, using machine translation output as a start. Research on this subject is fairly limited and empirical studies are scarce. The main research question raised was: is post-editing of medical texts best performed by translators, by doctors or by a combination of both and, if so, in which order? The hypothesis was that translators and doctors have different, likely complementary strengths, suggesting that a 2-step approach would be ideal, with a first round of post-editing by one type of professional, followed by a second round of revision by the other type of professional. The best order would need to be determined. The ongoing pilot study presented in this paper is a first attempt at researching these questions but, most of all, to figure out the best way of finding the answers. In this small-scale study, we work with four medical translations (Dutch to French), two doctors who understand Dutch and are native French-speakers and two experienced translators. We test two workflows: post-editing by a doctor, revision by a translator and vice versa.

While errors remained in both workflows, there were indications that the doctor-translator workflow might result in slightly better final target text. However, this difference is only noticeable at the revision step, since, after a single round of post-editing, translators and doctors performed almost identically, the only difference being the types of errors they solved most. Taking into account both the post-editing and revision steps, the translators in this study were able to correct more errors than the doctors, but at the cost of many unnecessary edits and more over-revisions as well. This is due at least in part to the translators' mistrust of the MT, which was made very clear when they regularly deleted the original MT output completely, to then rewrite the exact same translation. Hypotheses formulated about the expected behaviour of the different groups could only partially be confirmed. The intuition was that doctors would be better at solving mistranslations relating to the specialised medical content thanks to their expertise, while translators were expected to be more attentive to style & fluency. The counts of the error types solved by both groups was in line with the hypothesis: doctors did solve relatively more mistranslation errors, and translators more style & fluency errors. However, there were also surprising examples where doctors made serious mistakes regarding the medical

content and translators made badly-written translations. The patient-friendliness of the translations was difficult to assess, though the current experiments suggest that both groups struggle with this.

The main contributions of this pilot study are the lessons learnt for future, more elaborate research. First and foremost, even though this pilot study provided interesting findings and inspiration for further research, more participants of each group need to be tested to be able to draw well-founded conclusions. Second, the participants need to be screened more carefully. The two most important factors that need to be taken into account in this respect are experience with post-editing and experience with medical translation. It is suspected that both factors have a major impact on the results, especially for the translators. Translators without experience in post-editing clearly mistrusted the machine translation output and all participants were prone to making mechanical mistakes due to the editing of an existing text. Experience with medical translation could provide the significant advantage of trained research skills for domain-specific content and terminology. Further lessons learnt concern the evaluation. The machine error translation taxonomy (Tezcan *et al.*, 2017) has proven useful as a starting point for a maximally objective evaluation. Splitting the error types into two main groups allowed a nuanced evaluation when some of the individual error types were too infrequent to use. Since issues relating to patient-friendliness, such as readability and idiom are very difficult to assess, future research could focus on a limited number of items, which should be annotated in the machine translation output beforehand.

In conclusion, this pilot study was a first start towards trying to answer an issue that could help organisations such as iscientia make a more informed decision about a workflow for their translation project. The issues raised during this project provide valuable inspiration and lessons for larger-scale research on this topic.

## 6 Acknowledgements

Thanks are due to the two translators and the two doctors who took part in the experiment, for putting up with our instructions and our deadlines; and to Inez Vanoverschelde for her part in recruiting our test subjects. This research has been carried out as part of a PhD fellowship on the EXTRACT project, funded by the Research Foundation – Flanders.

## References

Andrea Cornwall and Karen Brock. 2005. What do buzzwords do for development policy? a critical look at ‘participation’, ‘empowerment’ and ‘poverty reduction’. *Third World Quarterly,* 26(7):1043–1060, October.

Joke Daems and Lieve Macken. 2019. Interactive adaptive SMT versus interactive adaptive NMT: a user experience evaluation. Machine Translation.

Maríıa Gonzâlez Davies. 1998. Student Assessment by Medical Specialists: An Experiment in Relating the Undergraduate to the Professional World in the Teaching of Medical Translation in Spain. In Translation and Medicine, volume 10 of American Translators Association Scholarly Monograph Series, pages 93–102. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Martha Monrad Hansen and Karen Korning Zethsen. 2018. Is Electronic Access to Medical Records an Empowering and Patient- centered Initiative? – A Qualitative Contextual and Linguistic Analysis of Danish Electronic Record. HERMES - Journal of Language and Communication in Business, (58):157–173.

Inger Holmström and Marta Röing. 2010. The relation between patient-centeredness and patient empowerment: A discussion on concepts. Patient Education and Counseling, 79(2):167–172, May.

Nisbeth Matilde Jensen and Korning Karen Zethsen. 2012. Translation of patient information leaflets: Trained translators and pharmacists-cum-translators – a comparison. Linguistica Antverpiensia New Series - Themes in Translation Studies, 11.

Marla O’Neill. 1998. Who Makes a Better Medical Translator: The Medically Knowledgeable Linguist or the Linguistically Knowledgeable Medical Professional? A Physician’s Perspective. In Translation and Medicine, number 10 in American Translators Association Scholarly Monograph Series, pages 69–80. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Ayla Rigouts Terryn, Isabelle S. Robert, Jim J. J. Ureel, Aline Remael, and Sabien Hanoulle. 2017. Conceptualizing translation revision competence: A pilot study on the acquisition of the knowledge about revision and strategic subcompetences. Across Languages and Cultures, 18(1):1–27, June.

Peter Salmon and George Hall. 2004. Patient empowerment or the emperor's new clothes. Journal of the Royal Society of Medicine, (97):53–56.

Arda Tezcan, Véronique Hoste, and Lieve Macken. 2017. SCATE Taxonomy and Corpus of Machine Translation Errors. In Gloria Corpas Pastor and Isabel Durán-Muñoz, editors, Trends in E-Tools and Resources for Translators and Interpreters. Brill, December.

Ana S. Vásquez, Ana B. Del Rio Lanza, Leticia Álvarez, and Rodolfo Casielles. 2017. Empower me? Yes, please, but in my way: different patterns of experiencing empowerment in patients with chronic conditions. Health Communication, 32(7):910–915.

Karen Korning Zethsen. 2018. Access is not the same as understanding. Why intralingual translation is crucial in a world of information overload. *Across Languages and Cultures,* 19(1):79–98.