The International Journal for Translation & Interpreting Research
trans-int.org

Back translation as a means of giving translators a voice

Uldis Ozolins
Royal Melbourne Institute of Technology University
uldis@languagesolutions.com.au

Abstract. While relatively little attention has been paid in translation literature to issues of back translation, and translators often express negative views about its validity, this methodology has become entrenched in a large field of translation particularly in international medical and psychosocial translation, and is promoted by bodies such as the World Health Organization. While some of the specific methodologies appear to have a mechanical view of back translation, this article argues that a suitable methodology of back translation can not only be a useful means of quality control if carefully used, but can perhaps unexpectedly put the translator in a position of having a voice vis a vis the clients, and be able to do what is too rarely the case in translation project management – establish an ongoing dialogue between a translator and client for mutual benefit. This article presents a case study of the translation of a medical diagnostic tool for prioritising hip and knee surgery in Melbourne, Australia, which employed a back translation methodology and resulted in robust interaction between authors and translators.

Keywords Translation; back-translation; medical translation; quality control.

Introduction

Back translation is a practice little studied in translation literature yet surprisingly prevalent in many areas of technical, particularly medical translation. Previous work by this author has studied the relative neglect or even vigorous disparagement of back translation in academic translation literature and among professional bodies (Ozolins 2008). Meanwhile, a substantial number of translators work on large projects where back translation has become a demanded methodology. Medical organisations, led by the WHO, have taken a particular interest in back translation as a quality control approach to achieve precise and comparable transfer of meanings across languages in international health studies.

This current article examines one translation project that may provide some useful methodological pointers to how back translation can be used in a considered way that both retains the ability of translators to present their work with integrity, and satisfies medical authors' demands for precision. A carefully used back translation process, as part of a complex checking process, can be based on transparent communication between translators and authors, enabling translators to have their voice heard by clients.

.

Back translation methodology in the medical field

Emanating from the work of Brislin (1970, 1986), the WHO and other international medical research and treatment organisations have seen back translation as a highly useful device in translating international questionnaires and surveys, as well as diagnostic and research instruments. The WHO recommended using forward translation, an expert group to assess the forward translation, back translation and pilot testing (www.who.int/substance_abuse/research_tools/translation/en/).

The context of such an approach is the increasing globalisation of medical practice and research, and the desire to find international comparative data on a myriad of health issues. Such views do not go unchallenged, and critiques

abound of attempts to translate medical and particularly psycho-medical and psychological data across cultural boundaries – an account of these debates over the validity of translation and cultural adaptation can be found in Bullinger (2003) and Ozolins (2008). Even here, however, few of the critiques look at the nitty-gritty of translation, aiming rather at larger questions of the validity of transferring usually Western-origin health concepts to other cultures.

The issue has also reached a more formal stage now as the evolving standards of translation in the USA or Europe do not include back translation as part of their repertoires of quality process. The US standard F2575-06 for example explicitly challenges what it considers a widely-held belief about the efficacy of back translation (ASTM International 2006). Translation agencies have to thread their way between such professional judgement and often mandated requirements of clients: one response, from the McElroy translation company in the USA, cites the standard's questioning of the

> validity of the back translation as an effective quality control step: According to the ASTM guide, "A back translation will not result in a text that is identical to the source text, and furthermore, a back translation is not necessarily a good indicator of the quality of the translation. We […] certainly agree, although we recognize the role that back translations play in our clients' needs to meet regulatory requirements. (Vitray, 2007)

Yet this cleavage between professional disdain of back translation, yet extensive practice of it, is no longer a mere issue of taste: despite critiques, the medical field now sees this methodology as something of a gold standard (eg Houchin and Wild, 2006). Such an ongoing cleavage indicates relatively hermetically sealed universes of practice and belief among different sectors of the translation profession.

The Multi-attribute Arthritis Prioritisation Tool

In Melbourne, Australia, in the early to mid 2000s the Victoria Centre for Rheumatic Diseases at the University of Melbourne had developed its Multi-attribute Arthritis Prioritisation Tool [MAPT], a questionnaire asking people suffering from severe arthritis a series of questions about their degree of pain, discomfort or problems associated with hip or knee pain.

Trials had shown that patients' responses to this questionnaire matched very closely actual clinical observations, making it a powerful diagnostic tool that could reduce the number of clinical visits necessary and give accurate rakings of priority for ultimate hip or knee replacement surgery, thus saving valuable clinical time and expense. The questionnaire had a series of questions each with a graded set of options, eg:

> Q: Do you have hip or knee pain that does not get better even while you rest (for example, while sitting)
>
> 1-None or mild pain
> 2-Moderate pain
> 3-Severe pain
> 4-Extremely severe pain
> 5-The pain is so severe that I cannot bear it.

The validity of the instruments rested on the fine gradations of patient response, and it was these gradations that were the essential items to maintain in any translation.

The particular methodology requested by MAPT was that of the European Organisation for Research and Treatment of Cancer, which had some interesting variations on the WHO model. The EORTC methodology involves

- Two independent forward translators
- Two independent back translators
- A coordinator to check back translations against source text, negotiate over discrepant items and resubmit them to further translation and/or back translation
- A report to the authors, who could then review the outcome and processes and ask for steps to be repeated again
- Pre-testing etc as for the WHO, plus more formal peer review on the most significant EORTC instruments. (Cull et al, 2002)

Interestingly, this methodology dispenses with any expert committee as in the WHO model, and runs the entire process at arm's length to the authors who only engage after the substantial work of securing appropriate back and forward translations has been accomplished. This model – seeing the translation process as a component of work largely outside authors' intentions - is clearly based on a double-blind model of clinical trials, where in this instance authors are removed from the toing and froing of the translation process; sufficient use of back translation will, it is assumed, eventually indicate the conceptually equivalent translation. Yet, how important the actual authors' intentions are, will become clear below.

Given that in most international instances translation would be into a single target language, the coordinator- presumably a bilingual project manager from the translation company or independent consultant – plays a crucial role in the EORTC methodology for translation from one language to another. But in Melbourne the agency was faced with a very different demand, and not for a single language. Translations were requested in 12 languages for use with non-English speaking patients in Melbourne hospitals: Arabic, Chinese, Croatian, Greek, Italian, Macedonian, Maltese, Polish, Russian, Spanish, Turkish and Vietnamese. What methodology would be appropriate here?

The translation agency approached was the Melbourne firm of All Graduates, who had previously completed back translations for clients but usually on highly technical items and usually only in extract translations. It was not initially favourable to the methodology proposed, arguing the well-rehearsed professional arguments against back translation, and proposing its own methodology involving a professionally accredited translator (Australia has a national accreditation system for translators and interpreters – www.naati.com.au) plus a detailed checking process using checking pro formas and consultation between translator and authors.

Any concern over theoretical dislike of back translation was added to by logistical issues: it would be no easy task to find say Turkish or Polish or Macedonian translators whose A language was English to perform the back translation. Nor was it considered appropriate here to go international in choice of translators: a knowledge of the likely readership population in

Australia and the language it understood was seen as essential. Moreover, the international literature had almost no reference at all to the application of back translation for the ultimate use of local, in this case immigrant populations; such methodologies had been overwhelmingly used for international translations and not for local multilingual target readers (Ashing-Giwa et al, 2008).

The researchers, by contrast, wanted a methodology that would fit in with international EORTC precedents and the demands of peer review. After negotiations, the following methodology was agreed on:

- Briefing by the MAPT team of the forward translators and checkers
- One forward translation
- Independent checking of the forward translation
- One back translation
- Comments (by email) by the MAPT team on any noted discrepancies between the English source text and the back translation
- Comments (by email) by each of the forward translator and back translator on the MAPT team's comments
- A final teleconference between MAPT team and forward translator on any unresolved items, and a further round of back translation if no resolution.
- Documentation on translators' qualifications and experience
- Pilot testing etc by the MAPT team.

While many issues were resolved by email, all languages did have a teleconference. Only one language ultimately had to have a second back translation. With the project underway, French was added as a 13th language, this time for an international reason as a group of the University of Melbourne's medical colleagues in France were also keen to try out this diagnostic tool.

Significantly, during briefing the attitude of translators and checkers was that the text of around 60 often short sentences would cause few problems; the language was natural and there was no difficult terminology; yes – the scales of intensity would require attention but would present no undue difficulty - and even the impossible title had been changed for the patients to 'Hip and Knee Questionnaire'. In some cases even stronger views were expressed: what was the fuss about, would a briefing be even necessary? By the end of the project these views in most cases were to have changed sharply.

The interaction - gradations and shades of intensity

We present here the written records of the interactions between the MAPT team and the translators involved. As anticipated, the issue of shades of meaning on the scales of options were a central focus of concern, but they were by no means the only issues that arose for back translations.

In some clear cases, the back translations were very useful in alerting the MAPT to shifts in meaning in the forward translation from the original, as in this case on pain stopping the patient going to sleep:

MAPT team: In the BT the frequency is related to pain, rather than to falling asleep or, as in the original, *going to sleep*. […] The emphasis needs to be on how often the person is unable to go to sleep because of pain. For example, for option 5, it should be *falling asleep all the time*, not *always feel pain*. That is, pain (that they may or may not always feel) stops the person from going to sleep every night when they first go to bed.

Forward translator: I get your point. I will rework these options accordingly.

[…] *Resolution after teleconference*: [The translator] has re-worked this so that the options read *I feel pain which sometimes prevents me from falling asleep* etc. for all options.

However, such occasions of clear error in emphasis in the forward translation were relatively rare. In other cases the problem arose in the back-translation itself, and was readily acknowledged:

MAPT team: Options 3, 4, 5. *Poorer* is not really the same as *worse*. It does not have the same intensity.

Forward translator: in FT, the word "tệ hơn" is used and means "worse", NOT "poorer"

Back translator: BT at fault! FT OK. I didn't express the FT translation with sufficient force in my BT.

MAPT team: Ok.

And while such responses were somewhat more common, the bulk of disputed instances for back translators were strongly defended by them, often with a pointer to the issue of how the back translator has to choose between a number of viable options:

Back translator: Please note in doing a BT always there will be a dilemma in choosing the words when the difference is very subtle, such as *mild* and *moderate*, *strong* and *severe*, as these words in Arabic are used in the same context. An example of this; the Arabic word used in the FT for option 2 means *moderate* and *mild*, and I chose to use the word *mild* but that does not mean the Arabic word used in the FT for moderate is not the right word.

Such exchanges properly consumed much time for MAPT team and translators as they worked through the scales of intensity; languages cut up these scales in different ways, and attention in these cases turned to the critical issue of whether there was an equivalent distance between the options to the distance between the options in English.

MAPT team: Check that it fits well as a pain intensity scale. Look at the gradations between options.

And this meant looking at all the options together, to ensure the correct gradation:

> *MAPT team*: Option 2. The word *rather* is too vague. Does it mean slightly or moderately? It is too subjective. The original is *moderately difficult*.
>
> […]
>
> *Resolution after teleconference:* the FT is not *rather* but *moderately*. The options are as follows:
>
> No or very little
> Yes, it is moderately difficult for me to enjoy my life
> Yes, it makes it very difficult…
> Yes, it makes it extremely difficult…
> Because of the condition of my hip or knee I cannot enjoy my life
>
> These are fine.

Linguistic issues and back translation noise

Any critic of back translation would however have been able to draw on many occasions in this project where the back-translation project led to wrong identification of discrepancies, usually responding to surface features brought about by differences between languages and giving false concern to the monolingual authors. The majority of alleged discrepancies were indeed linguistic issues where the back translation did not match the original, but where the forward translation was conveying the original in an appropriate manner. Such items included

- Apparent discrepancies in using singular or plural (there are no dedicated plural forms as such in e.g. Vietnamese)
- Plural forms used instead of singular where eg 'hip' and 'knee' have different grammatical genders in several languages
- Use of non-use of capital letters
- contractions (can't, it's, don't…) in the back translation when the original does not contract
- Apparently discrepant uses of tenses (again, tenses are often not marked in other languages in similar ways to that of English)
- 'hip or knee pain' (original) cf 'pain in the hip or knee' (most languages)

Many such apparent linguistic discrepancies are perfectly familiar to translators (and bilinguals in general) but often opaque to monolinguals. In sum they provide what may be termed 'back translation noise' in that surface linguistic features apparently signal non-conformity in translation where in fact this is not the case. Some brief examples can be given:

> *Use of the continuous tense:*
>
> *MAPT team*: Check *Are you having* with *Do you have*. Also, the BT has *that is not getting better*, whereas the original has *that does not get better*, which is different.

Forward translator: No difference in Greek - as is, if one considers the context, it says 'it does not get better'

Back Translator: The meaning in Greek is the same.

MAPT team: Ok.

'Any', *'some'* and '*my*'

An interesting case of linguistic difference was where the word 'any' was picked up in a back-translation where it did not appear in the original:

MAPT team: Q1. Ask about *any*. Make sure it is just *pain* in the FT.

Resolution after teleconference: Croatian does not use the word *any*. This is only in the back translation.

The reasons for this 'any' in the back translation is easy to understand: in English 'any' is not only a quantification term but also pragmatically a politeness term: for example a shopper is at least as likely to ask 'do you have any bread?' as to ask 'do you have bread?; which may be considered more abrupt; a distinction not uniform in other languages. And a doctor could certainly ask if a patient had 'any pain'. The back translator inserted 'any' to conform to this colloquial politeness maxim.

The opposite case of an original English word that did not appear in back translation was 'some': the original referred to the effect of pain and one option read *there are some things I cannot do.* When the lack of 'some' was picked up by the MAPT team, it brought this response:

Back translator: In the FT the word "some" is not used as it is not idiomatic and therefore in back translating it is not present. Not meaning any disrespect but if "there are things that one cannot do" does that not imply that they are some and not all?

The issue of 'my' as in the original *enjoy my life*, *my relationships* etc was picked up in several languages and affected a number of items:

MAPT team: The BT has *enjoy life* rather than *enjoy my life* (original). BT is less personal. What does the FT have?

Forward Translator: The FT is the same, following the logic that I can enjoy only my life and not somebody's else. There is no other natural way to say it in Macedonian.

Back translator: FT has enjoy life (which is more appropriate in the Macedonian language in terms of syntax)

MAPT team: Ok.

And where this issue was pursued nevertheless in this language, it brought a stronger response:

MAPT team: The idea of *my relationships* is not captured in the BT (*in relationships*). BT is less personal.

Forward Translator: See comment above. In Macedonian it is stylistically improper to use possessive pronouns more than once in the same sentence, unless necessary. In this particular case the logic is that ***my*** hip can cause difficulties only with ***my*** relationships and not with somebody else's, so it is superfluous to repeat the word ***my***. There is no other natural way to say it in Macedonian.

Authors' intentions – the case of 'some of the time'

In a few instances, however, an apparent linguistic difference revealed perhaps unexpected authors' intentions. In a number of languages *most of the time* (original) came back as 'in most cases' or variations thereof, where those languages did not employ the notion of time for such meanings. And on the item *some of the time* (original) back translations came back for most languages with 'sometimes', occasioning the MAPT team to repeatedly comment:

> *Sometimes* (BT) is close but it differs from *some of the time* (original) in that the latter implies more regular help than *sometimes*. *Sometimes* suggests that a person might receive help occasionally but not on a regular basis (i.e., a family member might drop in when they feel like it but not predictably), nor for most of the time. *Some of the time* implies that a person has regular help; for example, a family member or carer comes to help them take a shower every Wednesday and Friday.

This item occasioned considerable discussion in emails and the teleconferences. In perhaps the most interesting case, we had this exchange:

> *Forward translator*: I'd bet that most English speakers would miss the meaning of "help on a regular basis" when they read 'some of the time' here. To convey the idea that the person "has regular help", in Spanish we would have to say something like: *Con alguna regularidad* (With some regularity)
>
> *Back translator*: FT means both sometimes or some of the time
>
> *MAPT team's comments on translators' comments*: No, do not change from first translation.
>
> *Resolution after teleconference*: Stay with the first FT. Do not change.

This was an explicit questioning of whether what the MAPT team meant would be understood – by an English reader. Moreover, the translator offered to make the implicit understanding of regularity explicit, but this is not desired by the MAPT team, presumably as it would diverge from the English original. But that leaves the situation, perhaps not unusual in the construction

of questionnaires, where the authors have persuaded themselves an item has a certain implication but this implication may not be clear to a reader. And for the translator it is a peculiar challenge: the authors want the implication to come across in translation, but not be made explicit! In the end, the closest rendition in the target language is agreed on, which does not carry the implication of regularity– but then, does the English carry this implication?

Efficiency of back translation

The laborious though transparent process outlined in this project can be evaluated in various ways. A first question may be to what extent did this process lead to changes and improvements in the translations? Comparing the final version with what the checked forward translator had submitted in the first instance, changes varied significantly from language to language, from just one change in one language to nearly 20 changes in the most revised text, but with an average of around 5 changes per language. Relating this to the number of apparent discrepancies picked up by the MAPT team after back translation, an average of 24 discrepancies per language had been initially identified, ranging from 11 to nearly 50 for the various languages: we thus see that around a quarter to a fifth of identified discrepancies eventually resulted in changes to the translations.

A second question may be whether back translation is more effective than other checking processes.

By chance, this project did allow something of such a comparison. One of the languages translated was French, this time not for local patients but for a team of medical colleagues in France. The initial checked forward translation was sent to them, before having gone through the back translation process. The French team responded by suggesting some changes to the wording of some options in just one question, but otherwise approved the translation. Meanwhile, in Melbourne the back translation process was undertaken and led to a prolonged bout of communication between translators and MAPT team, longer than for any other language as points of language and meaning were discussed and eventually leading to a second back translation being undertaken - the only instance of a second back translation in the project. This resulted in a final translation arguably not significantly different from the first. Should the more favourable response of the French readers have been taken, saving effort?

A clear issue here was the degree of confidence that the authors wanted to have in the ultimate translation, and this confidence grew precisely through the interaction with translators, so that every nuance and variation could be explained to the authors' satisfaction, They were not prepared in this instance to rely on the opinion of an expert group (the French team) – interestingly, a procedure recommended as one step in the WHO methodology. Nor were they willing to have an arm's length relationship with the translation process, as implied in the EORTC methodology: they did not have a translation coordinator to do the initial sorting out of forward and back translations, but involved themselves directly with the translators. The aim was precision, pursued at every turn, with the ultimate thought not only of how the patients would read the questionnaire but how orthopaedic surgeons would interpret

it. This extract on an item asking how difficult hip or knee pain made it to look after oneself, shows this persistence:

> *MAPT team*: The idea of looking after oneself as being *harder* is not the concept here. The BT suggests that although it might be *harder* (than what – there is no comparator) the person is still able to do it. The surgeon needs to know the level to which patients can look after themselves; the amount to which the person's hip or knee causes them difficulties. That is, does the person's hip or knee *make it difficult* (and the options supply degrees of difficulty) to look after oneself? I know it might seem like a similar concept but it is important that it is specific to the surgeon's requirements.

Rather than an arm's length methodology, it was this hands-on engagement in the translation process that ultimately gave the MAPT team confidence in the translations.

Equally important to the MAPT team's confidence, is the question of how well translators could cope with this methodology, and how much they had confidence in having their own work understood and appreciated. As mentioned earlier, translators were not always favourably disposed to back translation, and were not shy in giving their own views on back translation more generally. Some took an academic approach:

> *Forward translator*: Back-translation as a checking tool has been successfully used effectively for many years. However it has certain inherent limitations. Firstly, back-translation works best when the languages and cultures involved are very close eg Dutch and German etc. Turkish is Ural-Altaic language. It has about 90,000 words whereas the English language has about 350,000. This alone shows that words in one language may not have exact equivalents in the other. Translation process is more than replacing words with the words in the other language. Often the same meaning needs to be expressed using different words, which back-translation will not reflect …

And this explanation continued through a number of different phrases and expressions that illustrated different syntactic as well as lexical items that could send a back-translator in many directions. Moreover, the translator continued, 'some errors (eg, choosing the right register, equivalent, spelling or typos) may occur in the back-translation process.' The difficulty of determining whether identified mistakes have been made in forward or back translation has been commented on in the little literature devoted to these questions (Grunwald & Goldfarb 2006).

A small number of back translators remained quite unconvinced. The back-translators of course never saw the full original document, so the feeling of being 'ambushed' was apparent in several instances. One back translator in particular saw this methodology as inevitably being flawed, breaking from a detailed comment on an item to make a more general point:

> Dear [MAPT team]: I strongly recommend you speak to your translators face to face. What is coming through is lack of meaning you wanted, which is exactly what is to be expected if you do not talk to either the FT or the BT. Suggest that you talk to your FT to explain

> the SENSE you want to convey, not try to match the words. Word matching will not work.

And this was repeated in various ways at a number of other items:

> Once again, I would much rather the Melbourne Uni team talk to a translator to explain the flavour they want to get across, and this can be done. Using FT/BT/ blind will simply lead to the wrong nuances being imported.

And finally the comments became cryptic:

> Once again, garbage in garbage out. If you want meaning, you need to explain what you want prospectively.

Yet the way in which the translators engaged in the task, the flow of communication and the ability to have their voice heard, however critically, and to explain their choices meant that they too had confidence in the final product. Overwhelmingly, the translators engaged in this project – even the last cited - were willing to do further translations for the same clients using similar methodologies.

Moreover, the responses from the translators also led to changes of understanding by the MAPT team. Certain stages of a translation project may open the way for client education, and the use of this particular methodology makes these occasions more frequent: at one point the MAPT team noted its understanding of the issues faced by the Chinese translator in this context:

> Interesting detail from [the translator] about the way Chinese functions. It seems the language needs to provide pictures (is the way I understood it). This means, for example, that concepts like *length* are explained as *how short* and *how long*. It is difficult for me to understand (and then explain) but it is worth keeping in mind when going over these options.

And this awakening understanding was used in specific items:

> *Resolution after teleconference*: Discussed gradations of options and I'm happy with [the translator's] responses. *Very difficult* (original) translates into *large difficulties* in Chinese. Again, I think this stems from the need for Chinese to express in pictures – *very difficult* is a concept. In some way I can see that *large difficulties* can express a picture…

Conclusion

The MAPT project thus provided a methodology that ensured great confidence for the authors in the translations produced, while also proving an unusually rich occasion for translators to provide clients with an understanding of their tasks and challenges, providing a learning experience for both sides. This was brought about by the willingness of the MAPT team to engage in this hands-on interaction, rather than expect an outcome from an arm's length methodology, and translators taking a perhaps all too rare opportunity to be able to explain their choices and challenges in depth.

Further case studies of other back translation projects may provide valuable understanding of whether they provide an opportunity for such useful interaction, or are limited to a double-blind methodology with little interest in the translation process.

Notes: The Multi-attribute Arthritis Prioritisation Tool [MAPT] is a copyrighted and proprietary instrument. For enquiries about its use and translations see www.crd.unimelb.edu.au/academic/projects/oahipknee.html

All Graduates agency is at www.allgraduates.com.au

References

Ashing-Giwa et al (2008) Measuring quality of life among cervical cancer survivors: preliminary assessment of instrumentation validity in a cross-cultural study. *Quality of Life Research*, 17.

ASTM International (2006) *F2575-06 Standard Guide for Quality Assurance in Translation*. http://www.astm.org/Standards/F2575.htm

Brislin, R.W.(1970) Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology* 1.

--- 1986 TheWording and Translation of Research Instruments. In W.L. Lonner and J.W. Berry, eds *Field Methods in Cross-Cultural Research*. Newbury Park, CA: Sage.

Bullinger, M. (2003) International comparability of health interview surveys: An overview of methods and approaches In A. Nosikov and C. Gudex (Eds.) *EUROHIS: Developing Common Instruments for Health Surveys. Amsterdam:* IOS Press.

Cull, et al (2002) *EORTC Quality of Life Group Translation Procedure.* Available at: http://www.eortc.be/home/qol/Manuals/Translation%20Manual%202002.pdf

Grunwald, D and N.M. Goldfarb (2006) Back Translation for Quality Control of Informed Consent, Journal of Clinical Research Best Practices, 2 (2). Available at www.translationdirectory.com/article1043.htm.

Houchin, C and D. Wild The Translation and Linguistic Validation of the Impact of Weight on Quality of Life (IWQoL-LITE) Questionnaire. Paper presented at the international conference of the ISOQOL. Available at http://www.isoqol.org/2006AbstractsBook.pdf

Ozolins, U. (2008) Issues of back translation methodology in medical translations. *Proceedings, FIT [International Federation of Translators] XVIII Congress*, Shanghai.

Vitray, Kim (2007) 'The New ASTM Translation Guide' Available at http://www.translationdirectory.com/articles/article1315.php