The FAR model: assessing quality in interlingual subtitling
Jan Pedersen, Stockholm University

ABSTRACT

To this day, there exists only a generalized quality assessment model for intralingual (live) subtitles for the deaf and hard of hearing – the NER model. Translated subtitles seem to be quality assessed mainly using in-house guidelines. This paper contains an attempt at creating a generalized model for assessing quality in interlingual subtitling. The FAR model assesses subtitle quality in three areas: Functional equivalence (do the subtitles convey speaker meaning?); Acceptability (do the subtitles sound correct and natural in the target language?); and Readability (can the subtitles be read in a fluent and non-intrusive way?). The FAR model is based on error analysis and has a penalty score system that allows the assessor to pinpoint which area(s) need(s) improvement, which should make it useful for education and feedback. It is a tentative and generalised model that can be localised using norms from guidelines, commissioner specs, best practice etc. The model was developed using existing models, empirical data, best practice and recent eye-tracking studies and it was tried and tested on Swedish fansubs.

KEYWORDS

Subtitling, quality assessment, NER model, functional equivalence, acceptability, readability, error analysis.

1. Introduction

What is quality in translation? Quality is about as elusive an idea as 'happiness,' or indeed, 'translation.' Quality means very many different things depending on your perspective. To those in translation management, the concept is often associated with processes, work flows and deadlines. To professionals, quality is often a balancing act between input and efficiency. To academics, it is often a question of equivalence and language use. Often, quality assessment descriptions end up as descriptions of quality procedures, where quality can be something as indirect as the translator's competence (cf. e.g. ISO EN 17100 or Robert & Remael: forthcoming). Quality in the translated product itself is often deliberately left vague – particularly for subtitling. This is hardly surprising, considering how elusive the concept is, and there is a general reluctance to pinpoint what actually constitutes quality. Still, many people have to judge translation quality on a daily basis: revisers, editors, evaluators, teachers, not to mention the subtitlers themselves, and of course: the viewers.

A great deal of work has gone into assessing intralingual subtitles for the deaf and hard of hearing, particular for live subtitling, and successful models have been built to ensure that there are objective ways of assessing quality for this form of subtitling (particularly the NER model; cf. Romero Fresco & Martinez 2015). However, these models are not very well equipped for handling the added complexities of translation between two natural languages in a subtitling situation. Again, there are models for assessment in machine translation, mainly for the benefit of post-editing, and these have been used for subtitling as well (cf. e.g. Volk & Harder 2007). These

models do not go into great detail about what the quality assessor is supposed to investigate, however.

This paper presents an attempt at producing a tentative model for assessing that elusive beast, the quality of interlingual subtitles (as a product, not a process). The FAR model is generic, but is meant to be localised by including the appropriate norms. The model is tripartite: the first part assesses Functional equivalence. The second part assesses Acceptability: grammaticality, idiomaticity etc. The third part assesses Readability: technical aspects, such as reading speed, the use of italics and subtitling punctuation and so on. The FAR model is based on error analysis, and each error is given a penalty point, which means that each subtitled version gets a score that makes it is possible to compare the quality of subtitles from different films or TV programmes.

We have often been told that interlingual subtitling quality is too complex to measure; still people do it every day, so apparently it is possible. This model is an attempt at pinpointing the silent knowledge that lets us as professionals and academics do just that. The model has been tried in the evaluation of Swedish fansubs of English-language films (cf Pedersen: in preparation), so it was constructed using input from real data.

2. Assessing translation quality

There have been many attempts at creating models for translation assessment, one of the earliest and most thorough being that of Juliane House in her monograph *A Model for Translation Quality Assessment* from1981. One of the main problems encountered is how to define a useful unit of translation, from Vinay & Darbelnet's (1958/2000) 'thought units' (unités de pensée) to Leuwen Schwarz's 'transemes' (1989: 155–157), the nature of the problem being if we cannot decide what should be equivalent to what, how can we then assess if equivalence has been achieved? And there we have that other problematic notion: equivalence. This concept has been used in so many senses that Gideon Toury (1995) passed over the issue by assuming that equivalence was there, and then focused on showing what forms it took. That works well if you are doing descriptive work, but if you want to build a model for quality assessment, you have to make a choice. What sort of equivalence is relevant for your assessment? The answer depends on many factors, such as the genre of the source text (ST) and skopos (cf. Vermeer 1989/2000) of the target text (TT). In the academic world, there has been a great deal of development to solve these issues and much new knowledge about the nature of translation has been the result of this.

In the business world, the focus has been mainly on translation processes and work-flow management. The ISO standard that regulates translation quality (EN 17001) has its main focus on the process, stipulating what

competences a translator should have, and how the revision process should be carried out and so on. It is quite vague when it comes to pinpointing quality in the actual translation product. This goes for many process-based models of translation quality. They are very useful for managing translation quality, but not always very useful for assessing it in a product. Quality in the translation process is very important, but as this paper is concerned with quality in the end-result of that process – the product in the form of subtitles – these models will thus not feature very prominently here.

Much work has been carried out in the field of machine translation, where quality is central to the process of post-editing (cf. e.g. Temizöz: forthcoming). For this purpose, models such as the LISA QA Metric have been developed to quantify the errors of automated translation. This model comes from the localisation industry and language is but a small part in the model, others being areas such as formatting and functionality.

Another problem with translation assessment is that different companies and other stake-holders tend to have their own system and metrics for assessing it (O'Brien 2012). It was in response to this that the EU Multidimensional Quality Metrics (MQM) model (Burchardt & Lommel 2014) was created to try combine the different systems and create a comparable system of translation quality metrics. This is a very good attempt at commensurability, and some of its features will be used in the present paper, but its comprehensive nature makes it a somewhat unwieldy system that is perhaps not always very easy to apply to actual texts (and it is also somewhat process-oriented).

The main problem with general translation quality assessment models when applied to subtitling is that they are difficult to adapt to the special conditions of the medium. They thus often see e.g. omissions and paraphrases as errors. This is very rarely the case in subtitling, where these are necessary and useful strategies for handling the condensation that is almost inevitable (cf. e.g. De Linde and Kay 1999: 51 or Gottlieb 1997: 73), or which may be a necessary feature when going from speech to writing (cf. e.g. Gottlieb 2001: 20).

3. Assessing subtitling quality

As part of various countries' efforts to increase audiovisual accessibility, there has been a boom in intralingual subtitling in the form of subtitling for the deaf and hard of hearing (SDH). Once the quotas for SDH had started to be met and the quantity issue was settled, so to speak, concern started to shift towards quality instead (cf. e.g. Romero-Fresco 2012). For this purpose, the NER model was constructed (Romero-Fresco & Martinez 2011). This is loosely based on previous models that focus on WER (Word Error Rate). The NER model, however, is adapted for respoken intralingual subtitles, and uses "idea units" as the translation units, rather than words.

The model has been used very successfully by the British regulator Ofcom for its three year study of SDH quality in English-language subtitles (Ofcom 2014; 2015). Not only that, but the model has been used in many other countries as well (Romero Fresco & Martinez 2015). The model is product-oriented, has high assessor intersubjectivity and works very well for intralingual subtitles. It has thus been used as inspiration for the model presented in this paper.

There are some very important differences that make the NER model hard to apply to interlingual subtitles, however. First and foremost, interlingual subtitles involve actual translation between two natural languages, and not only the shift in medium from spoken to written language. This means that the whole issue of equivalence takes on a completely different dimension. The NER model takes accuracy errors into account, but has not dealt much with equivalence errors. Secondly, the preconditions of SDH and interlingual subtitles tend to be very different. A great deal of SDH (and particularly that which the NER model was designed for; Romero Fresco & Martinez 2015) is in the form of live subtitles, whether produced via respeaking or special typing equipment. Interlingual subtitles are almost always prepared (and also spotted) ahead of airing, which means that issues such as latency (cf. e.g. Ofcom 2014: 9) which are a major concern in SDH, tend not to be an issue for interlingual subtitles. Thirdly, it could be argued that viewer expectations of (live) SDH and (prepared) subtitles are different. There may be less tolerance for errors in interlingual subtitles, as they are produced in a less stressful situation. Also, the shift in language means that reformulations can be seen by viewers as language shifts, or translation solutions (or errors, if they are less aware of language differences or less kind), whereas drops in accuracy levels are often seen by SDH viewers as "lies" (Romero Fresco 2011: 5).

Since the NER model, despite its many advantages, is not in its current state very well suited for interlingual subtitles, what is then used to assure quality in interlingual subtitling? To ascertain what models exist for assessing the quality of interlingual subtitles, I interviewed executives and editors of SDI Media in Sweden (Nilsson, Norberg, Björk & Kyrö) and Ericsson (Sánchez) in Spain. Their replies were rather process-oriented and concerned with in-house training, collaboration, and revision procedures, which are certainly very important strategic points at which to achieve quality. When it comes to defining product quality assessment, the answers were more geared towards discussions and guidelines, even if SDI Media produced a short document entitled "GTS Quality Specifications – Subtitling"[1].

Sánchez said that subtitling is an art, and how do you measure the quality of art? She then conceded that it had to be done nevertheless (personal communication), and then mainly by using guidelines. It would appear that in-house guidelines are the most common product-oriented quality tools

that many companies have (e.g. at SDI-Media, Nilsson *et al*, personal communication). In-house guidelines can vary a great deal in detail, but they are undoubtedly very useful, not only for training, but also for telling right from wrong when it comes to quality control. One problem with these guidelines, however, is that they can sometimes be hard to come by for outsiders, as they are in-house material, and they can also not be used when subtitlers themselves have not used them, as when investigating fansubbing, for instance. Also, since each company uses its own set of guidelines, we have no way to compare results[2].

In an attempt to chart subtitling quality, Robert and Remael (forthcoming 2016) carried out a recent survey on how translation quality was seen in interlingual subtitling by subtitlers and commissioners. The survey was fairly process-oriented, charting, as it did, pre-translation, translation and post-translation procedures. It did, however, give a very good overview of what areas subtitlers and the commissioners considered to be important quality areas. Incidentally, it is particularly interesting to note that subtitlers are more thorough and think that more areas are worthy of quality attention than commissioners do. Robert and Remael surveyed technical and translation quality parameters. They consider content, grammar, readability and appropriateness to be translation parameters, and style guide, speed, spotting and formatting as technical parameters. One might argue that many of the technical parameters are in fact related to readability. If the speed of the subtitles is too high, they become unreadable, and that can also be the case if the formatting or spotting is erroneous. As outside sources of quality assessment they quote the EN15038 standard, the precursor of the EN 17001, and also the Code of Good Subtitling Practice (Ivarsson and Carroll 1998: 157–159). This Code, which has been adopted by the European Association for Studies in Screen Translation (ESIST), is an early, but still valid, attempt at formulating what is high-quality subtitling practice. To be generally applicable, however, it is deliberately vague, and says more about subtitling processes than subtitled products.

4. A tentative model for interlingual subtitle quality assessment

As stated in the introduction, the point of this paper is to present a tentative general model for the assessment of quality in interlingual subtitles. That, in itself, presents two limitations to the scope. It does not purport to be useful for measuring the quality of intralingual subtitles (SDH) as there is already a very useful model for that (see above). Nor does it purport to have anything to do with the subtitling process, even though quality in the process is the precondition for quality in the subtitles as a product. Instead it looks on the finalised products, i.e. the subtitles themselves, in relation to the polysemiotic source text, including non-verbal as well as verbal semiotic channels. The model could be useful for anyone involved in assessing the quality of subtitles, be they made by professionals, amateurs, students or whomever. The model is partly based on the practices used for

assessing student subtitles at Tolk- och översättarinsitutet at Stockholm University (TÖI) and partly on the NER model, and it has been tested on a sizeable sample of fansubbed English-language films (cf. Pedersen: in preparation). The model is viewer-centred in that it takes account of reception studies using eye-tracking technology (cf. e.g. Ghia 2012; Caffrey 2012; Lång *et al* 2013 or Romero Fresco 2015) and also in that it is based on the notion of a tacit 'contract of illusion' (cf. Pedersen 2007: 46–47 and below) between subtitler and viewer. In an attempt at being comprehensive, it purports to investigate all areas of quality that can affect the viewers. Finally, it is intended to be a general model that can be localised by feeding it parameters with data from in-house guidelines, best practice or national subtitling norms. Like many other translation quality assessment models (cf. O'Brien 2012) it is rather depressingly, if quite necessarily, based on error analysis. That means that there will be no plus points for good translation solutions, no matter how brilliant. This is, of course, unfair, but necessary if subjectivity is to be kept at a minimum.

Having made all these bold claims, it is probably prudent to make a few caveats. It is true of subtitles, like of any form of translation, that "there is no such thing as a single correct translation of any text as opposed to one-for-one correct renderings of individual terms; there is no single end product to serve as a standard for objective assessment" (Graham 1989:59). It is probably too much to hope for that objectivity can be reached in the assessment process; we can only be thankful if some measure of assessor intersubjectivity can be achieved, even of it may not be as high as in NER model (Romero Fresco & Martinez 2015), as that actually has a "standard" of sorts: a source text in the same language.

4.1. Contract of illusion

Before describing the proposed model itself, it could be useful to say a few words on interlingual subtitles and their relationship with the end consumer. In Pedersen 2007 (46–47), I invented a metaphor for this relationship, that of a contract of illusion, between subtitler and viewer. This is (like quite a few other things in this paper) based on Romero Fresco's work, in this case his adaptation of Coleridge's famous lines of suspension of disbelief, which Romero Fresco applied to dubbing in 2009. He means that even though the audience knows that they are hearing dubbing actors, they suspend this knowledge and pretend that they hear the original lines. In the case of subtitling, there is a similar suspension of disbelief which is part of the contract of illusion: viewers pretend that subtitles are the actual dialogue, which in fact they are not. Subtitles are a partial written representation of a translation of the dialogue and text on screen, so there is a great deal of difference. In fact, due to the fact that the reading of subtitles becomes semi-automated (cf. e.g. d'Ydewalle & Gielen 1992), the viewers' side of the contract extends even further than pretending that the subtitles are the dialogue. The viewers even do not notice (or suspend their noticing) the

subtitles. To my mind, that is quite a feat of suspension of disbelief, and very accommodating of the viewers. In return, as their part of the contract, the subtitlers assist the viewers in suspending their disbelief by making their subtitles as unobtrusive as possible. This is the reason why subtitling tends to favour a transparent and fluent translation ideal (plus the fact that it is hard for viewers to reread a resistant subtitle). This also explains prolific subtitling aphorisms extolling the fluency ideal, like "the good subtitle is the one you never notice" (Lindberg 1989; my translation) or Sánchez (personal communication) saying that things that make you aware that you are reading subtitles are errors. For a further discussion on how AVT may affect the viewers' immersion in audiovisual media, the reader is referred to e.g. Wissmath & Weibel (2012).

Recently, the contract of illusion has been challenged in certain genres, such as fansubbing and certain art films, which experiment with subtitling placement, pop-ups, fonts and other exciting new things (cf. e.g Denison 2011 or McClarty 2015). This is a welcome development in some ways, but as by far the vast majority of subtitling still adheres to the fluency ideal, the contract of illusion can still be used for assessing the quality of subtitles that purport to follow this ideal.

4.2. Basic unit of assessment

WER (Word Error Rate) models tend to be used in speech recognition assessment (Romero Fresco & Martinez 2015: 30), which, simply put, means that you divide the number of words in a text by the number of errors in it. That is clearly not very useful for subtitling with its need for verbal condensation. Hence the NER model uses dependent and independent idea units as the basis of its rating system, which seems to work well. It is uncertain whether that is a suitable unit for translated subtitles, however, as the concept is somewhat vague, and also, omitting idea units is sometimes necessary in interlingual subtitling, without normally making a subtitle worse. The NER model addresses this problem by adding another step that analyses the nature of the omissions (or "correct editings" 2015: 33). At Ericsson, they use a minute of air time as the basic unit (Sánchez, personal communication), but that seems rather crude, as dialogue intensity varies enormously. In translation theory, there are a multitude of units presented as the basic unit of translation, as mentioned above, but in the present model, I would like to put forward that the most natural unit to use in subtitling is the (one or two-line) subtitle itself. This cannot be used for live subtitling, as these are often of the "rolling" kind (cf. Romero Fresco 2012), and live subtitlers have no control over segmentation. For interlingual prepared subtitles, the subtitle as unit of assessment is not only intuitive, but also has other advantages. Firstly, it is a clearly and easily defined unit, which is also ideally semantically and syntactically self-contained (cf. Code of Good Subtitling Practice). Secondly, an error in a subtitle breaks the contract of illusion and makes the viewer

aware that they are reading subtitles and that may affect not only a local word or phrase, but the processing of information in the whole subtitle. This is indicated by Ghia's eye-tracking study of deflections in subtitling, where viewers find that they have to return to complicated subtitles after watching the image (2012: 171ff).

5. The FAR model

In homage to the NER model (originally the NERD model, cf. Romero-Fresco 2011: 150) and its success in assessing live same-language subtitles, I would like to name this tentative model the FAR model. This is because it looks at renderings of languages that are not "near" you (i.e. your own) but "far" from you (i.e. foreign). Apart from the mnemonic advantages of the name, the letters stand for the three areas that the model assesses. The first area is Functional equivalence, i.e. how well the message or meaning is rendered in the subtitled translation. The second area is the Acceptability of the subtitles, i.e. how well the subtitles adhere to target language norms. The third area is Readability, i.e. how easy the subtitles are for the viewer to process. Actually, "how well" is somewhat misleading as the model is based on error analysis. This is the most common way of assessing the quality of translated texts, which is shown in O'Brien's 2012 study where 10 out of eleven models used error analysis. Two of them also highlighted particularly good translation solutions, but did not award any bonus points.

For each of the FAR areas, ways of finding errors and classifying the severity of them as intersubjectively as possible will be laid out here, and a penalty point system will also be proposed. This enables the users to assess each subtitled text from these three different perspectives. The penalty point system makes it possible to say in which area a subtitle's text has problems, and it can therefore be used to provide constructive feedback to subtitlers, which would be useful in a teaching situation. The error labels and scores are imported from the NER model and they are 'minor,' 'standard' or 'serious' (Romero Fresco & Martinez 2015: 34–41). The penalty points vary according to the norms applied to the model, but unless otherwise stated, the proposed scores are 0.25, 0.5 and 1 respectively. The penalty points an error receives is supposed to indicate the impact the error might have on the contract of illusion with the end user. Thus, minor errors might go unnoticed, and only break the illusion if the viewers are attentive. Standard errors are those that are likely to break the contract and ruin the subtitle for most viewers. Serious errors may affect their comprehension not only of that subtitle, but also of the following one(s), either because of misinformation, or by being so blatant that it takes a while for the user to let go of it and resume automated reading of subtitles.

The Code of Good Subtitling Practice improved its generalisability by being deliberately vague. The FAR model does this by being incomplete, in that it should be fed local norms, as presented in in-house guidelines, best

practices or national norms. In the following sections we will deal with each area in turn, with examples based on Swedish national norms for television (cf. Pedersen 2007) and Swedish fansubs.

5.1 Functional equivalence

It is easy to go astray when discussing equivalence, as discussed above. I want to avoid a lengthy discussion on the various forms of equivalence that abound explicitly in Translation Studies and implicitly in the practice of users. Instead, I will simply state that for subtitling, with its many constraints of time and space etc. the best form of equivalence for subtitling is pragmatic equivalence. I proposed this in Pedersen 2008, and I still hold it to be true, and I know this is shared by others, e.g. Ben Slamia (2015) or Gottlieb who claims that subtitling is the translation of speech acts (2001: 19). Without going too deep into speech act theory (for this, the reader is referred to Pedersen (2008)), I would like to say that in subtitling, it is not so much what you say, as what you want to communicate that matters. This means that the actual words spoken are not as important as what you intend to get across. The obvious reason for this is that there is not always room to replicate the original utterances, and often it is not a very good idea to do so, as it would affect the readability of the subtitle.

Ideally, a subtitle would convey both what is said and what is meant. If neither what is said nor what is meant is rendered, the result would be an obvious error. If only what is meant is conveyed, this is not an error; it is just standard subtitling practice, and could be preferred to verbatim renderings. If only what is said is rendered (and not what is meant), that would be counted as an error too, because that would be misleading. Equivalence errors are of two kinds: semantic and stylistic.

5.1.1. Semantic errors

To reflect how central semantic equivalence is in interlingual subtitling, and the assumed lower tolerance for errors that the users of interlingual subtitles have, the penalty points for semantic equivalence are minor: 0.5, standard: 1, and serious: 2.

The following is an example of a minor semantic error taken from a corpus of Swedish fansubs of English-language films (cf. Pedersen: in preparation)[3]. In *The Number 23*, the protagonist finds a smallish book in a second-hand bookshop. The title of the book is:

(1)   ST: "The Number 23"
"A Novel of Obsession by Topsy Kretts¨
         TT: "Nummer 23."
"En novel av besatthet av Topsy Kretts."
         BT: "Number 23."

> "A short storry [sic] of obsession
> by Topsy Kretts"
> (*The Number 23*: 10:52[4])

Novel/novell[5] is a false friend, as the word means 'short story' in Swedish, so translating *novel* as 'novell' is an error. It is not a serious error, as both words refer to a book, and the book on screen is rather short. It is thus a minor error, which gets an error score of 0.5. Minor functional equivalence errors are basically lexical errors, including terminology errors which do not affect the plot of the film.

Standard errors, with a score of 1 are exemplified by (2), which includes two such errors. Once again, from *The Number 23*, the voice-over narrator philosophises about the meaning of time:

> (2)   ST: Time is just a counting system; numbers with meanings attached to them.
>   TT: *"Tiden är bara ett räknande system, nummer med betydelse som slår ihop dem."*
>   BT: *"The time is just a system that counts, numbers with meaning that bang them together"*
> (*The Number 23*: 41.07)

The Swedish translation is an almost verbatim rendition of the words of the original utterance, but the meaning is completely lost, which illustrates that errors can be made even if the words are translated. This need not be the case, however, as there are also other ways of making a standard error. The definition of a standard semantic equivalence error would be a subtitle that contains errors, but still has bearing on the actual meaning and does not seriously hamper the viewers' progress beyond that single subtitle. Standard semantic errors would also be cases where utterances that are important to the plot are left unsubtitled.

A serious semantic equivalence error scores 2 penalty points and is defined as a subtitle that is so erroneous that it makes the viewers' understanding of the subtitle nil and would hamper the viewers' progress beyond that subtitle, either by leading to plot misunderstandings or by being so serious as to disturb the contract of illusion for more than just one subtitle. The latter is exemplified (3) from the same film, where the protagonist muses on how his life has not turned out the way the stars had foretold:

> (3)   ST: I am living proof of the fallacy of astrology
>   TT: *Jag lever ständigt av en orimligt hög av "lustingar."*
>   BT: *I am constantly living off an Unreasonably*[sic] *pile of "lusties."*
> (*The Number 23*: 11.31)

The error in (3) is so serious that it renders the subtitle impossible to understand and would presumably cause frustration for more than that subtitle.

5.1.2. Stylistic errors

Stylistic errors are not as serious as semantic errors, as they cause nuisance, rather than misunderstandings. The score for these are thus the same as for the NER model. Examples of stylistic errors would be erroneous terms of address, using the wrong register (too high or too low) or any other use of language that is out of tune with the style of the original (e.g. using modern language in historic films).

5.2. Acceptability

The previous area is related to Toury's notion of adequacy, i.e. being true to the source text, and this area, acceptability (1995: 74), is to do with how well the target text conforms to target language norms. The errors in this area are those that make the subtitles sound foreign or otherwise unnatural. These errors also upset the contract of illusion as they draw attention to the subtitles. These errors are of three kinds: 1) grammar errors 2) spelling errors, 3) errors of idiomaticity.

5.2.1. Grammar errors

These are simply errors of target language grammar in various forms. It would make little sense to list them here as they are language-specific. However, it should be pointed out that it is the target language grammar as adapted for subtitling that is relevant here. Subtitling can be seen as hybrid form of spoken and written language (cf. Pedersen 2011:115), which means that a strict application of written language grammar rules may be misguided. Many languages, e.g. Swedish, allow for certain typical spoken-language features in subtitling that would normally be frowned upon in many other written text genres. Such features are e.g. certain cases of subject deletion, incomplete sentences and shortened forms of pronouns.

A serious grammar error makes the subtitle hard to read and/or comprehend. Minor errors are the pet peeves that annoy purists (e.g. misusing 'whom' in English). Standard errors fall in between.

5.2.2. Spelling errors

Spelling errors could be judged according to gravity in the following way: a minor error is any spelling error (like the one in example (1)), standard errors change the meaning of the word, and serious errors would make a word impossible to read. This perspective differs from that of the NER model, which considers change in meaning to be worse than unintelligibility (Romero Fresco & Martínez 2015: 34). The reason for this is due to

differences in viewer expectations: consumers of prepared interlingual subtitles have less tolerance for errors than viewers of live subtitling have and also have access to the dialogue, which is often not completely unknown to them.

5.2.3. Idiomaticity errors

In this model, idiomaticity is not meant to signify only the use of idioms, but the natural use of language; i.e. that which would sound natural to a native speaker of that language. In the words of Romero Fresco "idiomaticity is described […] as nativelike selection of expression in a given context" (2009: 51; emphasis removed). Errors that fall into this category are not grammar errors, but errors which sound unnatural in the target language. The main cause of these error is source text interference, so that the result is "translationese" (cf . Gellerstam 1989), but there may be other causes as well. Also, garden-path sentences may be penalised under this heading, as they cause regressions (re-reading), hamper understanding and thus affect reading speed (Schotter & Rayner 2012: 91). It should be pointed out that sometimes source text interference can become so serious that it becomes an equivalence issue, as illustrated in example (2).

5.3. Readability

In this area, we find things that elsewhere (e.g. Robert & Remael: forthcoming or Pedersen 2011: 181) are called technical norms or issues. The reason why they fall under readability here is that the FAR model has a viewer focus, and presumably, viewers are not very interested in the technical side of things, only with being able to read the subtitles effortlessly. Readability issues are the following: errors of segmentation and spotting, punctuation and reading speed and line length.

5.3.1. Segmentation and spotting

Literature on subtitling stresses the importance of correct segmentation and spotting (cf. e.g. Ivarsson & Carroll 1998; Díaz Cintas & Remael 2007; Tveit 2004). *The Code of Good Subtitling Practice* (Ivarsson & Carroll 1998) devotes a full eleven separate points to various aspects of this, and the details of it are too many to go into here. Suffice it to say that flawed segmentation may distract the viewer, as it has been proved in eye-tracking studies that unusual segmentation "increases considerably the time in the subtitled area" (d'Ydewalle *et al* 1989: 42). Spotting errors are caused by bad synchronisation with speech, (subtitles appear too soon or disappear later than the permitted lag on out-times) or image (subtitles do not respect hard cuts). Lång *et al*'s eye-tracking study (2013: 78) has shown that delayed subtitles make viewers search for subtitles before they appear, so these are errors of more than aesthetic importance.

Segmentation errors are when the semantic or syntactic structure of the message is not respected (cf. Karamitroglou 1998 on segmenting at the highest syntactic node). This applies to what Gottlieb (2012: 41) calls macro- as well as micro-segmentation, i.e. between subtitles (standard error) and between the lines of subtitles (minor error). The errors of segmentation between subtitles are counted as more serious. As one subtitler put it regarding the importance of good synchronisation in subtitling "Good synchronisation = good flow = reading comfort = subs becoming nearly invisible = happy viewer" (van Turnhout, personal communication). This quote also illustrates the fluency ideal that is a prerequisite for the contract of illusion.

Serious errors are only to do with spotting and not segmentation, and a serious spotting error would be when subtitles are out of synch by more than one utterance. A minor spotting error would be less than a second off, and a standard error in between these two extremes.

5.3.2. Punctuation and graphics

It may seem nit-picking to have a subcategory of its own for punctuation, but the fact is that punctuation in subtitling is more important than in other texts – *Eats, shoots and leaves* aside (Truss 2003). The 'irrealis' use of italics is a good example: Italics are used in many countries to mark a voice or text that is 'not there:' voices on the phone, on TV, on PA systems, in dreams, in people's heads, in flashbacks, in hallucinations etc. In many places, this has become standard use and thus part of the contract of illusion, and the erroneous use of it should be considered a standard error. The same goes for the use of dashes. There is much variation in use of dashes. They are used for speaker indication, for continuation of utterances between subtitles and (rarer) for indicating the speaker's addressing a different person. Many of these usages are arbitrary: for example, in Danish television norms, there is a blank space after the "speaker dash" whereas there is not one in Swedish television norms (Pedersen 2007: 86). Similarly, some practices involve a speaker dash for each speaker in a dialogue subtitle, whereas some have it only for speaker number two. There is a hard rule in Swedish subtitles that each speaker must always have her or his own line, whereas in Finland Swedish subtitles (which are often bilingual, with one line in Finnish and one in Swedish) this is sometimes permitted (Pedersen 2007: 87). What decides how severe these errors are depends on which guidelines are used to feed the model (in that some allow variation) and consistency of use.

5.3.3. Reading speed and line length

The length of a subtitle line varies a great deal between media and systems. Also, it matters if the system that is used for viewing subtitles is character or pixel based. This is, however, something which is always regulated in

guidelines, normally in characters, so it is something that is easy and automatic to measure. The reason for not having too long lines is that these get slashed (so that the end is not shown), halved (so that there can be more than two lines) or represented in a smaller font (which reduces legibility), depending on the software.

Reading speeds in subtitling is also a varied and often contested issue. In reading research (cf. e.g. Schotter and Rayner 2012), speed is often measured in words per minute (wpm), and this is also the case for the NER model (cf. Romero Fresco & Martinez 2015: 47). However, in interlingual subtitling, the preferred measure is characters per second (cps), which brings with it an issue of conversion. Word length is also language-specific. For instance, in English, five characters per word is considered an average (Romero Fresco 2011: 113), whereas in Swedish, this is closer to six (cf. e.g. www.larare.at). That is but a trifling issue, however, compared with the many variations one finds in people's reading speed and in the pace at which dialogue is set. A further issue is how complicated the syntax and lexis is. For instance, Moran (2012: 183) has proven that high frequency words can be read at a significantly higher pace than low-frequency ones. Much more can be (and has been) said about this, but there is not enough space here. Suffice it to say that even though the long-held standard of 12 cps (which was proved to be an appropriate setting by d'Ydewalle *et al* in 1987 using eye-tracking) for interlingual subtitles for television is now under attack, and reading speeds of 15, or even 17 for some genres of streamed television are now used (Nilsson *et al*: personal communication). Recent research shows that this is not necessarily for the benefit of the viewers. According to Romero Fresco's eye tracking studies (2015: 338), if viewers are to spend less than half of their attention on the subtitles, the 12 cps rule should be followed, and this tallies with Jensema's (1997) findings as well. The time spent reading subtitles increases with reading speeds, so already at 15 cps, viewers spend on average about two thirds of their time in the subtitle area and at 16.5 cps, they spend 80% of their time reading subtitles. That may not be so much of a problem for relatively static genres such as interviews or news, but for a feature film, that leaves very little attention for the on-screen action. Or rather, the viewer has to choose between the subtitles and the action, and there is eye-tracking evidence (Caffrey 2012: 254) that viewers give some subtitles a miss, or stop reading in mid-subtitle (Lång *et al* 2013: 80). There is some evidence that too slow subtitles also draw the viewers' gaze (Lång *et al* 2013: 79), but that is a very rare error. It is, however, too intricate to devise a model that takes such things as individual reading speed, genre, text complexity and so on into account, so for practical purposes, this measure should also be fed norms from guidelines or national conventions. When no such norms exist, I suggest penalising anything higher than 15 cps, and increasingly so up to a level of 20 cps (or 240 wpm), which is a level where most people would probably do nothing else but reading subtitles (or stop using them). Thus,

20 cps could be considered a standard error, unless the norms tell you otherwise.

6. Discussion and conclusion

When having extracted the errors from the subtitles using the FAR model, a score is calculated in the three areas. In the first, Functional equivalence is calculated. The penalty points here are higher (for semantic errors) than in other areas, due to this arguably affecting the viewers' comprehension and ability to follow the plot the most. In the second, acceptability rate is calculated by adding up the penalty points for grammar, spelling and idiomaticity. In the third, the readability is calculated by summarising the errors of spotting and segmentation, punctuation and reading speed and line length. The next step is to divide the penalty score by the number of subtitles, and you then get a score for each area, which tells you to what degree a subtitled translation is acceptable, readable and/or functionally equivalent. By adding the penalty scores before you divide by the number of subtitles, the total score of the subtitles is calculated. This total score is unfortunately not immediately comparable to e.g. scores from the NER model, and probably much lower than the percentage scores from that model, as those are based on smaller units. If necessary, it can be compared to word-based scores by doing a word count of the subtitles, or by multiplying the number of subtitles by an average words per subtitle score (in Sweden that would be 7, according to an investigation carried out by the present author). This is really not recommended, however, as many errors affect more than one word.

The FAR model may seem time-consuming and complicated, but it need not be. It can be applied to whole films and TV programmes or just extracts and some of the data can be extracted automatically (line lengths and reading speeds will be provided by subtitling software). The greatest advantage of the model is that it gives you individual scores for the three areas, and that can be useful as subtitler feedback and as a didactic tool. Another advantage is that it can be localised using norms from guidelines and best practice, which means that it is pliable. This is important, as it would be deeply unfair to judge subtitles made under a certain set of conditions by norms that apply to different conditions, for instance by judging abusive (cf. Nornes 2009) or creative subtitles (cf. McClarty 2015) using norms of mainstream commercial subtitling.

There are several weaknesses in the model. One weakness is that it is based on error analysis, which means that it does not reward excellent solutions. Another is that it has a strong fluency bias, as it is based on the contract of illusion. The greatest weakness is probably subjectivity when it comes to judging equivalence and idiomaticity errors (which is also the case for models investigating SDH quality; cf. Ofcom 2014: 26). There is also a degree of fuzziness when it comes to judging the severity of the errors and

assigning them a numerical penalty score. It is hard to see how that could be remedied, however, given the nature of language and translation.

Bibliography

Informants
- **Nilsson, Patrik, Norberg, Johan, Björk, Jenny & Kyrö, Anna**. Managers and quality editors at SDI-Media. Sweden. Interviewed on June 2, 2015.

- **Sánchez, Diana**. Head of Operations Europe, Access Services, Broadcast and Media Services at Ericsson. Interviewed on April 6, 2016.

- **Van Turnhout, Guillaume**, subtitler at VRT. Belgium. Conversation on Twitter April 2, 2016.

Other references
- **Ben Slamia, Fatma** (2015). "Pragmatic and Semantic Errors of Illocutionary Force Subtitling." *Arab World English Journal* (AWEJ) Special Issue on Translation, 4, May, 42–52.

- **Burchardt, Aljoscha & Arle Lommel** (2014). "Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality." http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf. (consulted 19.05.2016).

- **Caffrey, Colm** (2012)."Using an eye-tracking tool to measure the effects of experimental subtitling procedures on viewer perception of subtitles AV content." In Perego, Elisa (ed.). *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 223–258.

- **De Linde, Zoé & Neil Kay** (1999). *The Semiotics of Subtitling*. Manchester: St Jerome.

- **Denison, Rayna** (2011). "Anime Fandom and the Liminal Spaces between Fan Creativity and Piracy." *International Journal of Cultural Studies*. 14(5), 449–466.

- **D'Ydewalle, Géry, Johan van Rensbergen and Joris Pollet** (1987). "Reading a Message When the Same Message is Available Auditorily in Another Language: The Case of Subtitling." In John Kevin O'Reagan and Ariane Lévy-Schoen (eds) (1987). *Eye Movements: From Physiology to Cognition.* Amsterdam & New York: Elsevier Science, 13–321.

- **D'Ydewalle, Gery, Luc Warslop and Johan van Rensbergen** (1989). "Differences between Young and Older Adults in the Division of Attention over Different Sources of TV Information." In *Medienpsychologie* 1, 42–57.

- **D'Ydewalle, Géry, and Ingrid Gielen** (1992). "Attention Allocation with Overlapping Sound, Image, and Text." In Rayner, Keith (ed.) (1992). *Eye Movements and Visual Cognition*. New York: Springer, 415–427.

- **Gellerstam, Martin** (1986). "Translationese in Swedish Novels translated from English". In Wollin, Lars & Hans Lindquist (eds). *Translation Studies in Scandinavia. Proceedings From the Scandinavian Symposium on Translation Theory (SSOTT) II. Lund 14 - 15 June, 1985*. (Lund Studies in English 75). Malmö: Liber/Gleerup, 88–95.

- **Ghia, Elisa** (2012). "The impact of translation strategies on subtitles reading". In Perego, Elisa (ed.). *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 157–182.

- **Gottlieb, Henrik** (1997). *Subtitles, Translation & Idioms.* Copenhagen: Center for Translation Studies, University of Copenhagen.

- (2001). *Screen Translation: Six studies in subtitling, dubbing and voice-over.* Copenhagen: Center for Translation Studies, University of Copenhagen.
- (2012). "Subtitles – Readable dialogue?" In Perego, Elisa (ed.). *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 37–82.

- **Graham, John D** (1989). "Checking, revision and editing." *In*: Catriona Picken, ed. *The translator's handbook*. London: Aslib, 59–70.

- **House, Juliane** (1981). *A Model for Translation Quality Assessment.* Tübingen: Gunter   Narr.

- **ISO 17100** *Quality Standard – Requirements for Translation Services*.

- **Jensema, Carl** (1997). "Viewer Reaction to Different Captioned Television Speeds". Institute for Disabilities Research and Training. https://www.dcmp.org/caai/nadh30.pdf. (consulted 19.05.2016).

- **Karamitroglou, Fotios** (1998). "A Proposed Set of Subtitling Standards for Europe". *The Translation Journal* 2 (2). http://translationjournal.net/journal/04stndrd.htm. (consulted 08.05.2016).

- **Lindberg, Ib** (1989). *Nogle regler om TV-teksting* [A few rules about TV subtitling].

- **Lång, Juha, Jukka Mäkisalo, Tersia Gowases & Sami Pietinen** (2013). "Using eye tracking to study the effect of badly synchronized subtitles on the gazer paths of television viewers." *New Voices in Translation Studies* 10, 72 – 86.

- **McClarty, Rebecca** (2015). "In support of creative subtitling: contemporary context and theoretical framework." *Perspectives – Studies in Translatology*. 22:4, 592–606.

- **Moran, Siobhan** (2012). "The effect of linguistic variation on subtitle reception." In Perego, Elisa (ed.). *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 183–222.

- **Nornes, Abé Mark** (1999). "For an abusive subtitling". In *Film Quarterly*, vol 52: 3, 17 – 34.

- **O'Brien, Sharon** (2012). "Towards a dynamic quality evaluation model for translation". In *Journal of Specialized Translation* 17. http://www.jostrans.org/issue17/art_obrien.php (consulted 19.05.2016).

- **Ofcom** (2014). Measuring live subtitling quality: Results from the first sampling exercise. http://stakeholders.ofcom.org.uk/binaries/research/tv-research/1529007/sampling-report.pdf (consulted 19.05.2016)

- (2015). Measuring live subtitling quality: Results from the fourth sampling exercise. http://stakeholders.ofcom.org.uk/binaries/research/tv-research/1529007/QoS_4th_Report.pdf (consulted 19.05.2016)

- **Pedersen, Jan** (2007). *Scandinavian Subtitles: A Comparative Study of Subtitling Norms in Sweden and Denmark with a Focus on Extralinguistic Cultural References* Doctoral thesis. Department of English, Stockholm University.

- — (2008). "High Felicity: a speech act approach to quality assessment in subtitling". Chiaro, Delia, Christina Heiss, & Chiara Bucaria (eds). *Updating Research in Screen Translation.* Amsterdam/Philadelphia: John Benjamins, 101–116.

- — (2011a). *Subtitling norms for television: an exploration focusing on extralinguistic cultural references* Amsterdam/Philadelphia: John Benjamins.

- — (In preparation). "Swedish fansubs – what are they good for?"[working title]

- **Robert, Isabelle S. and Aline Remael** (2016). "Quality control in the subtitling industry: an exploratory survey study." *Meta* 61(3), 578-605.

- **Romero Fresco, Pablo** (2009). "Naturalness in the Spanish Dubbing Language: A case of not-so-close *Friends.*" *Meta* 54(1), 49–72.

- — (2011). *Subtitling through speech recognition: respeaking.* Manchester: St Jerome.

- — (2012). "Quality in Live Subtitling: The Reception of Respoken Subtitles in the UK." Remael, Aline, Pilar Orero & Mary Carroll (2012). *Audiovisual Translation and Media Accessibility at the Crossroads: Media for All 3*. Amsterdam & New York: Rodopi, 25–41.

- — (2015). "Final thoughts: viewing speeds in subtitling." Romero Fresco, Pablo (ed.). (2015) *The Reception of Subtitling for the Deaf and Hard of Hearing in Europe*. Bern: Peter Lang

- **Romero-Fresco, Pablo & Juan Martínez** (2015). "Accuracy Rate in Live Subtitling: The NER model." Jorge Díaz Cintas and Rocío Baños (eds) *Audiovisual Translation in a Global Context: Mapping an Ever-changing Landscape.* Basingstoke & New York: Palgrave Macmillan, 28–50.

- **Schotter, Elisabeth R. and Keith Rayner** (2012). "Eye movements in reading. Implications for reading subtitles." Perego, Elisa (ed.) (2012) *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 83–104.

- **SDI Media** (No year). "GTS Quality Specifications – Subtitling." In-house material.

- **SDL** (no year). "LISA QA Metric'," http://producthelp.sdl.com/SDL_TMS_2011/en/Creating_and_Maintaining_Organizations/Managing_QA_Models/LISA_QA_Model.htm (consulted 09.05.2016).

- **Temizöz, Özlem** (Forthcoming). "Postediting machine translation output: Subject-matter experts versus professional translators." *Perspectives: Studies in Translatology.*

- **Toury, Gideon** (1995). *Descriptive Translation Studies – And Beyond*. Amsterdam & Philadelphia: John Benjamins.

- **Truss, Lynne** (2003). *Eats, Shoots & Leaves: The Zero Tolerance Approach to Punctuation.* London: Profile.

- **Tveit, Jan Emil** (2004). *Translating for Television: A Handbook in Screen Translation*. Bergen: Kolofon.

- **Van Leuven-Zwart, Kitty M** (1989). "Translation and original: Similarities and dissimilarities." *Target* 1.2, 151–81.

- **Vermeer, Hans J** (1989/2000). "Skopos and commission in translational action". In Venuti, Lawrence. 2000. *The Translation Studies Reader,* London & New York: Routledge, 221–231.

- **Vinay, Jean-Paul & Jean Darbelnet** (1958/2000). "A methodology for translation." Translated by Sager, Juan C. & Hamel M.-J. In: Venuti, Lawrence. *The Translation Studies Reader*, London & New York: Routledge, 84–93.

- **Wissmath, Bartholomäus and David Weibel** (2012). "Translating movies and the sensation of being there." Perego, Elisa (ed.). *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 277–293.

- **Volk, Martin and Harder Søren** (2007). "Evaluating MT with Translations or Translators. What Is the Difference?" *Proceedings of MT Summit*. Copenhagen. http://www.zora.uzh.ch/20406/2/Volk_Harder_2007V.pdf (consulted 15.03.17).

Websites
- **ESIST**, www.esist.org (consulted 19.05.2016)

- **SDL**, http://www.sdl.com/ (consulted 19.05.2016)

- **Språkstatistik**,   http://larare.at/pedagogik/sprakstatistik.html   (consulted 19.05.2016)

- **Stockholm   University,   Tolk-   och   översättarinstitutet**   (TÖI) http://www.tolk.su.se/ (consulted 19.05.2016)

Biography

Dr. Jan Pedersen is Associate Professor and Director of the Institute for Interpreting and Translation Studies at the Department of Swedish and Multilingualism at Stockholm University, Sweden, where he researches and teaches audiovisual translation. He has worked as a subtitler for many years and is the president of ESIST, the European Association for Studies in Screen Translation, and Associate Editor of *Perspectives: Studies in Translatology*.

Email: jan.pedersen@su.se

[1] GTS (Global Titling System) is SDI Media's in-house subtitling software.

[2] I am grateful to Dr. Romero Fresco of the University of Roehampton for this, as well as other comments that have improved the quality of this paper on quality.

[3] Due to space restrictions, examples will only be given for the most complex errors; those of functional equivalence. More examples will be found in Pedersen (in preparation).

[4] The examples will, in this paper, be given with the source text (ST) dialogue first, followed by the target text (TT) subtitles and a back translation (BT), and referenced by name of film and time of utterance).

[5] There is a minor spelling error here as well, as novel is spelt with two Ls in Swedish.