

## **A functional approach to translation quality assessment: Categorizing sources of translational distortion in medical abstracts**

**Hanna Martikainen**

Université Paris Diderot, Sorbonne Paris Cité, France  
hmarti@eila.univ-paris-diderot.fr

*The translation of Cochrane Systematic Review abstracts plays an important role in ensuring the communication of medical research results to the non-English-speaking public. The presentation of results should, of course, be free of errors and as objective as possible. Translation, however, is a highly subjective activity requiring extensive interpretation and is prone to errors. Translated Cochrane abstracts therefore contain elements that could affect readers' interpretation of the results of a Review, and that specifically could have an impact on the effectiveness of the intervention studied. Guided by the function of these translated texts, we have categorized such sources of distortion into a context-specific typology that will be used to measure translational distortion in Cochrane abstracts. Building on previous research and empirical corpus analysis, the typology accounts not only for translation errors with considerable potential impact, but also for biased translations of phraseological and modal structures that can markedly affect readers' interpretation of the degree of certainty expressed by authors.*

### **1. Introduction**

#### **1.1. Project background**

Cochrane is an independent international not-for-profit organization that aims to provide consumers – health professionals and service users alike – with impartial and up-to-date information on the effectiveness of healthcare interventions in order to help them make informed choices. Cochrane's main activity is the production of Systematic Reviews on the evaluation of healthcare interventions, which are published online in the Cochrane Library. A Systematic Review (SR) consists in identifying, assessing and synthesizing all existing studies on a given intervention with the aim of answering a precise question: On the basis of available evidence, what are the benefits and risks of the intervention? Often considered the gold standard in Evidence-Based Medicine, Cochrane SRs are influential in shaping medical decision-making.

Each Cochrane SR is accompanied by a scientific abstract and a plain language summary (PLS) both of which are freely available online and translated into different languages by regional Cochrane centres. In this study we have chosen to focus on the scientific abstract, which is aimed at practitioners and medical decision-makers. In the medical domain, the abstract could be considered “probably the most important section of a research article” (Yavchitz, 2015, p. 24, our translation). Indeed, it has been shown that medical decisions are sometimes made on the strength of the abstract alone (Yavchitz, 2015, p. 25), and the distorted reporting of results in the abstract is therefore of particular concern.

It should be noted, however, that Cochrane SR abstracts represent a distinct text type, one quite different from the much more studied research article abstract. Written as standalone documents, SR abstracts are structured in clearly defined sections<sup>1</sup> that summarize essential elements of the Review, for example, its objectives, key methods used for data collection and analysis and the main results of the statistical comparisons on predefined outcomes. The stated purpose is to present information, and for this reason Cochrane authors are specifically instructed to avoid making recommendations in their conclusions.<sup>2</sup> The text type of SR abstracts could then be defined, according to Reiss's (1981) classification, as purely informative, where the objective is the communication of content without any persuasive character. They respond to the ideal of an

objective presentation of results in the medical domain, of “let[ting] the results speak for themselves” (Cummings & Rivara, 2012, p.1099). Indeed, it could be argued that Cochrane authors have no need for persuasive rhetoric, since they are guarded against any potential conflicts of interest by the rigorous production process of SRs.

In order to provide access to high-quality medical information to a potential worldwide audience of over 220 million French-speakers,<sup>3</sup> Cochrane France has been translating Cochrane SR abstracts and PLS into French since 2010, and approximately 5,000 French translations are available today. These translations cater to a genuine demand from the French-language readers, as demonstrated by their 160,000 monthly views on average on *cochrane.org* in 2016 (Cochrane Translation Annual Report, 2016, p. 6).

Translating highly specialized texts is complex, time-consuming and costly, as it involves using specialized translators and target-language content experts for proofreading. The translation volumes are high, with up to 80 new and updated SRs being published every month in English. To tackle these important needs with limited resources, Cochrane France opted in 2013 for a solution based on statistical machine translation (MT). MT has proven to be one useful CAT tool among others, accelerating the translation process and thereby reducing costs (Koponen, 2016, p. 10). Highly structured and standardized, Cochrane SR abstracts are well suited to MT, which benefits from the existence of medical databases and of a human-translated corpus of abstracts for training the engine.

Publishing-quality translation of such highly specialized texts requires careful post-editing (PE) to ensure the effective diffusion of medical knowledge. Cochrane post-editing guidelines, however, correspond more to what is called ‘light’ or ‘fast’ PE, the objective of which is an accurate translation of essential content in which stylistic considerations are secondary. The expected level of quality is “good enough” or “fit for purpose”, which the Translation Automation User Society (TAUS) defines as follows: “The text may sound like it was generated by a computer, syntax might be somewhat unusual, grammar may not be perfect, but the message is accurate” (TAUS MT Post-Editing Guidelines, 2010, p. 17). Such guidelines are commonly adopted alongside the introduction of MT into the translation process in order to maintain the benefits of rapidity by reusing as much of the MT output as possible. Also, Cochrane’s community approach to content production is reflected in the translation process: post-editing at Cochrane France is partly done by volunteers, many of whom come from medical backgrounds without any linguistic training.

Our previous work (Martikainen & Kübler, 2016) suggests that this new mode of production of Cochrane SR abstract and PLS translations seems acceptable to end-users and could even offer the target audience some advantages regarding clarity; but the influence of MT may also give rise to specific errors that could influence readers’ interpretation of the message.

## 1.2. Objectives

The function of these translated texts can be defined as the effective communication of medical knowledge to help informed decision-making and transfer of research results into clinical practice. Accurate interpretation of the results requires that they be presented in a clear, complete and transparent manner (Yavchitz, 2015, p. 36). We define sources of translational distortion, then, as elements in the translation that can interfere with readers’ interpretation of research results or of the authors’ degree of confidence in them. Two kinds of element were deemed to have a potential impact on the accurate interpretation of translated Cochrane SR abstracts, namely (1) certain translation errors and (2) biased translations, especially of modal expressions. Example 1 illustrates an error encountered in texts produced by human translation and post-edited MT alike – the addition or omission of negation – that obviously has a direct bearing on the accurate interpretation of results.

- (1) The yoga group showed significant improvement in the Satisfaction With Life Scale (...)  
*Le groupe pratiquant le yoga n'a révélé aucune amélioration significative de l'échelle SWLS (...)*  
 = [The group practising yoga revealed *no* significant improvement of the SWLS scale]  
 (Source: Cochrane Error Corpus)

Besides certain translation errors, biased<sup>4</sup> translations can also be sources of distorted interpretation for the reader. For instance, modal markers such as 'slightly', 'probably' or 'may' are used in these texts to denote different levels of proof and authors' confidence in the results (Glenton et al., 2010, p. 572). The translation of modal markers poses a problem in conveying signals of mitigated claims (Pilegaard, 1997, p. 176), among others. For the modal auxiliary verb 'may', translations observed in the Cochrane Abstracts Corpus are mainly divided between the indicative and the conditional moods of the verb '*pouvoir*'. The indicative mood ('*peut*') is naturally more affirmative than the conditional mood ('*pourrait*'), and can potentially be interpreted as expressing actual rather than hypothetical capability (equivalent of 'can'). In some contexts, it is therefore likely to bias readers' interpretation of results positively, potentially inducing more certainty than was intended by the authors (Example 2(b)).

- (2) (a) In patients with cardiovascular disease, influenza vaccination *may* reduce cardiovascular mortality (...)  
*Chez les patients atteints de maladie cardiovasculaire, la vaccination antigrippale pourrait réduire la mortalité cardiovasculaire (...)*  
 = [In patients with cardiovascular disease, anti-influenza vaccination *MAY/MIGHT* reduce cardiovascular mortality]  
 vs  
 (b) In septic adults it *may* reduce mortality (...)  
*Chez les adultes septiques, elle peut réduire la mortalité (...)*  
 = [In septic adults, it *MAY/CAN* reduce mortality]  
 (Source: Cochrane Abstracts Corpus)

The objective of this article is to categorize these potential sources of distortion in translated Cochrane SR abstracts with the aim of subsequently evaluating their frequency and type distribution in the corpus and, finally, measuring their impact. Our typology builds, first, on existing research on the two aspects deemed to be essential in terms of translational distortion: translation error assessment and the expression of modality.

### 1.3. Evaluating translation quality through error assessment

The error typology approach developed in the contemporary period in response to the need for an objective and generalizable method for evaluating pragmatic, professional translation. The aim of this approach is to provide systematic and consistent classification models for human evaluators to rely on in order to reduce the inherently subjective nature of the evaluation process (Secară, 2005, p. 39). Error typologies are the preferred method of evaluating translation as a product in the translation industry today (O'Brien, 2012, p. 74), and they share some common characteristics.

#### 1.3.1 Common characteristics of error typologies

The typologies used today in the professional sector list varying numbers of error categories – up to the 675 possible error types for the Sical scale developed by the Canadian Government's Translation Bureau (Secară, 2005, p. 40). Generally, a fundamental distinction is made between language-related and content-related errors (see, for instance, Castagnoli, Ciobanu, Kunz, Kübler, & Volanschi, 2011). This is the case, for instance, in the Sical model, which distinguishes between transfer and

language errors. The ATA certification scheme also distinguishes “mechanical” errors that do not affect understanding or the usefulness of target text (TT) from potentially more serious “translation/strategy/transfer” errors (Koby & Champe, 2013, p. 166).

There is also usually some kind of gradation of error severity, which is used for setting a pass/fail threshold (O’Brien, 2012, p.58). The Sical model, for instance, distinguishes between minor and major errors, the latter being those that count for judging the acceptability of a translation (Secară, 2005, p. 40). In the ATA grading scale, mechanical errors can be assigned a score between 0 and 4 points, whereas transfer errors can count up to 16 points (Koby & Champe, 2013, p. 166). Despite differences in grading scales, there is general agreement on the definition of minor and major errors. In most evaluation models, minor errors are defined as those having little or no negative impact on meaning, nor do they have the potential for misleading the user (O’Brien, 2012, p. 62). On the contrary, major errors, in some instances termed “critical” or “serious”, are considered to have a negative impact on meaning and may have major effects on “product usability, company liability, consumer health, safety and behaviour” (O’Brien, 2012, p. 62). These errors can therefore result in “death, human injury, damage to equipment, product recall or wasted repair and operation time” or, at a more general level, “embarrassment to a client” (Secară, 2005, p. 41).

In MT evaluation, when quality assessment is not performed through automated measures such as BLEU and HTER, the error typology approach is frequently adopted. The objective of error analysis is generally to improve the MT engine and categories are therefore kept to a minimum so as to detect recurring errors, for instance: lexical, morphological, syntactic, semantic, formatting, other (Wisniewski, Kübler, & Yvon, 2012). Errors are often graded according to the post-editing effort needed to correct them, since interest in MT is largely dependent on its potential for saving time and money in comparison to human translation (Koponen, 2016, p. 21). In MT evaluation, the distinction between minor and major errors – or between mechanical and transfer errors – is also frequently based on the cognitive effort required of the post-editor to correct the error, for instance (in descending order of effort): mistranslation, omission or addition, syntax, word form, punctuation (Lacruz, Denkowski, & Lavie, 2014, p. 78). Finally, at its simplest, MT quality assessment is performed by attributing a grade for the parameters of adequacy and fluency, generally on a five-point scale (O’Brien, 2012, p. 66).

### 1.3.2 Function of the translated text in quality assessment

Ever since the 1959 FIT Congress that stressed the importance of considering the purpose of the translation rather than the abstract and vague criteria used up to that point (Martínez Melis & Hurtado Albir, 2001, p. 273), the function of the translated text has acquired growing importance in quality assessment, and it is widely acknowledged that “a translation (and hence the evaluation of its quality) in the 21<sup>st</sup> century should be adapted to the communicative context of use” (Depraetere, 2011, p. 2). Although central in Vermeer’s (1989) *skopos* theory and the approaches inspired by it, the notion of function is not clearly defined (House, 2015, p. 11): By what linguistic means is this function expressed in the source text (ST), and in the translated text? Most evaluation models do, however, take into account the function of the translated text at least implicitly, since it is the purpose of the text that determines the effect a translation error might have on the user experience and, therefore, on the degree of severity of the error (O’Brien, 2012, p.59).

Existing models have nonetheless been criticized for being too static, as evaluation consists in counting errors in random samples without considering client feedback, while the notion of quality is inherently linked to user opinion (O’Brien, 2012, p. 55). Alternative models have emerged that have in common a definition of quality as being context-dependent: there is no universal standard of quality, only translations that are more or less appropriate for their intended purpose (Secară, 2005, p. 39), i.e. “functionally appropriate” (Schäffner, 1997, p. 2). End-user needs are at the centre of this fit-for-purpose quality (see, for instance, O’Brien, 2012). In the functionalist approach, the linguistic format of the translated text is determined, above all,

by target language text-typological conventions, as well as by aspects of the communicative situation in the target culture in which the [translation] is to fulfil its function (e.g. time and place, knowledge and expectations of the [target text] addressees) (Schäffner, 1997, p. 2).

Initially developed in 1977 and later revised on two occasions (in 1997 and 2015), the complex model proposed by House (2015) seeks to capture the function of the translated text in all aspects of its situation. The objective is to detect mismatches in functional equivalence between the ST and the TT using different parameters derived from the various situational dimensions of language use. Evaluation based on House's model supposes a preliminary analysis of the functions of the ST and the lexical, syntactic and textual means used to express them, followed by a comparison with the TT using these parameters.

The function of the translated text is also central in Williams's (2001) approach, which applies argumentation theory to translation quality assessment with the objective of reconciling the quantitative and textological dimensions of other models. Williams proposes comparing ST and TT argument macrostructures as an efficient means of determining transfer adequacy and, therefore, translation quality. Williams's model, then, leads to a new definition of major error, which is the failure to render accurately in the TT the basic argument elements present in the ST, for instance, modalizers (Williams, 2001, p. 68).

#### **1.4. Expression of modality in specialized discourse**

Modality can be defined, in a broad sense, as speakers' positioning themselves in regard to the content of their proposition (Chuquet & Paillard, 1987, p. 108), or as a "gradable qualification" (Vihla, 1999, p. 19). Epistemic modality unfolds on the axis of uncertainty: the speaker evaluates the chances of realization (e.g. possible, probable, necessary) of the process in the extra-linguistic reality (Chuquet & Paillard, 1987, p. 109).

##### **1.4.1 Modality in scientific writing: hedging and boosting**

Modality occurs frequently in scientific writing, since it serves to construct the rhetorical strategies of hedging and boosting that enable writers to position themselves on the axis of uncertainty through the use of doubt and certainty markers. Hedging enables writers to distance themselves from a proposition. Hedges such as 'possible', 'might' and 'perhaps' represent a weakening of the claim for the purpose of showing doubt and indicating that the "information is presented as opinion rather than accredited fact" (Hyland, 1998, p. 2). Indeed, writers will carefully "consider the amount of certainty they should put in a particular statement according to the amount of reliable data backing this statement" (Vázquez Orta & Giner, 2008, p. 174). Hedging is mainly performed through epistemic and modal verbs, adverbs and adjectives (see, for instance, Kilicoglu & Bergler, 2008).

The main functions of boosters are related to persuading and convincing the audience (Vázquez Orta & Giner, 2009, p. 223). Boosters such as 'clearly', 'obviously' and 'highly' express conviction and confidence, marking involvement with an audience through shared information and group membership (Hyland, 1998, p. 2). These certainty markers allow writers to project "a credible image of authority, decisiveness, and conviction in their views" (Vázquez Orta & Giner, 2009, p. 221).

Hyland's (1998) work shows the importance of these strategies in academic discourse, with an average of 120 occurrences per paper. Hedges occur almost three times as often as boosters, which for Hyland reflects the importance, in scientific discourse, of distinguishing facts from opinions and presenting tentative suggestions rather than firm claims (Hyland, 1998, p. 7).

### 1.4.2 Modality in medical LSP

Medical writers “tend to make claims mainly in a tentative, reserved and objective way”, and “prefer to express uncertainty rather than certainty about their stated propositions” (Yang, Zheng, & Ge, 2015, p. 1-5). Hedges, then, are characteristic of medical discourse (Varttala, 1999, p. 177). The use of these markers of uncertainty is influenced, among other things, by the communicative purpose of the discourse and the level of claim the authors wish to make (Salager-Meyer, 1994, p. 1), as well as their degree of subjective certainty and the prospective reader (Pilegaard, 1997, p. 176). Also, the type of medical text influences the use of modality: modalization is significantly more frequent in review articles, which suggests that “the more universal the pretension of a claim, the more hedged the discourse” (Salager-Meyer, 1992, p. 93). It has also been argued that hedges are actually used in medical LSP to achieve a higher degree of precision by allowing authors to signal their genuine uncertainty regarding the results (see, for instance, Salager-Meyer, 1994, p. 1 and Vold, 2006, p. 81).

The use of boosters has also been evoked in the medical literature, in studies on “spin”, which is defined as a distorted presentation of research results through the use of conscious or unconscious strategies stressing positive results that can result in an interpretation bias (Yavchitz, 2015, p. 3). It has been shown that spin can influence physicians’ interpretation of research results and make them more susceptible to considering the intervention as beneficial in spite of non-significant results on the primary outcome of the study (Boutron et al., 2014, p. 6). Linguistic spin is “any word or expression stressing the beneficial effect of the intervention” (Lazarus et al., 2015, p. 5), such as ‘excellent’ or ‘encouraging’ results, ‘great potential’ or ‘a trend towards significance’ (Lazarus et al., 2015, p. 3). Some even recommend that authors, to avoid succumbing to this “literary sin” of “using words to convince, not illuminate”, use a minimum of adjectives to describe the topic and eliminate from their discourse any pretension to clarity (“our results clearly show”) or strength of conclusions (“this study shows that [X] is very effective”) (Cummings & Rivara, 2012, p. 1099). Linguistic spin has been observed in some 20 per cent (Boutron et al., 2014) to 25 per cent (Lazarus et al., 2015) of research article abstracts.

The interpretation of modal markers in medical LSP can be complex, as they can receive different and overlapping readings. While their epistemic and deontic uses relate, respectively, to the scientific and practical aspects of medicine (Vihla, 1999, p. 42), some markers (e.g. “can”) will also in some instances either express dynamic possibility (Vihla, 1999, p. 28) or combine epistemic and evidential readings (e.g. “may”) (Alonso-Almeida & Cruz-García, 2011, p. 70).

### 1.4.3 Inter-linguistic and cultural differences in expressing modality

French-language researchers have a reputation for being “much more prescriptive, authoritative and categorical than their English-speaking colleagues”, to the point of seeming “arrogant”, while their English-language peers have in turn been criticized for their “hyper-modesty” (Salager-Meyer, Ariza, & Zambrano, 2003, p. 10). It is therefore not surprising that English-language authors tend to use more epistemic modal markers of uncertainty, that is, hedges. In a comparative inter-linguistic and inter-disciplinary study, Vold (2006) establishes the relative frequency of hedges per 1,000 words at 3.7 and 2.1 respectively in English and in French. Different thresholds of appearance of modal markers can also be explained by the use of diverging modalization strategies (Chuquet & Paillard, 1987, p. 126). Modality can, indeed, be marked very differently from one language to the other (Guillemin-Flescher, 1984, p. 462). While auxiliary verbs are the privileged vehicle of modality in English, French tends to use other strategies, such as simple assertions (Chuquet & Paillard, 1987, p. 127).

Such inter-linguistic and intercultural differences need to be taken into account when translating modal structures. Genre- and culture-specific conventions regarding modality should guide translation choices (Pilegaard, 1997, p. 180). For instance, an assertive syntactic pattern in French (“*Il viendra demain*”) will often be translated by a non-assertive structure in English (“I believe he’s coming tomorrow”) (Guillemin-Flescher, 1984, p. 462). At the level of syntax, it is frequent to go from a simple sentence in English (“He may be there”) to a complex subordinated

sentence in French (“*Il se peut qu’il y soit*”). Also, the combination of declarative verbs indicating probability with modal verbs in English is particularly relevant to translation, given the importance of hedging and modalization in decoding the target language (Pilegaard, 1997, p. 178).

## 2. Methods

### 2.1. Analysis of error corpora

In order to define the categories for a context-specific typology based on actual sources of distortion found in these texts, we started out by analysing two existing corpora of translated Cochrane abstracts that had previously been reviewed for errors. The typology needs to cover errors typical to human and machine translation alike, which is why 50 translated abstracts produced by each of these processes were included in the Cochrane Error Corpus. For the first sub-corpus (HT), the abstracts were selected from a pool of 450 texts translated by professional translators and reviewed by Cochrane medical experts for evaluation purposes. The second sub-corpus (PE) contains machine-translated abstracts post-edited mainly by professional translators and in part by medical volunteers. Designed for evaluating the performances of the two MT engines in use at Cochrane France,<sup>5</sup> this sub-corpus includes 25 abstracts produced by each engine, reviewed for errors by a linguist.

We present here (Table 1) some results from this preliminary analysis that have guided the design of the error-related categories in the typology. The results are only indicative and cannot be used to compare the two sub-corpora or draw conclusions because of differences in text-selection and error-annotation methods, different annotator profiles and the lack of blinding. Errors taken into account are those likely to distort interpretation, namely those considered serious by the original annotators. Some categories used in the original error annotations are not reproduced here, since only minor errors were observed.

Table 1: Major errors by type in the Cochrane Error Corpus

Error Type	HT sub-corpus	PE sub-corpus
Grammar	7	14
Syntax	9	6
Terminology/lexis	48	18
Omission	6	19
Addition	7	1
Distortion <sup>6</sup>	n/a	14
TOTAL	77	73

The total number of major errors observed is very similar in the two error sub-corpora, approximately 1.5 per text, although their type distribution is more varied. The high number of terminology errors in the HT sub-corpus could be due in part to the original reviewers’ having been medical experts, and they were possibly more critical of the terminology used in translations than linguist error annotators. In the PE sub-corpus, the number of major grammar errors and omissions is relatively large compared to the HT sub-corpus, presumably due to the under-editing of MT output. These potential tendencies in error type distribution will be explored in a further corpus study using the typology drafted.

### 2.3. Corpus study on the translation of modal markers

A second corpus analysis was then conducted in order to determine how to account for biased translations of modal expressions in the typology. A list of relevant markers for study was first defined through a literature search in specialized databases<sup>7</sup> using domain keywords in English and in French. The most frequently quoted English markers were then searched for in a corpus of translated Cochrane SR abstracts and their most frequent translations listed. The Cochrane Abstracts Corpus is also divided into two sub-corpora produced by the two translation processes, each comprising 998 English abstracts and their French-language translations, for a total of 1,002,330 words in the STs and 1,213,559 in the translations. Searches were conducted using IMS Corpus Workbench.

Table 2 shows the results that guided the elaboration of the categories related to modality in the typology. The data, originating from automatic analysis, have not been sorted by type of modality (epistemic or deontic); in addition, the corpus is aligned at sentence level and translational correspondences have not been verified at word level. The examples presented are nonetheless representative of general tendencies observed in the corpus.

Table 2: French translations of “may” in the Cochrane Abstracts Corpus

French translation	Occurrences (%)
Verb <i>'pouvoir'</i> + indicative mood	805 (49)
Verb <i>'pouvoir'</i> + conditional mood	569 (35)
Other translations	260 (16)
TOTAL	1,634

Modal auxiliary verbs are frequent in the corpus and, unsurprisingly, “may” is the first among them with 1,634 occurrences. It is translated mainly by the indicative mood of the verb “*pouvoir*” (49 per cent), which we have defined as positively biased in certain contexts (see section 1.2), in comparison to the conditional mood, used in 35 per cent of the cases.

The positive bias is further reinforced when the conditional auxiliary “might” is rendered in the translation by the indicative form “*peut*”. This is used in 17 per cent of its occurrences and is illustrated in Example 3.

- (3) Antiplatelet therapy *might* reduce the volume of brain damaged by ischaemia and reduce the risk of early recurrent \*ischaemic stroke.

*Un traitement par antiagrégants plaquettaires peut réduire le volume cérébral endommagé par l'ischémie et réduire le risque de récurrence précoce 'd'AVC.*

= [Antiplatelet therapy *MAY/CAN* reduce the cerebral volume damaged by ischaemia and reduce the risk of early recurrent stroke.]

(Source: Cochrane Abstracts Corpus)

Other modal markers for which biased translations were frequently observed in the Cochrane Abstracts Corpus were evidential verbs, adjectives and adverbs. This first analysis thus yielded some indications of potential sources of modal bias in translated Cochrane SR abstracts but both their actual frequency and their respective role among sources of translational distortion will need to be established in a further corpus study using the typology drafted.

### 3. Results and Discussion

We have defined the function of these translations as the efficient diffusion of medical information through an accurate and non-biased presentation of research results, and sources of distortion as elements of the translation likely to interfere with this communicative objective (section 1.2). Accordingly, although two of the larger meta-categories in the typology (Figure 1) correspond to language errors commonly used in error typologies, our definition of error is guided by the function of these texts. Errors that lead to translational distortion are those that affect readers' interpretation of the characteristics of trials (e.g. interventions, participants, methods) or their results. Usual distinctions in error typologies have not been reproduced since they do not apply in this context. First, no distinction is made between language and content errors: the typology is designed to cover only errors that interfere with content transfer and, as such, may distort interpretation. These instances should nonetheless be attributable to precise linguistic items in the translated text (i.e. lexical, grammatical or relevant sub-categories of these two meta-categories). Similarly, the distinction between minor and major errors is not relevant here: we are interested only in the latter because of their potential for distorting essential content. Of course, the effect of a translation error must be assessed in context – errors that could be considered minor in some instances can naturally lead to distorted interpretation in other contexts.

Likewise, biased translations of phraseological and modal structures in the third meta-category are relevant only insofar as they are capable of affecting the readers' interpretation of results. Indeed, the more affirmative style of French authors is naturally likely to be visible in translations as well, and these shifts on the axis of uncertainty are also likely to be expected and appropriately interpreted in most cases by the readers. For instance, our corpus observations show that the modal auxiliary "should" is frequently translated into French as the indicative form of the verb "*devoir*" in the present or the future tense (the equivalent of "must"), instead of the corresponding conditional form. In most cases, however, such a translation does not lead to any relevant distortion, as this marker is typically used in the context of recommendations regarding future studies. Instances of translational distortion are occurrences where these more affirmative translations can potentially influence readers' interpretation of the authors' level of certainty, and therefore of the effectiveness of the intervention studied.

The expanding organization of the typology allows for different levels of annotation ranging from the three meta-categories (Lexis, Grammar, Lexicogrammar) to the twelve optional sub-categories (Figure 1) on the most specific possible level of annotation. Also, the impact of translation errors and of biased translations is encoded as an optional attribute in the typology (positive, negative, n/a). As we hypothesize that the less nuanced style of French scientific writing will also be perceptible in these translated texts, we expect to see a tendency towards positively biased translations of markers of uncertainty.

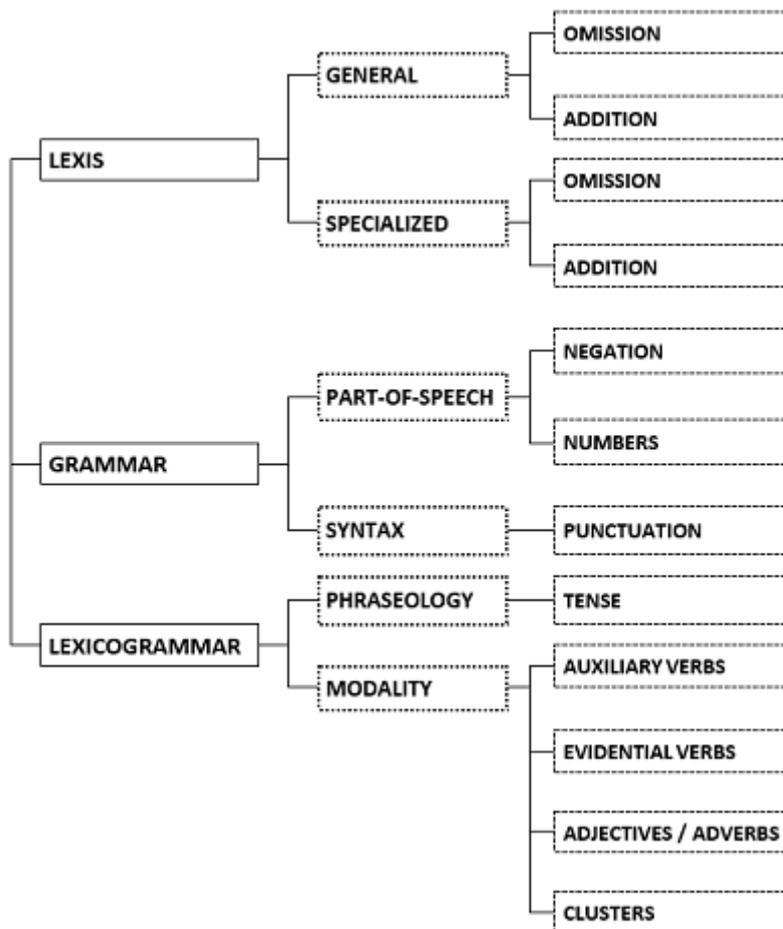


Figure 1: Typology of sources of translational distortion

Translation errors constitute random sources of distortion and are covered by the first two meta-categories of the typology. The meta-category of Lexis is divided into two sub-categories for distinguishing between errors of general vocabulary and those of specialized terminology. Both of these sub-categories are further divided into two additional sub-categories that cater for errors pertaining more specifically to the omission and addition of lexical items. In Example 4, the positive distortion due to the lexical error transforming six “weeks” of follow-up into six “months” is further reinforced by the omission of the adverb “only” from the translation.

- (4) One large study with *only* six weeks follow-up (...)  
*Une étude à grande échelle avec un suivi de six mois (...)*  
 = [One large-scale study with a follow-up of six *months*]  
 (Source: Cochrane Error Corpus)

It must be stressed that omission and addition can, of course, also be part of various translation strategies. As with other types of error, the omission and addition of lexical items will be taken into account only in instances where they lead to a distorted interpretation of the results. This is the case with the omission of one element of a composed term in Example 5.

- (5) (...) angiotensin-converting enzyme (ACE) *inhibitors* (...)  
 (...) *l'enzyme de conversion de l'angiotensine (ECA)* (...)  
 = [angiotensin-converting enzyme (ACE)]  
 (Source: Cochrane Abstracts Corpus)

The second meta-category, Grammar, is again divided into two sub-categories, the first of which covers errors pertaining to individual parts of speech (e.g. prepositions, verb inflection or the agreement of adjectives and participles). Within this category, errors related to the inversion of negation have been further distinguished in a sub-category of their own, because of their considerable impact on the accurate interpretation of results (see Example 1). Also, even though the rare occurrences of negation errors observed in the Cochrane Error Corpus were all encountered in the HT sub-corpus, the inversion of negation is typical of statistical MT and, as such, is expected to be more frequent in reality. Finally, errors related to numbers (although only a few were observed in the preliminary analysis of error corpora) are again likely to have a considerable impact on the interpretation of results, and have therefore also been allocated a sub-category of their own within Part-of-Speech errors.

The second sub-category of grammatical errors, Syntax, concerns errors pertaining to sentence, clause or phrase structure. We suspect that syntax errors are more frequent than what was observed in the preliminary analysis of error corpora, again partly due to the influence of statistical MT. In Example 6, for instance, the syntax error transforms a restrictive clause into a circumstantial modifier, as a result introducing a positive distortion in favour of the treatment.

- (6) This indicated significant recovery in motor function after methylprednisolone therapy, *when administration commenced* within eight hours of injury.  
*Elle révélait une amélioration significative de la fonction motrice après un traitement à base de méthylprednisolone, dès le début de son administration dans les huit heures suivant le traumatisme.*  
 = [It revealed a significant improvement of motor function after treatment with methylprednisolone, *from the beginning of its administration* within eight hours of the injury.]  
 (Source: Cochrane Error Corpus)

A further sub-category makes it possible to distinguish, within syntax errors, those pertaining specifically to punctuation. In Example 7, it is not the elliptical omission of the relative pronoun “that” but the addition of a comma that affects the sentence structure, again transforming a restrictive clause into a circumstantial modifier. As a result, a positive distortion in favour of the intervention is introduced.

- (7) There is a lack of evidence for treatment and prevention strategies *that are* safe for mothers and infants.  
*Il existe un manque de preuves pour les stratégies de traitement et de prévention, sans danger pour les mères et les nourrissons.*  
 = [There is a lack of evidence for treatment and prevention strategies, safe for mothers and infants.]  
 (Source: Cochrane Abstracts Corpus)

While the error-related categories in the typology deal with random instances of distortion, the third meta-category covers systematic sources of distortion, namely biased translations of lexico-grammatical structures. Gledhill and Kübler (2016) define lexico-grammatical patterns as “chains of meaningful interlocking lexical and grammatical structures” that are “the basic building-blocks” of LSP (Gledhill & Kübler, 2016: 75). For instance (Gledhill & Kübler, 2016: 78):

“FAILURE TO + V [comply with instructions] CAN/MAY + RESULT IN/LEAD TO + N [injury/death].”

The distortion here is of a systematic nature, since the tendency towards biased translation is rooted in the writing tradition of the target culture (e.g. frequent affirmative translation of the modal auxiliary “may” discussed in sections 1.2 and 2.3).

The meta-category Lexicogrammar is also divided into two sub-categories, the first of which concerns phraseological structures in general, while the second deals more specifically with patterns containing modal markers. Although the category Phraseology was included in the analysis of the Cochrane Error Corpus, no major errors had been observed (section 2.1). Because biased translations of phraseological structures are not errors per se, they would not have been detected by the analysis, in spite of their potential for distorting readers’ interpretation of results. The potential impact of phraseological choices on translation quality has been discussed previously (see, for instance, Gledhill, 2011). Moreover, expert reviewers’ comments in the HT error sub-corpus suggest that biased translations of certain phraseological structures common to scientific discourse are likely to affect the accurate interpretation of results. This is particularly the case with the presentation of inconclusive evidence, in which a common mistake is to confuse “no evidence of an effect” with “evidence of no effect”<sup>8</sup> (Higgins & Green, 2011, para. 12.7.4). Therefore, in the medical domain it is generally considered that “[i]t is never correct to claim that treatments have no effect or that there is no difference in the effects of treatments” (Alderson & Chalmers, 2003, p. 475).<sup>9</sup> That is why, for Cochrane medical experts, patterns such as

“[There was] + NEGATION + DIFFERENCE/EFFECT”

should be translated into French using the simple negation “*ne ... pas*” (Example 8(a)). In these contexts, the use of the reinforced negation “*ne ... aucun*” (Example 8(b)), although extremely frequent, is considered to be negatively biased.

- (8) (a) The intervention *had no effect* on CD4 percentage (...)  
*L’intervention n’avait pas d’effet sur le pourcentage de CD4 (...)*  
 = [The intervention *did not have an effect* on CD4 percentage]  
 vs  
 (b) [The intervention showed] *no effect* on CD4 count (...)  
 (...) *aucun effet n’a été constaté au niveau du nombre de CD4 (...)*  
 = [*no effect* was observed in terms of CD4 count]  
 (Source: Cochrane Abstracts Corpus)

Within phraseological structures, an optional sub-category distinguishes instances where the biased translation more specifically involves verb tense. These will typically be patterns such as

“[The intervention] + [REDUCE/INCREASE/IMPROVE + past tense] + [effect]”,

where a change to the present tense in the translation distorts the interpretation by conferring a value of general truth to the claim (Example 9).

- (9) Treatment with anticoagulants *reduced* recurrent stroke (...)  
*Le traitement par anticoagulants réduit la récurrence d’AVS (...)*  
 = [Anticoagulant treatment *reduces* recurrence of stroke]  
 (Source: Cochrane Abstracts Corpus)

A second sub-category within the meta-category Lexicogrammar makes it possible to distinguish patterns that include modal markers. Coming back to the example of “may”, approximately half of its occurrences in the Cochrane Abstracts Corpus are translated by the more affirmative solution (Table 2). As discussed, this translation can potentially lead to distortion, since the same form expresses the kind of actual capacity conveyed by “can” when used in a dynamic sense (see section 1.2). This potential for distortion is, however, realized only in contexts where the marker relates to treatment effectiveness and is associated with a desired effect. When the pattern

“[The intervention] + MAY + [reduce/increase/improve] + [effect]”

is associated with a negative consequence (e.g. treatment side-effect), the more affirmative translation is not interpretable as expressing an actual capacity to provoke a desired effect and therefore does not distort the interpretation. The category Modality is further divided into optional sub-categories that serve to distinguish the type of marker responsible for the distortion within the larger pattern. Besides modal auxiliaries, which have already been discussed, these include evidential verbs, which are understood to be reporting verbs that convey the source of knowledge (e.g. “show” and “demonstrate”), and adjectives and adverbs, which have been grouped together for ease of annotation. Finally, the sub-category of modal clusters has been added to account for instances where several markers within a pattern are translated in a biased manner (Example 10).

- (10) A \*single-centre trial in Sweden with 56 participants *showed that* excimer laser ablation in addition to mechanical debridement *may* reduce the number of erosions and improve symptoms.

*Un essai \*multicentrique réalisé en Suède et portant sur 56 participants a démontré que l'ablation au laser excimère en plus d'un débridement mécanique pouvait réduire le nombre d'érosions et améliorer les symptômes.*

= [A \*multicenter trial conducted in Sweden with 56 participants *demonstrated that* excimer laser ablation in addition to mechanical debridement *could* [CAN + past tense] reduce the number of erosions and improve symptoms.]

(Source: Cochrane Abstracts Corpus)

#### 4. Conclusion

We have argued that a functionally appropriate translation of medical abstracts requires careful consideration of various aspects, namely lexical, grammatical and lexico-grammatical. For the translation of lexico-grammatical patterns, in particular, it is necessary to balance target-language writing conventions against an accurate rendition of the level of proof. While we acknowledge that the categorical distinctions between sources of translational distortion remain arbitrary to some extent, and that translation evaluation is at its core always subjective, we hope that the typology drafted on the basis of empirical corpus research will help to achieve a more comprehensive assessment of what might constitute functional quality in the translation of medical LSP.

This operative typology will now be used to annotate a representative corpus sample in order to establish the frequency of translational distortion by type and by text section in translated Cochrane SR abstracts. The typology will also be used to compare the extent of distortion in texts produced by the two processes used for the French-language translation of Cochrane abstracts. The annotation phase will also be an opportunity to test and further adapt the typology. Finally, the effect of translational distortion in Cochrane abstracts will need to be confirmed in a qualitative study to be conducted among French-language end-users of these translated texts.

#### References

- Alderson, P., & Chalmers, I. (2003). Survey of claims of no effect in abstracts of Cochrane reviews. *British Medical Journal*, 326(7387), 475.
- Alonso-Almeida, F., & Cruz-García, L. (2011). The value of *may* as an evidential and epistemic marker in English medical abstracts. *Studia Anglica Posnaniensia*, 46(3), 59–73.
- Boutron, I., Altman, D. G., Hopewell, S., Vera-Badillo, F., Tannock, I., & Ravaud, P. (2014). Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: The SPIIN randomized controlled trial. *Journal of Clinical Oncology*, 32(36), 4120–4126.

- Castagnoli, S., Ciobanu, D., Kunz, K., Kübler, N., & Volanschi, A. (2011). Designing a learner translator corpus for training purposes. In N. Kübler (Ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice* (Vol. 12, pp. 221-248). Bern, Switzerland: Peter Lang.
- Chuquet, H., & Paillard, M. (1987). *Approche linguistique des problèmes de traduction anglais-français*. Paris, France: Ophrys.
- Cochrane Translation Annual Report (2016). Retrieved from <http://community.cochrane.org/>.
- Cummings, P., & Rivara, F. P. (2012). Spin and boasting in research articles. *Archives of Pediatrics & Adolescent Medicine*, 166(12), 1099–1100.
- Depraetere, I. (Ed.). (2011). *Perspectives on translation quality* (Vol. 9). Berlin, Germany: Walter de Gruyter.
- Gledhill, C. (2011). A lexicogrammar approach to checking quality: Looking at one or two cases of comparative translation. In I. Depraetere (Ed.), *Perspectives on translation quality* (Vol. 9, pp. 71–97). Berlin, Germany: Walter de Gruyter.
- Gledhill, C., & Kübler, N. (2016). What can linguistic approaches bring to English for Specific Purposes? *ASp. la revue du GERAS*, 69, 65–95.
- Glenton, C., Santesso, N., Rosenbaum, S., Nilsen, E. S., Rader, T., Ciapponi, A., & Dilkes, H. (2010). Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. *Medical Decision Making*, 30(5), 566–577.
- Guillemin-Flescher, J. (1981). *Syntaxe comparée du français et de l'anglais: Problèmes de traduction*. Paris, France: Ophrys.
- Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane Handbook for systematic reviews of interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Retrieved from <http://handbook-5-1.cochrane.org/>.
- House, J. (2015). *Translation quality assessment: Past and present*. London, England: Routledge.
- Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *Text-Interdisciplinary Journal for the Study of Discourse*, 18(3), 349–382.
- Kilicoglu, H., & Bergler, S. (2008). Recognizing speculative language in biomedical research articles: A linguistically motivated perspective. *BMC Bioinformatics*, 9(11), S10. doi:10.1186/1471-2105-9-S11-S10
- Koby, G. S., & Champe, G. G. (2013). Welcome to the real world: Professional-level translator certification. *Translation & Interpreting*, 5(1), 156–173.
- Koponen, M. (2016). *Machine translation post-editing and effort: Empirical studies on the post-editing process* (Doctoral dissertation). Retrieved from <https://helda.helsinki.fi/>. (Accession No. 160256)
- Lacruz, I., Denkowski, M., & Lavie, A. (2014). Cognitive demand and cognitive effort in post-editing. In S. O'Brien, M. Simard, & L. Specia (Eds.), *Proceedings of the Third Workshop on Post-editing Technology and Practice (WPTP-3), AMTA Workshop, Vancouver, Canada, October 26 2014*, 73-84. Retrieved from [https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings\\_PEWorkshop\\_final.pdf](https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf).
- Lazarus, C., Haneef, R., Ravaut, P., & Boutron, I. (2015). Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Medical Research Methodology*, 15(1), 85. doi: 10.1186/s12874-015-0079-x
- Martikainen, H., & Kübler, N. (2016). Ergonomie cognitive de la post-édition de traduction automatique: Enjeux pour la qualité des traductions. *ILCEA*, 27. Retrieved from <http://journals.openedition.org/ilcea/3863>.
- Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta: Journal des traducteurs – Meta: Translators' Journal*, 46(2), 272–287.
- O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *The Journal of Specialised Translation*, 17, 55–77.
- Pilegaard, M. (1997). Translation of medical research articles. In A. Trosborg (Ed.), *Text typology and translation* (pp. 159–184). Amsterdam, The Netherlands: John Benjamins.
- Reiss, K. (1981). Type, kind and individuality of text: Decision making in translation. *Poetics Today*, 2(4), 121–131.
- Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*, 11(2), 93–113.
- Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, 13(2), 149–170.
- Salager-Meyer, F., Ariza, M. Á. A., & Zambrano, N. (2003). The scimitar, the dagger and the glove: Intercultural differences in the rhetoric of criticism in Spanish, French and English medical discourse (1930–1995). *English for Specific Purposes*, 22(3), 223–247.

- Schäffner, C. (1997). From ‘good’ to ‘functionally appropriate’: Assessing translation quality. *Current Issues in Language & Society*, 4(1), 1–5.
- Secară, A. (2005). Translation evaluation: A state of the art survey. *Proceedings of the eCoLoRe-Mellange Workshop, Resources and tools for e-learning in translation and localisation, Leeds 21–23 March 2005*, 39–44. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.126.3654>.
- TAUS MT Post-Editing Guidelines (2010). Retrieved from <https://www.taus.net/>.
- Varttala, T. (1999). Remarks on the communicative functions of hedging in popular scientific and specialist research articles on medicine. *English for Specific Purposes*, 18(2), 177–200.
- Vázquez Orta, I., & Giner, D. (2008). Beyond mood and modality: Epistemic modality markers as hedges in research articles: A cross-disciplinary study. *Revista Alicantina de Estudios Ingleses*, 21, 171–190.
- Vázquez Orta, I., & Giner, D. (2009). Writing with conviction: The use of boosters in modelling persuasion in academic discourses. *Revista Alicantina de Estudios Ingleses*, 22, 219–237.
- Vihla, M. (1999). *Medical writing: Modality in focus*. Amsterdam, The Netherlands: Rodopi.
- Vold, E. T. (2006). Epistemic modality markers in research articles: A cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics*, 16(1), 61–87.
- Williams, M. (2001). The application of argumentation theory to translation quality assessment. *Meta: Journal des traducteurs – Meta: Translators’ Journal*, 46(2), 326–344.
- Wisniewski, G., Kübler, N., & Yvon, F. (2014). A corpus of machine translation errors extracted from translation students exercises. *Proceedings of LREC 2014*, 3585–3588. Retrieved from [http://www.lrec-conf.org/proceedings/lrec2014/pdf/1115\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/1115_Paper.pdf).
- Yang, A., Zheng, S. Y., & Ge, G. C. (2015). Epistemic modality in English-medium medical research articles: A systemic functional perspective. *English for Specific Purposes*, 38, 1–10.
- Yavchitz, A. (2015). *Communication des résultats scientifiques dans les essais randomisés contrôlés et les revues systématiques* (Doctoral dissertation). Retrieved from [www.theses.fr](http://www.theses.fr). (Accession No. 2015USPCC272)

- 
- 1 Cochrane SR abstract sections: Title, Background, Objectives, Search methods, Selection criteria, Data collection and analysis, Main results, Authors’ conclusions.
  - 2 Cochrane Handbook for Systematic Reviews of Interventions, 11.8. Writing an abstract: “The primary purpose of the review should be to present information, rather than to offer advice or recommendations. The Authors’ conclusions should be succinct and drawn directly from the findings of the review so that they directly and obviously reflect the main results. Assumptions should generally not be made about practice circumstances, values, preferences, tradeoffs; and the giving of advice or recommendations should generally be avoided.”
  - 3 Source: <https://www.francophonie.org/>.
  - 4 The term ‘bias’ is used in this article in the statistical sense of a systematic distortion in the presentation of results.
  - 5 MT engines used at Cochrane France: Google Translate and an in-house engine based on the open-source toolkit, Moses.
  - 6 The category ‘Distortion’ was specific to the original error annotation of the post-edited MT sub-corpus and therefore does not apply to the human-translated error sub-corpus.
  - 7 Databases used for the literature search: Persée, Research Gate, PubMed, Google Scholar.
  - 8 Cochrane Handbook for Systematic Reviews of Interventions, 12.7.4. Common errors in reaching conclusions: “A common mistake when there is inconclusive evidence is to confuse ‘no evidence of an effect’ with ‘evidence of no effect’. When there is inconclusive evidence, it is wrong to claim that it shows that an intervention has ‘no effect’ or is ‘no different’ from the control intervention.”
  - 9 Complete quotation from Alderson & Chalmers (2003, p. 475): “It is never correct to claim that treatments have no effect or that there is no difference in the effects of treatments. It is impossible to prove a negative or that two treatments have the same

---

effect. There will always be some uncertainty surrounding estimates of treatment effects, and a small difference can never be excluded.”