Discourse markers in English and European Portuguese translations: establishing functional equivalents and types of omission

*Marcadores discursivos em traduções em inglês e português europeu: definição de equivalentes funcionais e tipos de omissão**

Milana A. Morozova[**]
Universidade Nova de Lisboa, Lisboa, Portugal

FLP22(1)

Abstract: Based on the translations of a bidirectional English-Portuguese parallel corpus, this paper examines some English discourse markers (henceforth 'DMs', such as *well*, *you know*, *I mean*). The goal is twofold: firstly, the analysis of the translations establishes functional equivalents of the English DMs in European Portuguese, thus complementing the existing studies on translation of DMs in parallel corpus. Secondly and most importantly, this paper aims to approach the phenomenon of DMs omission frequently observed in translations from the empirical, rather than theoretical point of view. In particular, the study focuses on omission of DMs in the target languages. The corpus analysis resulted in the identification of three most common types of omission: DM deletion (i.e. a common DM deletion or omission in the target language), partial DM deletion (i.e. when one of the two DMs in the original language drops, resulting in translation of only one of them in the target language), DM addition (i.e. when there is no DM in the original language, but the translator has added it).

Keywords: Discourse markers. Omission. Translation. English. Portuguese.

Resumo: Tendo como base traduções de um corpus paralelo bidirecional inglês-português, este artigo visa a examinar alguns marcadores discursivos (daqui em adiante MDs) em inglês (tais como *bem*, *sabe*, *quer dizer*). O artigo tem dois objetivos. Primariamente, a análise das traduções estabelece equivalentes funcionais de MDs de inglês para português europeu, complementando, desta forma, os estudos existentes sobre traduções de MD em corpus paralelo. Por outro lado, e mais importante, este trabalho procura abordar o fenômeno de omissão de MDs frequentemente observado em traduções do ponto de vista empírico e não teórico. Em particular, o estudo focaliza a omissão dos marcadores discursivos em inglês e português. A análise do corpus resultou na identificação de três tipos mais comuns de omissão: eliminação de marcador discursivo (ou seja, uma exclusão ou omissão simples do marcador), eliminação parcial de marcador (ou seja, quando um dos dois marcadores foram omitidos na tradução, ficando apenas um deles) e adição de marcador (ou seja, quando não há marcador no idioma original, mas o tradutor o adicionou).

Palavras-chave: Marcadores discursivos. Omissão. Tradução. Inglês. Português.

[**] Doutorada em Linguística, Faculdade de Ciências Sociais e Humanas, Centro de Linguística da Universidade Nova de Lisboa, Portugal; milana1002@gmail.com

1 INTRODUCTION

The last three decades have seen a large amount of research on discourse markers (henceforth, DMs) in various theoretical frameworks (Schiffrin, 1987; Brinton, 1996; Fraser, 1999; Fischer, 1998, 1999; Östman, 2006; Blakemore, 2002). Nonetheless, the popularity of DMs is still growing, especially in a multi-language contrastive perspective. The recent emergence of parallel corpora has boosted research both in cross-cultural studies and translation studies (Dupont & Zufferey, 2017). A predominant majority of the research based on translations in parallel corpora attested a frequent omission of DMs in the target language (Aijmer, 2007; Aijmer & Simon-Vandenbergen, 2003; Cuenca, 2008; Ramón, 2015; Kafipour, 2016; Mattsson, 2006). However, little or no explanation has been provided to account for the phenomenon.

This paper pursues a twofold objective. On the one hand, it contributes to the existing studies on DMs in parallel corpora by focusing on translation equivalents of several English and Portuguese DMs. On the other hand, this paper provides a deeper qualitative analysis into the cases of the omitted DMs attested in the parallel corpus. The typification of omission of DMs elaborated here will demonstrate, however, that omission is not always about 'omitting' a DM; there may be other types of changes involved. The intention of this paper is, thus, to approach the phenomenon of omission and to provide a corpus-driven typification of omission in the given language pair.

The focus is on some DM in English such as *well, you know* and *I mean* and their functional equivalents in Portuguese. Their conversational nature, procedural meaning, pragmatic multi-functionality and, to some extent, unpredictability make them especially interesting for the analysis of omission in translations.

This paper is constructed as follows. Section 2 provides a brief overview of the two bipolar strategies frequently employed in translation: omission (or 'implicitation') and addition (or 'explicitation'). Section 3 focuses on omission of DMs in previous works and provides the existing explanations for this phenomenon. In Section 4, the methodology and corpus of the present research will be described. Section 5 presents some relevant research about Portuguese DMs. Section 6 features quantitative and qualitative data analysis; while the quantitative analysis provides numerical data on the number of occurrences and their back translations, the qualitative analysis dives into the established types of omission. Finally, Section 7 will round this paper by providing an overall discussion of the findings and by drawing some general conclusions.

2 COMMON STRATEGIES IN TRANSLATION

In the literature, two common bipolar strategies frequently employed by translators can be found: omission (or 'implicitation') or addition (or 'explicitation'). Regarding the first strategy, Asr and Demberg (2012, p. 2669) propose an information-theoretic perspective, according to which discourse connectives are more likely to be omitted when they are making a relation that is expected or predictable. The Uniform Information Based Theory (Ibid.) suggests that speakers communicate and choose the linguistic items available in the language according to the principle of uniformity of information transmission. In other words, speakers choose information and evenly distribute it in discourse for a successful and more optimized communication. The

connectives become 'optional' and can be easily omitted if they lead to information density, which is, according to the proposal of Asr and Demberg, undesirable. The information-theoretic perspective seems to echo back to a cognitive principle in relevance-theoretic terms, in other words "…the individual automatically aims at maximum relevance" (Sperber & Wilson 1986, p. 144). An input is relevant to an individual only when its processing yields positive cognitive effects. Similarly, if a discourse connective does not contribute to positive cognitive effects, it is considered irrelevant and, therefore, can be removed from the utterance. In this respect, omission or 'optionality' is similar to their "implicitation" in the translated discourse (Pym, 2005; Becher, 2010, 2011).

Besides omission, another common process is addition or 'explicitation'. Originally proposed by Blum-Kulka (1986), the "explicitation hypothesis" suggests that translations are more explicit than their corresponding source texts. Some authors believe it is a "universal strategy" employed by translators (Baker 1993, p. 243; Blum-Kulka 1986, p. 21). Others disagree with this point of view and believe that explicitation in the target language occurs due to obvious differences in linguistic and stylistic conventions between source language and target language texts (House 2004, p. 193; Becher, 2010, p. 2, 2011). According to Klaudy (2008) (cited in Becher, 2011), 'explicitation' may be of the following types:

(a) obligatory (caused by grammatical differences between the languages; the translator has to compensate for those differences by making certain parts of text explicit);

(b) optional (explicitations that are not obligatory and rely on translators' personal choice);

(c) pragmatic (explicitations that are required due to some cultural or world knowledge differences);

(d) translation-inherent (result from the process of translation itself); as stated in Becher (2010:3), it is quite difficult to provide examples for the last type.

Roughly speaking, these types of explicitation can be subcategorized into two larger groups based on the criterion of obligation: obligatory and voluntary (i.e. non-obligatory) explicitation (Frankenberg-Garcia, 2004, p. 2).

Section 6 (especially, in 6.3.) will reflect on the applicability of these strategies to the analysis of DMs carried out on the basis of the translated literary texts from English into Portuguese.

3 OMISSION OF DMs IN CROSS-LINGUISTIC RESEARCH

Many papers on DMs traditionally yield to explanation of the existing rich terminological range that define this linguistic category, be it discourse markers (Fraser, 1999; Schiffrin, 1987; Lopes, 2014), pragmatic markers (Brinton, 1996; Aijmer, 2004; Cuenca, 2008), pragmatic particles (Östman, 2006), discourse connectives (Blakemore, 2002), etc., well-known labels, which certainly are not limited by this short exemplification list[1]. In this respect, it is important to mention here that the analysis

---

[1] For a more complete list of terms, consult Lutzky (2012, p. 9-10).

presented in this paper does not refer to the omission of 'discourse markers' only, but to omission of the categories also defined as 'discourse connectives', 'pragmatic markers', and so on. In other words, and at this stage of the research, the paper keeps the terminological debate aside, focusing on the process of omission.

As stated in Aijmer (2007, p. 50), omission of DMs seems to be a general phenomenon. Indeed, a large amount of cross-linguistic research on DMs in translation that have been carried out so far necessarily refers to omission as a most frequent occurrence in the analyzed data. Omission (or 'zero correspondence', as it is called in Aijmer & Altenberg, 2002) is a frequent phenomenon in the studies of translation of DMs from English into a range of European languages: Cuenca (2008) for English, Spanish and Catalan; Ramón (2015) and Chaume (2004) for Spanish and English; Becher (2011) for English and German; Mattsson (2006) for Swedish and English; Dupont & Zufferey (2017) for English and French; Hauge (2014) for English and Bulgarian; and the contributions on this issue by Aijmer (2007) for German and Swedish, and Aijmer & Simon-Vandenbergen (2003) for English, Swedish and Dutch; and also Kafipour (2016) for Persian and English.

In translation studies based on parallel corpora omission seems to be far more interesting. It is frequent, relevant and omnipresent. However, omission has not been studied on its own. It is commonly stated as a fact, rather than analyzed, classified, etc. in an attempt to account for such cases. As Torabi & Demberg (2015, p. 118) state, there is no theory to explain "when and why a discourse relation is marked explicitly or when the connective is omitted".

The translation of *well* into Spanish and Catalan based on a film with subtitles demonstrated that omission is frequent (Cuenca, 2008, p. 1377). In the Catalan version it is the preferred option (around 46%), and in Spanish it is quite frequent as well (25%). Another interesting observation is that omission is frequent when a DM combines with some other DM or an adverb (Cuenca, 2008, p. 1387). Omission is seen as a specific and frequent translation 'strategy' (Cuenca, 2008, p. 1378). Moreover, Cuenca adds that omission can be explained by the presence of different factors. One of them is the fact the DMs do not carry meaning and their presence in discourse is rather optional (Cuenca, 2008, p. 1379).

The analysis of Persian and English movies with subtitles revealed that omission is the most frequent error made by translators (Kafipour, 2016, p. 104). The second most frequent error was translation using a different DM. It has been suggested that omission can be a frequent phenomenon in the translated texts due to some kind of interdependency between the nature of the source language and the target language. Namely, a DM is likely to be omitted in the target language that is not a DM-rich language (like Persian).

Mattsson (2006) focuses on the subtitling of Swedish/English DMs of one particular movie in three different TV formats: public television, commercial television and DVD. The analysis revealed that DMs are omitted in all three texts. This is explained by the fact that different sets of translational norms are employed in the target culture and its media channels. For instance, different TV channel translation policies on subtitling, different working conditions or translating agencies can be used (Mattsson, 2006, p. 9).

In the Spanish/English language pair, Ramón (2015, p. 342) studies the English *oh* and its Spanish translations, based on a parallel corpus. Omission (or other

options selected by translators) of the DM *oh* is explained by the fact that *oh* does not cover the same range of pragmatic functions in English and in Spanish, which results in employment of different translational patterns.

The most significant contribution probably belongs to Aijmer (2007), who focused on the German/Swedish pair and on the Swedish marker *alltså* (*also* in German). As stated in her paper (Aijmer, 2007, p. 50), omission was a frequent option when *alltså* had a medial position (41%). In this respect, Aijmer hypothesizes that omissions could be a general 'universal' translation strategy (Ibid.). In addition to this hypothesis, the author believes that omission can occur when there is clash of a DM with some other linguistic items that have a similar or an identical meaning. The DM is omitted in order to avoid this redundancy of meaning. Finally, omission could be the signal pragmaticalization of a DM. Once the meaning of a DM has been weakened, it is no longer used (Aimer, 2007, p. 54).

Besides, concessive connectors are often left untranslated (Aijmer & Altenberg, 2002, p. 22). The authors believe that "most connectors can be omitted if the context is clear enough" (Ibid.). It comes down to the fact the DMs are known to have a procedural, rather than a core semantic meaning. Their procedural pragmatic meaning is helpful, but sometimes unnecessary. In this respect, Aijmer (2007, p. 52) further explains that omission is not necessarily a lack of translation pragmatic equivalent in the target language. Sometimes, the context is clear enough and there is no necessity to add an extra word.

Overall, omission of DMs is a widely encountered phenomenon in cross-linguistic research based on parallel corpus. Regarding the explanations of this phenomenon, so far, we can say that:

    (a) omission can be a common translation strategy;

    (b) omission can occur when there is clash of a DM with other linguistic items that have a similar or an identical meaning[2];

    (c) omission can be the signal pragmaticalization of a DM, i.e. if the meaning of a DM has been weakened, it is no longer used;

    (d) omission can occur when the context is clear enough, i.e. to avoid information density;

    (g) omission can occur in case a DM does not cover the same range of pragmatic functions in the target language;

    (e) omission can occur if the target language is not a DM-rich language;

    (f) omission can be explained by different sets of translational norms that are employed in the target culture.

---

[2] Section 6 will provide more details on this point.

4 METHODOLOGY AND CORPUS OF THE PRESENT RESEARCH

Based on the assumption that DMs have a core semantic meaning and a context-dependent pragmatic function [3], Aijmer & Simon-Vandenbergen (2003) suggested that a DM can have various translations (or pragmatic functional equivalents) in the target language. In this respect, Aijmer (2004) proposed a methodology and a model for the analysis of DMs based on translations, or 'translation method', which has already been implemented in some other studies[4]. Likewise, this method has been adopted for the present analysis.

In short, the translation method

> contributes to specifying how markers function intra-linguistically, how they relate to other, semantically and pragmatically similar items in the same language, and how semantic fields in different languages relate to one another (Aijmer, 2004, p. 1782).

Translations function as means to establish paradigms between languages (Aijmer, 2004, p. 1785). As Baker (1993, p. 235) stated: "the question is no longer how equivalence might be achieved but, increasingly, what kind of equivalence can be achieved and in what contexts".

The proposal of this methodology was illustrated by the analysis of markers of the semantic field of expectation (*actually*, *in fact* and *really*) in English and their pragmatic equivalents in Swedish and Dutch. The analysis was carried out on the basis of the English-Swedish Parallel Corpus for English/Swedish and of the Triptic Corpus for English/Dutch.

The translation method has a number of benefits. As stated in Aijmer (2004, p. 1786), translation data provides a more detailed description of the polysemous meanings of a lexical item. In other words, it helps to reveal hidden meanings and can be a practical instrument for a researcher who wants to test new hypotheses about DMs (Cuenca, 2008, p. 1389). Secondly, the resulted semantic map helps to see which equivalents in the target language are more distant and which ones are closer to the translated token in the source language. Last but not least, the translation method provides information not only about the core meaning, but also about the pragmatic interpretations of the tokens.

Regarding the selected corpus, the present study is based on a bidirectional parallel corpus of English and Portuguese - *Compara* (Santos, 2000; Frankenberg-Garcia & Santos, 2003; for more details on *Compara* see http://www.linguateca.pt/COMPARA), which is a database with original and translated texts in these two languages. The texts have been linked together sentence by sentence. *Compara* is a useful tool for researchers, lexicographers, translators and editors, teachers, students and engineers related to natural language processing. It allows the study of human translation by contrasting English and Portuguese by means of an automated search.

---

[3] For more on this point, see Aijmer (2004), Blakemore (2002), Brinton (2008), Cuenca (2008), Fischer (1998), Fraser (1999), Lopes & Sousa (2014).

[4] For instance, Cuenca (2008) based her analysis of the translation of the DM *well* into Spanish and Catalan on this method.

Compara is currently the largest post-edited English-Portuguese/Portuguese-English parallel corpus in the world, totaling around three million words. At present, *Compara* is composed of 75 pairs of digitized literary parallel texts.

5 DMs IN EUROPEAN PORTUGUESE

Similarly to the English DM *well*, its Portuguese equivalent *bem* has several uses in the syntactic and pragmatic domains[5]:

(I)   In the syntactic domain:

  (a) as an adverb, for instance: "O João comeu **bem**"[6] ('John ate **well**'[7]), *bem* ('well') refers to mode or manner (Lopes, 2004, p. 7).

  (b) as an adjective modifier, for instance: "O homem é **bem** alto" ('The man is **very** tall'). The meaning of the marker is similar to the Portuguese adverb *muito* ('very'), compare with "O homem é muito alto" ('The man is very tall'). Also, it serves as an intensifier. Lopes (2004) defines these adjectives (*alto* or 'tall') as *pure* adjectives, i.e. they are nor derived from verbs.

  (c) as an adverbial modifier of mode or manner (with *verbal* adjectives), for instance: "uma mulher **bem** arranjada" ('a good looking woman').

  (d) as an adverbial modifier of quantity or degree, for instance: "Momento de decisões históricas que quase passavam para segundo plano depois da morte de dois soldados israelitas na zona de segurança junto ao Líbano – **bem** perto de Haifa" ('Moment of historical decisions that almost passed into the background after the death of two Israeli soldiers in the security zone near Lebanon – **very** close to Haifa') (Lopes, 2004, p. 8).

(II) In the pragmatic domain[8]:

  (a) it marks / signals discordance with the information stated before;

  (b) it initiates discourse, serving as a marker of turn-taking;

  (c) it marks / signals shift of topic;

  (d) it fulfills a mitigating function.

Lopes (2004, p. 8) provides various examples of pragmatic functioning of *bem*, one of them is as follows: "**Bem**, agora estão a lavar as grelhas, depois para o tempo da…" ('**Well**, now they are washing the grates, then for the time of...'). In this case, the DM clearly does not perform any syntactic function. It is not part of the propositional content of the utterance either. Prosodically speaking, *bem* in the

---

[5] These functions have been identified by Lopes (2004) on the basis of the 'Reference Corpus of Contemporary Portuguese' (*Corpus de Referência do Português Contemporâneo*).

[6] All the examples can be found in Lopes (2004, p. 7-8).

[7] These translations from Portuguese into English are my own.

[8] Lopes (2004, p. 21) explains that the wide range of functions *well* performs in the pragmatic domain is due to its interactional nature.

pragmatic domain demonstrates independence (preceding and anteceding pauses, i.e. a feature typically associated with DMs).

Following Traugott (1998), Lopes explains that the different uses of *bem* in the pragmatic domain are the result of gramaticalization (or 'semantic bleaching'), i.e. recategorization of a linguistic unit.

As for the Portuguese DM of reformulation *quer dizer* ('I mean'), it is a fully grammaticalized expression in contemporary European Portuguese. It operates in such a way that it implies a source utterance on the left and a reformulation on the right (Lopes, 2014, p. 34). In terms of prosody, it is highlighted by a pause (or a comma is case of written discourse).

But for reformulation, *quer dizer* ('I mean') can also express additional meanings in the context (Lopes, 2014). Additional meanings include mitigation, conclusion or 'filler' (the latter used in oral spontaneous discourse only).

Concerning the DM *sabes* ('you know'), so far there has been no study on this marker in Portuguese.

6 DATA ANALYSIS

6.1 Establishing Portuguese functional equivalents

Using the translation method described in Section 4, the present section provides the analysis of the English DMs *well*, *you know* and *I mean* and their functional equivalents in Portuguese (see Table 1, Table 2 and Table 3). It is noteworthy that *Compara* does not differentiate between the use of *well* as a DM and *well* as an adverb, for instance. In other words, the automated search provides all the possibilities. *Compara* has identified two principal modes of functioning of the DM *well*, compare:

(1) a. EBDL1T1(49): **Well**, I´ll tell you what to do. – **Bem**, vou dizer-lhes o que costumo fazer.

b. EBDL3T1(1090): He seems to know his material very **well**, but resents questions and discussion as they interrupt his train of thought. – Parece conhecer muito **bem** a matéria, mas não gosta de perguntas nem de discussão, porque lhe cortam o fio ao pensamento.

In 1a. *well* functions as a DM, while in 1b. it is an adverb of manner. In order to distinguish between these two modes of functioning of *well* (and other DMs), the table contains the information on the total number of the occurrences identified in the corpus, and the total number of DMs found. To facilitate the reading, the Portuguese pragmatic equivalents with the highest number of occurrences and the total number of occurrences of omission have been boldened.

Table 1. Translations of the English DM *well* into Portuguese.

Translations	N. of occurrences
bem	**29**
bom	9
enfim	9
ora esta	1
paciência	2
percebo	1
OK	1
ora	1
afinal	1
com que então	1
pois	1
então	2
Ø	**34**
Total n. of occurrences	1000 (1064)[9]
Total n. of DMs	91

Table 2. Translations of the English DM *I mean* into Portuguese.

Translations	N. of occurrences
quero dizer	**41**
quer dizer	20
quero eu dizer	5
quero dizer que	3
ou seja	4
por exemplo	1
isto é	6
afinal	1
aliás	1
a propósito	1
claro	1
o que eu quero dizer é que	2
Ø	**22**
Total n. of occurrences	188
Total n. of DMs	108

FLP22(1)

---

[9] For copyright reasons, *Compara* provides a random sample of 1000 occurrences out of 1064 that have been found.

Table 3. Translations of the English DM *you know* into Portuguese.

Translations	N. of occurrences
sabe	**17**
sabes	13
você sabe	3
sabem	2
como sabes	2
como sabe	2
como sabem	1
estás a ver	1
se quer saber	1
sabe disso	1
não sei se sabes	2
sabia	1
sabias	1
sabe como é	1
bem sabes	1
bem vê que	1
veja lá bem	1
imaginem lá	1
não sei se estão a ver	1
sabes que	3
sabem como é	1
percebem	1
percebe	1
percebes	2
compreende	1
compreenda	1
não é	1
claro	1
já se sabe	1
não acham	1
a senhora sabe	1
o senhor sabe	2
olha que	1
sei lá	1
Ø	**28**
Total n. of occurrences	388
Total n. of DMs	100

FLP22(1)

As it can be seen, the Portuguese DM *bem*, *bom* and *enfim* are predominant pragmatic equivalents of *well*. The least common but nevertheless identified options include the DMs *ora esta*, *paciência*, *percebo*, *OK*, *ora*, *afinal*, *com que então*, *pois* and *então*.

Concerning the DM *I mean*, it is predominantly translated as *quero dizer* or *quer dizer*. Options like *quero eu dizer*, *quero dizer que*, *ou seja*, *isto é* are less frequent, but nonetheless found. Single occurrences include DMs like *por exemplo*, *afinal*, *aliás*, *a propósito*, *claro*.

The English DM *you know* presents the widest range of possible translation equivalents in Portuguese. In total, 34 tokens have been identified. The most frequent translations include *sabes* and *sabe*. The majority of these tokens are single and

sometimes quite unexpected translations (for instance, *sei lá, já se sabe, o/a senhor(a) sabe, imaginem lá*, etc.)

In sum, the results of this analysis partially overlap with the conclusions found in Cuenca (2008) for Spanish and Catalan, namely, that:

(a) the English DMs *well, you know* and *I mean* have a wide range of functional equivalents in Portuguese. The DM you know got the highest number of possible translations in the target Portuguese language (34 translation equivalents altogether).

(b) the semantic correspondences of *well, you know* and *I mean* in Portuguese can be very distant from the literal translation (e.g. *you know = imaginem lá* or *imagine*). This proves the fact that translation can be misleading if it is merely based on establishing a corresponding token with the same core semantic meaning, without considering the pragmatic function it performs.

(c) literal Portuguese translations of the English DMs *well, you know* and *I mean* represent the highest percentage of occurrences, therefore, the tendency towards literal translation is maintained.

As for omission, and similarly to previous studies, it has proved to be a frequent option[10]. In fact, omission is a preferred option in the translations of DMs *well* and *you know* (34 occurrences of the omitted DM *well*, 22 occurrences of the omitted DM *I mean*, and 28 in case of the DM *you know*).

6.2 Back translations from Portuguese into English

The next step was to identify the translations that are 'mirrored back' into the original language (Table 4). This method, proposed by Dyvik and cited in Aijmer (2007, p. 1785) is especially useful, for it "implies that we look at the meaning of a lexical item as mirrored in its translations in another language". It is handy when a DM has an extremely wide range of translations in the target language (like in case of *you know*). By looking at the back translations we can see "which translations are more frequent or prototypical, and which are less frequent or even 'singleton' translations" (*Ibid.*). Table 4 presents back translations from Portuguese into English. Namely, it contains two most frequent equivalents of *well* (*bom* and *bem*), *you know* (*sabes* and *sabe*) and *I mean* (*quer dizer* and *quero dizer*) and highlights back translations and omissions.

FLP22(1)

---

[10] The examples of omissions will be closer examined in 6.3.

Filol. Linguíst. Port., São Paulo, v. 22, n. 1, p. 103-121, jan./jun. 2020

114

Table 4. Back translations from Portuguese into English.

DM	Translation (n. of occurrences)
bem	well (13) that's that (1) Ø (2)
bom	well (8) good (1) anyhow (1) all right (1) Ø (4)
sabes	you know (10) you see (1) Ø (0)
sabe	you know (17) you see (7) do you know (1) you understand (1) I hope you realize (1) Ø (2)
quer dizer	I mean (11) that is to say (5) that is (4) you mean (1) in other words (1) Ø (3)
quero dizer	I mean (10) I meant to say that (1) which is to say (1) Ø (4)

As it can be seen, back translations confirm the results of the first stage of the analysis. They allow us to state more confidently that the English-Portuguese DMs *well - bem*, *I mean - quero dizer* and *you know - sabe* are likely to be functional equivalents. As for omission, it is considerably less frequent in Portuguese-English translations. Besides, it is never a predominant option. This difference could be perhaps explained by the fact that in English-Portuguese language pair, English tends to signal discourse relations more explicitly, i.e. it requires the use of DMs for communicative goals.

6.3 Omissions of Portuguese DMs and their typification

So far, it has been stated that omission of DMs in translations is a common phenomenon and that it has been attested in the translations of many European language pairs. This section, however, aims to provide a deeper insight into the omitted cases of DMs attested during the first step of the analysis (Tables 1, 2 and 3). We have seen that in the case of the DMs *well* and *you know* omission is a preferred option (34 and 22 occurrences, respectively). In the case of the DM *I mean*, *quero dizer* is the predominant translation (41 examples); however, omission is the second most frequent option (22 examples).

Each case of the omitted DM has been analyzed, totaling 84 examples. The analysis resulted in identification of several main types of DMs omission, namely, DM deletion, DM partial deletion, DM addition[11]. Each of these types is presented with the corresponding examples below:

Type Nº1: *DM deletion*, i.e. a common DM deletion or 'omission' in the target language, for instance:

(2) a. **I mean**, how trivial can you get?
     b. Até onde pode ir a superficialidade de uma pessoa?
         (Ø)

(3) a. **I mean**, did you go anywhere interesting?
     b. Perguntei se foi a algum lado interessante.
         (Ø)

(4) a. "He is right, **you know**", Amy said...
     b. "Ele tem razão", disse a Amy...
          (Ø)

(5) a. He is obsessed by things like that, **you know**.
     b. É obcecado por coisas assim.
         (Ø)

(6) a. Yes, **well**, nice to have met you again.
     b. Gostei de voltar a vê-lo.
         (Ø)

(7) a. Yeah, **well**, it's like a fruit machine, y' know, you got to keep pulling the old lever.
     b. Pois, isto é como uma máquina de moedas, temos de ir puxando a alavanca.
         (Ø)

(8*) a. **Well**, it's like James, ah, well the guy wants to be a modem, I mean he has the symbolism bit and God is dead and all...
      b. É assim: O James, pronto, o que o tipo quer é ser moderno, tem isso do simbolismo, e Deus morreu, pronto.
          (Ø)
     bb. **Bem**, é como James, ah, bom, o cara quer ser moderno, quer dizer, ele sacou o lance do simbolismo, Deus está morto o tudo o mais.

Examples 2 and 3 illustrate the omitted English DM *I mean* in the corresponding Portuguese translation; examples 4 and 5 refer to the omission of *you know* (*sabes* in Portuguese); and examples 6 and 7 show cases where *well* (*bem*) has been fully omitted in the target translation.

Example 8* is identified as a rare and a special case found in *Compara*, when an original utterance has two possible interpretations provided by different translators. In this case, one of the possible translations includes the DM in the target language (bb), while the other one omits it (b). These cases with double translations and including two possibilities are quite rare in the corpus. Yet, they may clearly be indicators of

---

[11] This typification is not based on any previous study on omission of DMs; therefore, a further terminological revision may be required.

translators' professional expertise, which result in optionality and instability of the translated tokens.

DM deletion is the most frequent type of omission (73 out of 84 examples). Omission could be explained in terms of translation strategy of 'implicitation' (Pym, 2005, p. Becher, 2010, 2011). However, in the examples 2-7 it is rather difficult to draw the implicatures about particular DMs used in the context. Consider example 6(b) "Gostei de voltar a vê-lo.". This context does not permit any reconstruction of meaning. It does not have any clue for the reader or hearer to 'guess' about the implicit DM *well*. It could be *well*, as it could be *really*: "Yes, **really**, nice to have met you again.". That is why the analysis of the DMs is particularly puzzling. Their function is more likely to be defined as interactional, rather than textual. And they seem to exist beyond text level. If we looked at any kind of textual DMs (or 'connector'), whose function would be merely textual (i.e. a cohesive device), it would possibly be easier to trace or to 'reconstruct' the implicit DM. Consider the following random example with the DM *so* ('portanto'), extracted from Compara:

(9) a. I tried to grow a moustache once, but it turned out rather funny-looking, grey on one side and a sort of gingery-brown on the other, **so** I shaved it off quick.

   b. Uma vez tentei deixar crescer o bigode, mas ficou um bocado esquisito, pois metade era grisalha e a outra metade ligeiramente acastanhada, e **(portanto)** rapei-o logo.
   Ø

9(b) is an example of explicatory sequence and it contains a causal relationship. In 9(b), since causal relationship is of merely textual nature, the reader/listener can easier reconstruct the implicit DM *so*: "Uma vez tentei deixar crescer o bigode, mas ficou um bocado esquisito, pois metade era grisalha e a outra metade ligeiramente acastanhada, e **(portanto)** rapei-o logo".

Type Nº2: *partial DM deletion*, i.e. when one of the two DMs in the original language drops in the target language. In other words, just one of the DMs is translated. Consider the examples below:

(10) a. **I mean**, **perhaps** another evening if you're not free.
   b. Ou **talvez** outra noite, se hoje não está livre.
   (partial Ø)

(11) a. We didn't really want to start a family right away, **but** well, **you know** how it is...
   b. Não queríamos que a família aumentasse tão depressa, **mas** **sabes** como é...
   (partial Ø)

(12) a. We have a view, **too**, **you know**.
   b. Temos uma vista **também**.
   (partial Ø)

(13) a. **But**, **you know**, I really needed your help that day in the Lake District.
   b. **Mas** naquele dia, no Lake District, precisava mesmo da tua ajuda.
   (partial Ø)

As it can be seen, different kinds of DMs combinations have been observed: "*I mean, perhaps*" (example 9); "*too, you know*" (example 11) and "*but, you know*" (example 12). Sometimes, a larger group of DMs can operate together ("*but well, you now*" in example 10).

A similar type of omission has been identified by Cuenca (2008, p. 1387) in English, Spanish and Catalan. A DM combined with some other marker (a question tag or an adverb, for instance) was dropped. Moreover, Aijmer (2007, p. 50) confirmed that omission can occur when there is clash of a DM with other linguistic items that have a similar or an identical meaning. A DM is omitted in order to avoid this redundancy of meaning. If we look at the examples 10-13, this explanation could probably be applicable to the example 11. The DMs "*but well, you know*" all together do create some sort of clash. Besides, *well* and *you know* could have a similar meaning in the context.

As for the examples 10, 12 and 13 ("*I mean, perhaps*", "*too, you know*" and "*but, you know*"), it is difficult to speak of any redundancy in meaning. Each of the DM in these cases has its own meaning. Nonetheless, one of them displaces the other one.

Type N°3: *DM addition*, i.e. when there is no DM in the original language, but the translator has added it, for instance:

(14) a. **I mean**, you said so yourself.
    b. <u>Pelo menos</u> foi o que tu disseste.
       ('<u>at least</u>')

(15) a. **I mean**, I can't quite think how you can really love a man who thinks like that...
    b. <u>Também</u> não percebo como podes gostar de um homem que pensa dessa maneira...
       ('<u>also</u>')

(16) a. **I mean**, it didn't seem terribly normal to name a bird after your mother.
    b. <u>Realmente</u>, quem é que dava o nome da mãe a um passáro?
       ('<u>really</u>')

(17) a. I said, "I could walk, **you know**, in a dressing-gown"
    b. Disse-lhe que, <u>mesmo</u> em camisa de dormir, poderia ir a pé.
       ('<u>even</u>')

(18) a. Yes, **well**, if you wouldn't mind taking her off a minute, and passing through the gate again...
    b. Claro, <u>desculpe</u>. Se não se importa de tirar a medalha por um minuto e passar outra vez pelo detector...
       ('<u>sorry</u>')

Examples 13-17 illustrate cases in which a DM in the original sentence disappears in the target translation. Instead, the translator adds another (and different from the original sentence) word in order to substitute or to make up for the necessary meaning in the target language.

It is, however, difficult to speak of a 'voluntary explicitation' strategy (Frankenberg-Garcia, 2004, p. 2; Becker, 2011, p. 2-3), since there seems to be no 'explicitation'. In fact, the DM is completely removed from the target version of the sentence. This results in a difference of meanings between the two languages or even

its loss. Neither we can speak of 'implicitation' strategy, since the recovery of the DM based on the Portuguese translations is unlikely to occur.

This type of omission brings forward the problem of a relative subjectivity of translators and the various ways an original sentence can be interpreted. In this respect, some interesting cases can be found in Hauge (2014), who analyzed translations from English into Bulgarian. The results are quite similar. Where there was no DM in the source text, the translator added it him/herself. It is quite problematic to analyze such voluntary and free modifications due to lack of knowledge on the level of the translator's expertise. The modifications can be the result of the translator's expertise in these languages and a conscious strategic move in order to compensate for the lexico-syntactic differences (when, for instance, one of the languages is richer in DMs). Or, on the contrary, they could indicate a possible lack of translator's experience.

To sum up point 6.3, Table 4 contains the information on the three types of omission identified in translations of DMs from English into Portuguese.

Table 4. Types of omission identified in translations of DMs from English into Portuguese.

Type of omission (Ø)	DM (n. of occurrences)
Deletion (Ø)	well (32)
	I mean (19)
	you know (22)
Subtotal n.	73
Partial deletion (Ø)	well (1)
	I mean (1)
	you know (3)
Subtotal n.	5
Addition (Ø)	well (1)
	I mean (3)
	you know (2)
Subtotal n.	6
Total n. of omissions	84

FLP22(1)

7 CONCLUSION

Different stages of the analysis presented in Section 6 have a) illustrated the ways in which English DMs *well, you know* and *I mean* can be translated into Portuguese resulting in a wide range of tokens; b) demonstrated to what extent the back translations into English contribute to establish the functional equivalents between these languages; c) provided a deeper analysis into the frequent types of omission identified in the parallel bidirectional English-Portuguese corpus *Compara*. The last stage of the analysis resulted in the elaboration of a typification of recurrent patterns of the omitted DMs.

However, looking at a boarder picture, the analysis of omissions and functional equivalents in translations can be biased due a number of factors.

First of all, the nature of the language itself definitely predetermines its richness in DMs or any other linguistic category. We have seen that while for some languages it is acceptable to use a large number of DMs in discourse, others may be less rich in terms of DMs range and, therefore, more reluctant to avoid them (like English and Persian in Kafipour, 2016). However, the degree of 'richness' is relative. While English is seen as a DM-rich language in the English-Persian language pair, the same may not be necessarily true if we compare it with other languages. Dupont & Zufferey (2017) claim that French tends to require more explicit linking than English. In other words, DMs tend to be more explicit and frequent in French than in English. Thus, it is not easy to say to what extent certain languages are richer in terms of DM. The results can vary depending on the type of data that is compared.

Secondly, translation optionality and voluntary interpretation of the original data into the target language deserved a special attention. In this respect, Dupont & Zufferey (2017) conclude that certain factors (such as corpus register, directionality of corpus translation and translation expertise) have a crucial impact on the translators' selection of the options (including 'zero-correspondences' or omission) and that it is necessary to take these factors in to account while establishing cross-linguistic equivalents in a contrastive analysis based on parallel corpora.

Finally, as pointed out in Dupont & Zufferey (2017), there are other secondary factors to be considered like: a) whether the translations have been subject to revision; b) whether the translators work from or to the mother (native or dominant) language; c) whether the texts have been produced by a native speaker or not.

The results of this study suggest that in order to provide a trustworthy, valid and rigorous analysis of the data (be it analysis of DMs or any other linguistic category based on translations from one language into another), it is absolutely essential to reach translation homogeneity. To do so, there is the need to create a corpus from scratch, taking into a thorough consideration the above-mentioned factors in order to provide reliable results and reveal the strategies employed by the translators.

REFERENCES

Aijmer K, Altenberg B. Zero translations are cross-linguistic equivalence: Evidence from the English-Swedish Parallel Corpus. In: Stenstrom AB, From the Colt's mouth...and others'. Amsterdam/New York: Rodopi; 2002. p. 19-41.

Aijmer K, Simon-Vandenbergen AM. The discourse particle well and its equivalents in Swedish and Dutch. Linguistics. 2003;41(6):1123-1161.

Aijmer K. A method and a methodology for the study of pragmatic markers: the semantic field of expectation. Journal of Pragmatics. 2004;36:1781-1805.

Aijmer K. The meaning and functions of the Swedish discourse marker alltså - evidence from translation corpora. Catalan Jornal of Linguistics. 2007;6:31-59.

Asr FT, Demberg V. Implicitness of discourse relations. In Proceedings of the 25th International conference on computational linguistics COLING, Mumbai, India. 2012; 2669-2684.

Baker M. Corpus linguistics and translation studies: implications and applications. In: Studies in text and technology. In: Baker M, Francis G, Tognini-Bonelli E., Honour of John Sinclair. Amsterdam/Philadelphia: John Benjamins. 1993, p. 233-250.

Becher V. Towards a more rigorous treatment of the explicitation hypothesis in translation studies. Trans-Kom. 2010;3(1):1-25.

Becher V. When and why do translators add connectives? Target. 2011;23(1):26-47.

Blakemore D. Relevance and linguistic meaning: the semantics and pragmatics of discourse markers. Cambridge: Cambridge University Press; 2002.

Blum-Kulka S. Shifts of cohesion and coherence in translation. In: House J, Blum-Kulka S, Interlingual and intercultural communication. Tübingen: Narr; 1986. p. 17-35.

Brinton LJ. Pragmatic markers in English: grammaticalization and discourse functions. Berlin/New York: Mouton de Gruyter; 1996.

Brinton LJ. The comment clause in English. Syntactic origins and pragmatic developments. Cambridge: Cambridge University Press; 2008.

Chaume F. Discourse markers in audiovisual translating. Meta. 2004;49(4):843-855.

Cuenca MJ. Pragmatic markers in contrast: the case of well. Journal of Pragmatics. 2008;40:1373-1391.

Dupont M, Zufferey S. Methodological issues in the use of directional parallel corpora. A case study of English and French concessive connectives. International Jornal of Corpus Linguistics. 2017;22(2):270-297.

Fischer K. Validating semantic analysis of discourse particles. Journal of Pragmatics 1998;29:111-127.

Frankenberg-Garcia A, Santos D. Introducing COMPARA, the Portuguese-English parallel corpus. In: Zanettin F, Bernardini S, Stewart D, Corpora in translator education. Manchester: St. Jerome; 2003; p. 71-87.

Frankenberg-Garcia A. Are translations longer than source texts? A corpus-based study of explicitation. Paper presented to the Third international corpus use and learning to translate conference, Barcelona, January 2004.

Fraser B. What are discourse markers? Journal of Pragmatics. 1999;31:931-952.

Hauge KR. Found in translation - Discourse markers out of the blue. In: Ebeling G, Hauge and Santos, Corpus-based studies in contrastive linguistics, Oslo Studies in Language. 2014:6(1):43-52.

Kafipour R. Errors in the translation of discourse markers from English into Persian in movie subtitles. American Journal of Educational Research. 2016;4(15):1100-1105.

Klaudy K. Explicitation. In: Baker M, Saldanha G, Routledge encyclopedia of translation studies, London: Routledge; 2008, p. 80-85.

Lopes ACM. Contributo para o estudo sincrónico dos marcadores discursivos 'quer dizer', 'ou seja' e 'isto é' no português europeu contemporâneo. Diacrítica. 2004;28(1).

Lopes ACM, Sousa S. The discourse connectives ao invés and pelo contrário in contemporary European Portuguese. Journal of Portuguese Linguistics. 2014;3(1):3-27.

Lutzky U. Discourse markers in Early Modern English. Amsterdam/Philadelphia: John Benjamins Publishing Company; 2012.

FLP22(1)

Mattsson J. Linguistic Variation in subtitling. The subtitling of swearwords and discourse markers on public television, commercial television and DVD. MuTra 2006 - Audiovisual translation scenarios: Conference Proceedings.

Östman JO. Constructions in cross-language research: Verbs as pragmatic particles in Solv. In: Aijmer K, Simon-Vandenbergen AM, Pragmatic markers in contrast 2. Amsterdam: Elsevier; 2006. p. 237-257.

Pym A. Explaining explicitation. In: Karoly K, Foris A, Honour of Kinga Klaudy. New trends in translation studies. Budapest: Akadémia Kiadó; 2005. p. 29-34.

Ramón N. The English discourse particle 'oh' in Spanish translations: evidence from a parallel corpus. Procedia - Social and Behavioral Sciences. 2015;173:337-342.

Santos D. O projecto processamento computacional do português: balanço e perspectivas. V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000), São Paulo, Brasil, 19-22 de Novembro de 2000, p. 105-113. Disponível em: http://www.linguateca.pt/COMPARA.

Schiffrin D. Discourse markers. Cambridge: CUP; 1987.

Sperber D, Wilson D. Relevance: communication and cognition. Oxford UK & Cambridge USA: Blackwell; 1986.

Torabi F, Demberg V. Uniform information density at the level of discourse relations: negation markers and discourse connective omission. Proceedings of the 11th International conference on computational semantics, London, UK; 2015. p. 118-128.

FLP22(1)