

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Discourse markers in English as a target language: the use of so by simultaneous interpreters

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1705861> since 2019-07-08T19:06:38Z

Published version:

DOI:10.7370/93189

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Discourse Markers in English as a Target Language: The Use of *so* by Simultaneous Interpreters

Claudio Bendazzoli

Abstract

This paper investigates the distribution of a particular discourse marker, i.e. *so*, in the target speeches produced by professional simultaneous interpreters while translating from Italian into English. The objective is to examine the possible effect on discourse marker distribution of specific situational norms that are in play in simultaneous interpreter-mediated settings. The analysis is both quantitative and qualitative, and is based on a parallel corpus of three medical conferences with Italian and English (native and non-native) speakers along with the corresponding simultaneous interpretations. All the occurrences of zero correspondence (30% of all the occurrences of *so* in target speeches) are examined in detail and grouped into different macro-categories. Subsequently, there is a discussion of possible reasons behind the interpreters' decision to add "sequentially dependent elements which bracket units of talk" (Schiffrin 1987: 31), with a view to contributing to the description of English in interpreter-mediated communication.

Keywords: conference interpreting, interpreting from Italian into English, discourse markers, additions, simultaneous interpreting strategies, DIRSI corpus.

1. Introduction

English is both the *de facto* lingua franca of the medical sciences and, more generally, of globalisation and internationalisation processes – and this inevitably extends to the translation and interpreting (T&I) industry (Albl-Mikasa 2010). The primary role of English in these fields is reflected in the Directionality in Simultaneous Interpreting (DIRSI) Corpus. DIRSI is a parallel corpus of Italian and English speech events recorded at three medical conferences held in Italy and mediated by professional simultaneous interpreters (Bendazzoli 2010, 2012). Two of the conferences were also open to the general public (i.e. patients and their families) and one conference was part of the agenda of a transnational European project involving different partners, who were required to use English as their official language (Bendazzoli 2017). The fact that in these circumstances communication was envisaged in Italian as well as English explains why Italian/English simultaneous interpreters were hired.

As is common in interpreting markets outside international institutions, simultaneous interpreters work in a bidirectional mode, i.e. they translate both from and into the two working languages involved. Interpreters' working languages are generally classified as language A (their native language), language B (their 'active' foreign language, i.e. that they can interpret from *and* into) and language C (their 'passive' foreign language, i.e. that they can only interpret from) (AIIC 2012).

On account of the particular conditions in which simultaneous interpreters' language production takes place, distinguishing features of what is also known as *interpretese* have been highlighted by a number of scholars (e.g. Shlesinger 2008; He, Boyd-Graber, Daumé 2016; Kajzer-Wietrzny 2018). Among these features, Straniero Sergio (2012) highlighted the special role played by interpreter-generated discourse markers. Against this background, the aim of this study is to use corpus methods to carry out both quantitative and qualitative analyses of the use of discourse marker (DM) *so* in English as a target language, i.e. the language produced by interpreters.

The following research questions will be addressed: how do simultaneous interpreters use *so*? To what extent do they use it only in response to an equivalent or similar unit of meaning in the source text? Alternatively, do they add it and use it in a more independent way?

Section 2 gives a general overview of DM use in both non-mediated and interpreter-mediated communication. This is followed by a description of the DIRSI corpus (section 3) and of the

methodology of the study (section 4). Results are then presented and discussed (section 5). Section 6 contains the conclusion and future developments of this line of enquiry.

2. Discourse markers in mediated and non-mediated communication

DMs, also referred to as *connective items* or *linking words*, are individual words or multi-word units that “are used as discourse structuring elements for ideational, rhetorical and sequential relations” (Lenk 1998: 49). The main objective of the present paper is to study the use of DMs in English as a target language, i.e. English produced by simultaneous interpreters working from Italian into English (with English as their native or active working language). In particular, the focus is placed on DM *so* with respect to its discourse use and pragmatic meaning, thus excluding its primary use and propositional meaning (Bazzanella 1994; Blakemore 2002).

Various discourse functions (interactional and textual) of this particular DM have been highlighted by a number of scholars in response to the question “[h]ow do cultural, social, situational, and textual norms have an effect on the distribution of discourse markers?” (Schiffrin 2001: 62). The use and distribution of *so* have been investigated in different contexts and communities, for instance among native speakers as a marker of participation structures (e.g. turn exchange and speaker continuation), and of cause and result (Schiffrin 1987); in interviews between native English teachers and non-native learners of English as a marker of addition and continuity (Pulcini and Furiassi 2004); and in (experimentally prompted) narratives by native speakers and learners of English (Müller 2005), first on the textual plane, as a marker of additional and more fine-grained functions such as result or consequence, main idea unit, summarising, and rewording, and second on the interactional plane, with the functions of question, request, opinion, implied result, and transition relevance place.

The present study examines the possible effect on discourse marker use and distribution of specific situational norms that are in play in simultaneous interpreter-mediated settings. In previous T&I studies, contrasting effects have been documented. In audiovisual translation, for instance, deletion of DMs in target texts is reported (Chaume 2004), while additions have been recorded in literary translation (Hauge 2014). Although these findings refer to written target texts, it is interesting to note that shifts in DM use have also been observed. In legal interpreting, the treatment of DMs has received special attention owing to the potentially serious consequences deriving from their systematic omission by interpreters (for *well*, *see* and *now* in courtroom discourse, see Hale 1999). Conversely, Blakemore and Gallai (2014) investigated additions of DMs *well* and *so* in interpreters’ renditions of police interviews and framed these as a means “to encourage the audience to follow an inferential path which results in the representation of thoughts and thought processes of someone other than the interpreter” (Blakemore and Gallai 2014: 115), rather than as an explicitation-related device. As regards conference interpreting, within the specific context of the European Parliament (EP), Defrancq, Plevoets, and Magnifico (2015) examined both simultaneous interpretations and translations of EP plenary debates and verbatim reports. In their data, both omissions and additions of DMs were recorded, with interpreters omitting but also adding more DMs than translators. The authors point out that “the very fact that additions occur in interpretations is surprising *per se* ... as [making additions] requires cognitive resources that are already scarce” (p. 215), and they call for further investigations to ascertain whether this is evidence of “chaining strategies” or “delaying strategies” adopted by the interpreters (p. 217). The variety and frequency of interpreter-generated DMs have also been studied to explore interpreters’ style, as “[e]ach interpreter appears to have his or her own stock-in-trade, made up of a finite number of DMs” (Straniero Sergio 2012: 220).

While these studies adopted both monolingual comparable and parallel perspectives, here an exclusively parallel perspective is applied, due to space limitations. Additionally, the present analysis focuses on the relationship between interpretations (or target texts, henceforth TTs) and the related original speeches (or source texts, henceforth STs) only insofar as this helps explore the use of DM *so* in English as a target language.

3. The DIRSI corpus

The DIRSI corpus is made up of four sub-corpora including original speeches in Italian and English along with their simultaneous renditions. It includes approximately 136,000 running words from 9.5 hours of selected recordings from three different conferences, two about cystic fibrosis (CFF4 and CFF5) and one about elderly care (ELSA). Only the opening, presentation and closing sessions are transcribed in the corpus, thus excluding debates and question-and-answer sessions, which have different interactional formats (dialogic vs. monologic). The corpus is generally balanced, with each sub-corpus containing on average 33,900 words. As Table 1 shows, the largest sub-corpus contains the English STs (37,249 words¹, mostly paper presentations or lectures) while the smallest contains the English TTs (31,510 words, from Italian source texts, which range from paper presentations to opening/closing remarks, floor allocation, and procedural or housekeeping announcements).

TABLE 1

Five professional interpreters in total are represented in the corpus. In terms of working languages, four interpreters have Italian as their A (native) language and English as their B (active working) language (IT-01; IT-02; IT-03; IT-04); one interpreter has English as their A language and Italian as their B language (UK- 01). Their overall working time and distribution as represented in the corpus are summarised in Table 2. The figures in bold refer to the target texts considered in the analysis (in total: 280 minutes; 31,510 words).

TABLE 2

4. Methodology

The quantitative analysis was carried out by automatically extracting all the occurrences of *so* from the English TTs sub-corpus via the Corpus Workbench (CWB) suite of corpus query tools (Christ 1994). A qualitative analysis was performed by scrutinising the data on the online corpus interface (LLI-UAM²), where transcripts are aligned with the corresponding audio files and ST-TT content alignment is also available. LLI-UAM queries made it possible to retrieve the relevant transcript file for each occurrence of *so*, display it aligned with its source text, and listen to the audio recording to disambiguate all those cases that appeared unclear just by reading the transcript. All the occurrences were analysed in this way and eventually classified into three different categories, i.e., translation, addition or phrasal, depending on the kind of use detected in the interpreters' TTs.

¹ The English STs sub-corpus includes both English as a native language (31,525 words) and English as a foreign language (5,724 words).

² The online corpus interface is hosted on a server of the Computational Linguistics Laboratory of the *Universidad Autónoma de Madrid* and is freely accessible for research purposes (<http://cartago.llf.uam.es/static/dir-si/dir-si.html>).

- Translation: DM *so* is used in the TT in response to an equivalent or similar unit of meaning in the ST.
- Addition: DM *so* is used in the TT autonomously and independently of the corresponding segment in the ST, thus potentially signalling processing of the ST message or other strategies deployed by the interpreter.
- Phrasal: the use of *so* in the TT is due to the presence of lexicalised expressions or grammatical constructions which require its presence, as in the case of *so as to*, *so* + adjective, *and so on*, etc.

To classify the different types of occurrences, a spreadsheet was designed with six different columns reporting the following details: number of speech event, conference code, interpreter code, translation (specifying the ST term or expression corresponding to each instance of *so* in the TT), addition (a yes/no field), and phrasal (specifying the expression or construction used in the TT). Interesting examples encountered during the analysis were also included in the file. After filling in all the details in the spreadsheet, automatic filters were used to count the total number of each type of occurrence and retrieve the information necessary to link each target expression to its source.

5. Results and discussion

5.1. Quantitative analysis

The total number of occurrences of *so* in the English target texts of DIRSI is 257. These are more or less evenly distributed among the interpreters involved, ranging from a minimum of 39 occurrences (IT-04) to a maximum of 60 occurrences (IT-02). Interpreter IT-01 is present in two different conferences, CFF4 with a small number of occurrences (just 10) and CFF5 with a higher number of occurrences (45). UK-01, the only native English interpreter, ranks second in terms of total number of occurrences (57). However, these general results on their own are meaningless, as they are strictly dependent on each interpreter's working time and on the features of the STs they had to translate. What is interesting here is that all the interpreters are represented to some extent.

TABLE 3

The breakdown of all the occurrences per conference, per interpreter, and by analytical category (see §4) is displayed in Table 3 (the value in brackets is the normalised frequency per 10,000 words). Since additions are especially noteworthy, they are expressed not only in terms of number of occurrences, but also in terms of relative frequency per minute, which highlights 'how often' *so* was added by each interpreter.

A glance at the totals reported in Table 3 shows that more than half of the occurrences of *so* are produced by the interpreters in response to a similar or equivalent unit of meaning in the ST. However, by the same token, 30% of all the occurrences in the TTs are the result of addition or further processing of the ST by the interpreters, thus confirming similar results reported in the literature (§2). Finally, almost 20% of all the occurrences are due to the use of expressions or phrases for which the use of *so* is mandatory, though they do not function as DMs.

A more detailed examination of the Translation category revealed that the corresponding units of meaning in the STs include a limited range of words or expressions in Italian. The more frequent ones are: "quindi" (66), "e quindi" (12), "allora" (7), "ecco" (7), "così" (5), "perciò" (5), "cioè" (4), "per cui" (4), "appunto" (3), "effettivamente" (2). In addition to these, there are 15 further items occurring only once.

Moving on to the Addition category, what emerges is that not all the interpreters used *so* strategically. For instance, IT-01 displays 0 occurrences of addition in CFF4 (though this is counterbalanced by the use of additional *so* in the other conference where the same interpreter worked, i.e. CFF5), and IT-04 uses additional *so* only in four cases out of the total of 38. On the other hand, UK-01 and IT-03 show a number of instances of interpreter-generated *so* which, when compared to the relevant total number of occurrences, is substantial. The third category (i.e. Phrasal) is largely accounted for by the following constructions: “so + *adjective*” (14), “[and] so on” (12), “so that + *subordinate clause*” (8).

5.2. Qualitative analysis

The second analytical category (i.e. Addition) is obviously the most revealing in relation to the research questions. Here it was possible to identify several different uses of *so*. In some cases it appears to be used to help manage the structure of the ST; in other cases it comes with extra information or explicitation of the ST message; other instances are seemingly due to the reaction of the interpreters as they grasp the meaning of the source speaker’s message and verbalise this process of understanding. Below are some examples of these different uses. The examples show the transcripts in tabular form, with the Italian ST on the left and the English TT on the right. The time codes embedded in the transcripts are not indicative of interpreters’ *décalage* (i.e. the time lag between ST and TT production), serving only as references for audio alignment. Highlighting and underlining are meant to guide the reader in establishing visual correspondence between ST and TT more conveniently.

5.2.1. Target text chunking

Example (1) is taken from the ELSA conference, involving TT 004 produced by interpreter IT-03. The excerpt is the final part of the longest lecture (30’) presented in the opening session of the conference. Though delivered at an average of 115 words per minute³, it is full of abstract nouns and complex syntax. Indeed, at this stage the interpreter is at first lagging behind from the previous segment (there are some omissions of possibly redundant items), through struggling to keep TT production under control. Control is finally reasserted by chunking the last subordinate clause and making it a main clause introduced by *so*:

Example 1)

<p>poi credo che il tema della partecipazione vada declinato almeno a due livelli 27:37 // noi lo abbiamo spesso declinato sul versantepubblico 27:41 // ma vi è anche u- una questione molto più stretta personale individuale che dovrebbe porsi nel nostro operare qu tidiano 27:50 // <u>ed è il tema del rapporto</u> fra la valutazione tecnica il ruolo dei tecnici e il grado di libertà di autodeterminazione delle persone rispetto a quello che non è un piano di cura ma è essenzialmente un piano di vita 28:4 //</p>	<p>and then the two levels of participation the public level and the personal level 28:21 // <u>that is to say the le- relationship</u> between the role of experts and the degree of freedom and self-determination of people 28:35 // <u>so</u> we shouldn’t be seeing these themes in terms of a a care health care plan 28:44 // we should be thinking in terms of a life plan 28:48</p>
--	--

³ In simultaneous interpreting, an ST speech rate of 100-120 words per minute is considered optimal (Gerver 1969/2002) or easy (Setton & Dawrant 2016: 52).

There is a similar example in the following TT (Example 2) produced by interpreter UK-01 during the CFF4 conference (TT 097):

Example 2)

ma io credo che questo sia stato sufficientemente approfondito già dalla lettura del professor Durie </ Doering/> e e a meno che ci siano delle chiari- delle necessità di chiarificazione potrebbero eventualmente essere accantonati per il momento 1:0 //	but I think that this has already been sufficiently addressed by the lecture of professor Durie 0:51 // and I don't know if you have clear need as an audience for further detail in this respect 0:58 // so we could move on to other things for the moment 0:60 //
--	---

The ST shows several speech production inaccuracies, e.g. incorrect pronunciation of a proper name (“Durie” instead of “Doering”), an unfinished word (“chiari-”) followed by a reformulation, as well as a lack of cohesion and grammatical concord (“necessità” vs. “accantonati”). The interpreter tidies up the form of the ST to deliver a smoother TT. Again, the use of *so* seems to favour this chunking strategy and streamlining process.

5.2.2. Explication

Example (3) shows how the use of additional *so* comes with the presentation of information that had been already mentioned before but is re-stated in a more explicit way by the interpreter. The excerpt below is from TT 009 by interpreter UK-01 in the CFF4 conference.

Example 3)

prima però mi è stato detto che come ormai si usa ahimè c'è una pausa per la pubblicità e quindi inviterei un attimo il dottor Giulio Cabrini 0:38 // c'è sì 0:40 // deve dare un breve comunicato molto importante che si inserisce solo parzialmente però in questo discorso	there is a short break our # Giulio Cabrini has to give a brief announcement before we can actually have the discussion 0:47 // and so there is a brief interruption
--	---

Looking at the TT, the first part of the interpreter's delivery contains the main message expressed in the ST. However, the interpreter seems to feel the need to reformulate it in an alternative way, perhaps to prevent the audience from having the impression that important information is being omitted. Indeed, the source speaker continues to speak, adding some somewhat vague remarks, and the interpreter follows through by utilising *so* to add more explicit details of what is going to happen as announced by the source speaker.

Similar instances can be found in example (4), an excerpt from TT 004 spoken by interpreter IT-03 in the ELSA conference. The interpreter rephrases the information, using *so* to introduce a kind of explanation.

Example 4)

<p><u>quindi</u> la terza dimensione oltre quella appunto economica e sociale diventa quella ambientale 9:7 // <u>per la prima volta si inizia a parlare di attenzione alle relazioni intergenerazionali</u> 9:13 // e viene posto [e questo ci riguarda] forte attenzione su due aspetti 9:19 // quello dell'integrazione delle politiche e delle v- della necessaria valutazione dell'impatto di sostenibilità preventivo e intersettoriale 9:29 //</p>	<p><u>so</u> sustainability acquired an environmental meaning as well 9:20 // <u>for the first time intergeneration relationships are mentioned</u> 9:28 so the relationships between different generations of people 9:31 // and then the need to integrate policies and the need to assess sustainability impacts on sustainability 9:42 //</p>
---	--

Given the subsequent embedded clause used by the source speaker (“e questo ci riguarda”) along with the lack of grammatical cohesion (“viene posto” vs. “attenzione”), this kind of explicitation may also be due to the attempt by the interpreter to generate more units of meaning from the incoming source message (i.e. a delaying strategy). The interpreter also relies on the slides used by the presenter, where further information can be referenced, for the benefit of the audience.

5.2.3. Adding extra information

In example (5), interpreter-generated *so* in the English TTs occurs with the addition of information that is not explicitly present in the ST. The excerpt below is taken from TT 004 spoken by interpreter IT-03 during the ELSA conference:

Example 5)

<p>innanzitutto un attimo un un indice di quelle che saranno le le mie riflessioni centrate su sostanzialmente tre concetti fondamentali che sono al centro dell'incontro di oggi 1:13 // l'integrazione la partnership la partecipazione 1:18 //</p>	<p>I'll give you a brief overview of my thoughts this morning 1:5 // so if you can see on the slide I'm going to be dealing with three main topics 1:13 that is to say integration partnership and participation 1:19 //</p>
---	---

In this example the interpreter adds some situational instructions for the benefit of service users whose attention is drawn to the slides projected onto the screen, i.e. a contextual reference not present in the ST. This extra information is introduced by using *so*.

5.2.4. Other functions

In addition to introducing more explicit or new information, other occurrences of interpreter-generated *so* were observed in TT delivery whose function appeared to be to strengthen or focus the ST structure. The reasons behind such a choice by the interpreters are not entirely clear, as they seem to use *so* both as a structuring device, with a coordinating function, and as a verbalisation of successful comprehension of the ST message as it is constructed by the source speaker. Examples (6a) to (9b) below are excerpts from all the different conferences and interpreters following the presentation order of quantitative data in Table 3 above.

The first two examples under this heading (6a and 6b) come from the CFF4 conference and were produced by interpreter UK-01 in TT 137a, which is one of the main lectures presented at that conference:

Example 6a)

dall'altra parte invece dalla parte del professionista circa il settanta per cento dice che l'utilizzo la ricerca di informazioni sanitarie su internet possa aumentare il rischio di autogestione 8:8 // dipende dal punto di vista e dal tipo di autogestione ovviamente 8:14 //	from the professional's perspective about seventy per cent of professionals say that the research for health-based information on internet can have increased risks involved in self management 8:8 // so it all depends on what your point of view is and also it depends on the type of self management 8:12 //
--	--

Example 6b)

probabilmente però è anche a causa dell'aumento di insomma dell'ottimizzazione dei motori di ricerca 29:10 // continuando con la nostra revisione eravamo arrivati a dodicimilanovecento pagine sulla fibrosi cistica o che in qualche modo menzionassero la frase in italiano 29:21 //	probably however this is because of the optimization of search engines 29:8 // so now continuing with our revision we were able to get to twelve thousand of pages twelve thousand nine hundred pages with the title including the words fibrosi cistica 29:20 //
---	--

In both examples (6a) and (6b), the interpreter adds DM *so* without changing the ST structure and without providing new or more explicit information. When listening to the recording, this DM does not seem to signal a delaying strategy either. Among the possible functions already discussed in the literature, these additions appear to be more in line with the use of *so* as a marker of summarising and implied result (6a), and continuity (6b), although they belong more to the inferential path of the interpreter himself rather than the source speaker.

The next example (7) occurred in the TT production of interpreter IT-04 (TT 006) in the ELSA conference:

Example 7)

in aggiunta oggi abbiamo gli amici stranieri e e anche la loro presenza mi fa molto molto piacere 1:9 // siamo qui in tanti // io provo a dare [un attimo se ci riesco un po'] il senso di questa giornata 1:18 //	and in addition to that today we have our foreign guests here with us 1:8 // and the fact that they are here is something that I find even more pleasant 1:14 // so I don't know whether I manage but I'd like to provide you with a general overview of what we're going to do today and why we are here today 1:26 //
--	--

When listening to the recording, the analyst would tend to perceive this addition as a delaying strategy, and this might be due to the hesitant pace of the source speaker and the embedded clause separating the verb of the main clause ("io provo a dare") from its object ("il senso di questa giornata"). At that point, the interpreter seems to need greater focus on the upcoming structure of her rendition.

The next example (8) is from the CFF5 conference, interpreter IT- 01 (TT 050), where the interpreter omits a small unit of humour ("ho sempre più paura di prendere la scossa"). Despite this omission, the main message about the complexity of the tools used in the Genoese laboratories vs. the easy-to-understand presentation of the lecture comes across. By adding DM *so* at this particular point in the TT, the interpreter seems to verbalise his own understanding of the main message, even if he is aware of having omitted a unit of humour related to it:

Example 8)

io tutte le volte che vado a trovare Gino a Genova vedo degli strumenti di elettrofisiologia sempre più complicati // ho sempre più paura di prendere la scossa 0:41 // e con questa estrema complessità renderla semplice è veramente una cosa che solamente quelli bravi bravi riescono a fare 0:49 //	every time I go to Genoa and I saw very complex more and more complex tools every time I go to Genoa's laboratories so it was a pleasure to follow such a simplified and understandable presentation
--	---

The last two examples under this heading (9a and 9b) are from the CFF5 conference, interpreter IT-02, TT 012:

Example 9a)

sembra semplice oggi venire dall'America // per lei è stato molto difficile 0:22 // grazie anche per questa avventura transatlantica // prego	it seems easy to fly from America but from for her it was really hard 0:23 // so thank you for being here with us
---	--

Example 9b)

gli strumenti c- ora disponibili non permettono di dire ah ecco ora posso disegnare a tavolino una molecola che possa fare questo lavoro 2:47 // perché le informazioni sulla struttura della proteina disponibili finora che stanno crescendo man mano non sono comunque così imponenti da da f- permettere questo tipo di di di lavoro 3:1 //	the available tools da- do not allow us to say okay we need that so we design a drug that can do that // because information on the protein so far has is no not so developed now 2:57 // so we still have too little information to do this 3:1 //
---	--

In both examples, the interpreter adds *so* in a way that may be perceived as a verbalisation of her own understanding of the ST. Although in example (9b) some restructuring can be identified, these additions appear to mirror the inferential path of the source speaker.

6. Final remarks

This study examined one specific feature of English as a target language, the language spoken by simultaneous interpreters when interpreting from Italian into English. The aim was to shed light on discourse markers as used by interpreters, not so much in response to an equivalent unit of meaning in the source text but rather in a more autonomous way.

The investigation focused on DM *so* in the English target texts. It was clear that in addition to using it as an adverb, conjunction or part of lexicalised expressions in response to equivalent units of meaning in the ST, the interpreters were also adding this DM as a device to keep the English TT structure more under control (thus helping to manage their cognitive capacity) or to enhance the reception of their output by service users thanks to syntactic transformation. After all, there are some major differences between the two languages involved: English is a Germanic language with its SVO structure and is now more spoken by non-native than native speakers; Italian is a Romance, pro-drop language with far more flexible syntax. Chunking and syntax reshuffling – which affects, among other items, adverbs and DMs such as *so* – are hypothesised by theorists as helping to manage these differences (e.g. see Bartłomiejczyk 2006; Gile 2009: 205; and more specifically on interpreting from Italian into English, Snelling 1992).

This hypothesis was verified by studying the occurrences extracted from the DIRSI corpus both quantitatively and qualitatively. The analyses showed that 50% of all the occurrences of *so* found in the TTs are due to translation of an equivalent unit of meaning in the ST, while 20% are part of fixed expressions and grammatical constructions. The remaining 30% are interpreter-generated DMs, confirming similar trends encountered in other corpora, first of simultaneous interpreting (and translation) at the European Parliament, where DMs are actually omitted but also added (Defrancq, Plevoets, Magnifico2015), and second of police interpreting, where *so* is added “on the basis of her [the interpreter’s] own understanding of the utterance” (Blakemore and Gallai 2014: 116).

The specific distribution of interpreter-generated DM *so* found in the DIRSI corpus accounted for TT segmentation into more manageable units, explicitation of information already expressed in the TT (also present in the ST), addition of new information not present in the ST, and the addition of *so* alone as a possible way of strengthening the rhetorical structure of the TT, to make it more accessible to interpreting service users. Indeed, DMs are “effective as coherence indicators [...] when topical inconsistencies or topic changes seem to be threatening a coherent understanding of the overall discourse” (Lenk 1998: 46). This is all the more important considering how superdiversity is impacting audience composition and the profile of international conferencing in terms of working languages, where English, or rather many Englishes, are and will be increasingly used as a lingua franca along with the main local language (Albl-Mikasa 2014; Bendazzoli 2017).

The small size of the DIRSI corpus, and even more so of its sub-corpora, is obviously a limitation of this study and the results obtained cannot be generalised. Yet the use of corpus methods made it possible to retrieve and analyse occurrences in a systematic way, and the results can be contrasted to what is found in other interpreter-mediated communicative situations. An ST-oriented perspective is also missing in this investigation, which only focuses on use of *so* in TTs, but it would be extremely interesting to look at the occurrences of *so* in STs, and check how these were managed by the interpreters. This is just one example among many of the research opportunities afforded by the DIRSI corpus.

Acknowledgments

I am grateful to Bart Defrancq for his helpful comments on an earlier draft of this paper.

References

- AIIC, 2012, “Working Languages”, *aiic.net*, <http://aiic.net/p/4004>, last accessed June 11, 2018.
- Albl-Mikasa, Michaela, 2010, “Global English and English as a Lingua Franca (ELF): Implications for the Interpreting Profession”, *Trans-kom* 2 (3), pp. 126-48.
- Albl-Mikasa, Michaela, 2014, “The Imaginary Invalid. Conference Interpreters and English as a Lingua Franca”, *International Journal of Applied Linguistics* 24 (3), pp. 293-311.
- Bartłomiejczyk, Magdalena, 2006, “Strategies of Simultaneous Interpreting and Directionality”, *Interpreting* 8 (2), pp. 149-174.
- Bazzanella, Carla, 1994, *Le Facce del Parlare: Un Approccio Pragmatico all’Italiano Parlato*, La nuova Italia, Scandicci.
- Bendazzoli, Claudio, 2010, *Corpora e interpretazione simultanea*, Bologna, Asterisco. Also available on AlmaDL: <http://amsacta.unibo.it/2897/>, last accessed June 11, 2018.
- Bendazzoli, Claudio, 2012, “From International Conferences to Machine-readable Corpora and Back: An Ethnographic Approach to Simultaneous Interpreter-mediated Communicative Events”, In C. Falbo and F. Straniero Sergio (eds), *Breaking Ground in Corpus-Based Interpreting Studies*, Peter Lang, Frankfurt am Main, pp. 211- 30.
- Bendazzoli, Claudio, 2017, “Benefits and Drawbacks of English as a Lingua Franca and as a Working Language: The Case of Conferences Mediated by Simultaneous Interpreters”, in C. Boggio and A. Molino (eds), *English in Italy. Linguistic, Educational and Professional Challenges*, Franco Angeli, Milano, pp. 119-41.
- Blakemore, Diane and Gallai, Fabrizio 2014, “Discourse Markers in Free Indirect Style and Interpreting”, *Journal of Pragmatics* 60, pp. 106-20.
- Blakemore, Diane, 2002, *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*, CUP, Cambridge.
- Chaume, Frederic, 2004, “Discourse Markers in Audiovisual Translating”, *Meta* 49 (4), pp. 843-55.
- Christ, Oli, 1994, “A Modular and Flexible Architecture for an Integrated Corpus Query System”, *COMPLEX’94*, Budapest.
- Defrancq, Bart, Plevoets, Koen, Magnifico, Cédric, 2015, “Connective Items in Interpreting and Translation: Where do they come from?”, in J. Romero-Trillo (ed), *Yearbook of Corpus Linguistics and Pragmatics 2015*, Springer, Singapore, pp. 195-222.
- Gerver, David, 1969/2002, “The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters”, in F. Pöchhacker and M. Shlesinger (eds), *The Interpreting Studies Reader*, Routledge, London-New York, pp. 53-66.
- Gile, Daniel, 2009, *Basic Concepts and Models for Interpreter and Translator Training*, John Benjamins, Amsterdam.
- Hale, Sandra, 1999, “Interpreters’ Treatment of Discourse Markers in Courtroom Questions”, *Forensic Linguistics* 6, pp. 57-82.
- Hauge, Kjetil Rå, 2014, “Found in Translation – Discourse Markers out of the Blue”, in S. Oksefjell Ebeling, A. Grønn, K. R. Hauge and D. Santos (eds) *Corpus-based Studies in Contrastive Linguistics, Special issue of Oslo Studies in Language* 6 (1), pp. 43-52: <https://www.journals.uio.no/index.php/osla/article/view/720/808>, last accessed June 11, 2018.
- He, He, Jordan, Boyd-Graber and Daumé III, Hal, 2016, “Interpretese vs. Translationese: The Uniqueness of Human Strategies in Simultaneous Interpretation”, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 971-976.
- Kajzer-Wietrzny, Marta, 2018, “Interpretese vs. Non-native Language Use: The Case of Optional That” in M. Russo, C. Bendazzoli and B. Defrancq (eds), *Making Way in Corpus-based Interpreting Studies*, Springer, Singapore, pp. 97-113.
- Lenk, Uta, 1998, *Marking Discourse Coherence. Functions of Discourse Markers in Spoken English*, Gunter Narr Verlag, Tübingen.

- Müller, Simone, 2005, *Discourse Markers in Native and Non-native English Discourse*, John Benjamins, Amsterdam.
- Pulcini, Virginia and Furiassi, Cristiano, 2004, "Spoken Interaction and Discourse Markers in a Corpus of Learner English", in A. Partington, J. Morlwey and L. Haarman (eds), *Corpora and Discourse*, Peter Lang, Bern, pp. 107-120.
- Schiffrin, Deborah, 1987, *Discourse Markers*, CUP, Cambridge.
- Schiffrin, Deborah, 2001, "Discourse Markers: Language, Meaning, and Context", in D. Schiffrin, D. Tannen and E. H. Hamilton (eds), *The Handbook of Discourse Analysis*, Blackwell, Malden MA, pp. 54-75.
- Setton, Robin and Dawrant, Andrew, 2016, *Conference Interpreting. A Trainer's Guide*, Amsterdam-Philadelphia, John Benjamins.
- Shlesinger, Miriam, 2008, "Towards a Definition of Interpretese: An Intermodal, Corpus-based Study", in G. Hansen, A. Chesterman and H. Gerzymisch-Arbogast (eds), *Efforts and Models in Interpreting and Translation Research: A tribute to Daniel Gile*, John Benjamins, Amsterdam, pp. 237-253.
- Snelling, Clyde, 1992, *Strategies for Simultaneous Interpreters. From Romance Languages into English*, Campanotto, Udine.
- Straniero Sergio, Francesco, 2012, "Using Corpus Evidence to Discover Style in Interpreters' Performances", in F. Straniero Sergio and C. Falbo (eds), *Breaking Ground in Corpus-based Interpreting Studies*, Peter Lang, Bern, pp. 211-30.

Table 1
DIRSI Corpus size

<i>Sub-corpus</i>	<i>No. of speech events</i>	<i>No. of words</i>	<i>% of DIRSI</i>
ORG-IT	63	33,412	24.6
INT-IT-EN	63	31,510	23.2
ORG-EN	16	37,249	27.4
INT-EN-IT	16	33,664	24.8
TOTAL	158	135,835	100

Table 2
Interpreters' working time (in minutes) and speech production (no. of words) in DIRSI

<i>Conference</i>	<i>Interpreter</i>	<i>Language A</i>		<i>Language B</i>	
		<i>working time</i>	<i>no. of words</i>	<i>working time</i>	<i>no. of words</i>
CFF4	UK-01	58'	7,654	67'	8,257
CFF4	IT-01	43'	5,276	15'	1,828
ELSA	IT-03	40'	4,212	37'	3,700
ELSA	IT-04	27'	3,495	50'	5,536
CFF5	IT-01	38'	4,713	60'	6,359
CFF5	IT-02	69'	7,811	60'	6,433
Total		276' (4h 36')	33,061	289' (4h 49')	32,113

Table 3
Breakdown of the DIRSI occurrences of so in English TTs, by category

Conference	Interpreter	Translation	Addition			Other	Total	
			No. of occurrences	Relative freq. in minutes				
CFF4	UK-01	18 (24)	25 (33)	2.32	14 (17)	57 (69)		
	IT-01	7 (38)	0 (0)	0.0	3 (16)	10 (55)		
ELSA	IT-03	21 (57)	22 (59)	1.68	3 (8)	46 (124)		
	IT-04	23 (42)	4 (7)	12.5	11 (20)	38 (69)		
CFF5	IT-01	24 (38)	15 (24)	4	6 (9)	45 (71)		
	IT-02	39 (61)	12 (19)	5	9 (14)	60 (93)		
TOTAL		131 (42)	80 (25)	3.45	49 (16)	256 (81)		