English-Arabic Subtitling: A Relevance-Theoretic Approach

Sanaa Benmessaoud, University of Sharjah

Abstract

Several scholars have recently decried the dearth of research on Arabic audiovisual translation, including subtitling, calling for the exploration of appropriate theoretical frameworks that could support Arab translators in their profession. To contribute to filling in this gap, the paper will explore the usefulness of Gutt's (1991/2014) relevance-theoretic approach for English-Arabic film subtitling. The paper argues that given the "prescriptivism" in film subtitling of only translating what is deemed most relevant to the comprehension of the film dialogue, a relevance-theoretic approach best provides practitioners with a framework for making the appropriate decisions.

Keywords: *Subtitling, Arabic audiovisual translation, relevance theory, reduction*

Introduction

In 2017, a study commissioned by the MESA Europe Content Localization Council concluded that over-the-top services, such as Hulu, Amazon Prime and Netflix, were witnessing such a great boom that the volume of audiovisual translation, mainly subtitling, would exceed two billion dollars per year by 2020, including in the Middle East (Green 2018). Considering the dramatic surge in Netflix subscribers from all over the world in the first quarter of 2020 ('Netflix' 2020), this trend is set to spike even higher over the next few years, including in the Arab World. Such growing market of audiovisual translation in this region, however, has not been adequately reflected in research in Arab countries. Gamal (2019: 202), for instance, points out that despite existing research on both translation policy and translation practice in the Arab world, such research has fallen short of creating "a school of thought in Arabic translation with a developed philosophy, defined theoretical frameworks or a designed pedagogy." This, coupled with lack of interest and investment by schools and university departments of translation in the Arab world, has resulted in a dearth of research on Arabic AVT (208).

In response to this concern, the present paper explores one specific mode of audiovisual translation (AVT), namely subtitling, from English into Arabic. Applying Gutt's (2014) relevance-theoretic approach to the translation of a clip from the British film *Chicken Run* (Lord & Park 2000) into Arabic, the paper's objective is twofold. It aims to give insight into the main difficulties encountered in subtitling movies from English into Arabic, two linguistically and culturally remote languages. It also aims to highlight the usefulness of relevance theory as a tool for decision-making in subtitling and, therefore, for analysing and evaluating subtitles. The paper will thus focus on the two main difficulties generally associated with this mode of translation, namely the need for reduction, resulting mainly from the spatial and temporal constraints under which this translational mode is performed (Kovačič 1994; Matielo et al. 2015), and culture-specific elements. It argues that given the stringent conditions under which subtitling is practiced, and which impose the "prescriptivism" of only translating what is deemed most relevant to the comprehension of the film dialogue, a relevance-theoretic approach best provides practitioners with a framework for making the appropriate decisions.

Arabic audiovisual translation: an overview

Audiovisual translation (AVT) was born out of the need for films to conquer new linguistic markets, which locates the birth of this mode of translation in the beginning of the twentieth century. While it was the big screen that brought about AVT, it was the extraordinary advances in communication technologies, especially the Internet, video-streaming and portable players, that gave AVT momentum and significantly increased the need for it. According to Gambier (2013: 53), these technological advances have had several implications for AVT. On the one hand, they offer audiences a wider range of more specialized and personalized services, such as Pay TV and thematic TV channels, thus marking a shift from "broadcasting to narrowcasting" and creating new audiences with pronouncedly divergent needs and expectations. On the other hand, and because of the globalizing effect of the Internet and the ever-increasing amount of content finding its way through video streaming sites to a global audience, there is at once an increase in "fansubs" and "fandubs," online fan communities that translate AV content, and a strong need for TV broadcasters and film distributors to ensure their content reaches the wider audience before these fans download it and subtitle/dub it (54). Finally, automation is increasingly digitizing the profession and deeply changing the practice.

As a result of this significant increase in the demand for AVT, and of the deep changes this industry has undergone over the past couple of decades, AVT developed so quickly and so deeply from a "virgin area of research" as Delabastita (1989: 202) described it not that long ago, into what many scholars have started to consider as a discipline in its own right (see, for instance, Pérez González 2014). The varied terminology that has been used over the years to describe and discuss AVT, from the restrictive "film translation" and "screen translation" to the broader "versioning" and "multimedia translation," highlights not only the impact of technological developments on the practice but also what Gambier (2013: 46) aptly describes as "the vitality of the research domain and the diversity of practices."

The impact of technological developments, especially Internet penetration, the advent of social media platforms and the proliferation of portable players, coupled with political developments in the region, resulted in a similar diversity of AVT practices in the Arab world, albeit to a lesser extent due to high illiteracy rates. AVT, especially subtitling, is used as much for information and entertainment (see, for instance, Eldalees, Al-Adwan & Yahiaoui, 2017 on fansubbing), as for activism (cf. Baker 2016). This diversity of practices, however, has not been matched by a vitality of research. In his detailed and comprehensive account of the situation of audiovisual translation in the Arab World, Gamal (2019: 208) decries what he sees as a "dearth of publications on the subject despite the importance of language transfer on screen, particularly as screens dominate the way millions in the Arab World live, study, work, communicate socially and organize political opposition." While he acknowledges that "some academics at Arab universities have […] responded to the noticeable emergence of audiovisual translation studies," he points out that this response remains minimal and reflects nothing more than a "passing academic interest."

Gamal (2019: 209) argues that such lack of engagement with AVT studies on the part of Arab scholars is "directly linked to the absence of adequate theoretical frameworks." Without such frameworks, Arabic AVT will not "grow in its own environment to be professionally relevant and socially responsible." Echoing Gamal, Khuddro (2018: 20) contends that Arabic AVT is "still a relatively young field in translation studies," a field that will only expand with more research undergirded by the main theories and approaches to

translation, including the polysystems theory, the functional approaches, critical discourse analysis and relevance theory.

This lack of scholarly engagement and the need for research anchored in sound theoretical frameworks affect not only dubbing—a practice that only came into prominence and started garnering more audience acceptance in the Arab world at the turn of the twenty-first century with the dubbing of Turkish drama (Gamal 2019)—but also subtitling, a practice that "has been established as the preferred mode of film translation" ever since the arrival of the talking cinema to the Arab world (Gamal 2008: 8). More importantly, a quick review of the most recent research on Arabic subtitling gives credence to Khuddro's concerns above, as it reveals that very few studies have been conducted from within the framework of the main translation theories and approaches. One specific theoretic approach that is conspicuously absent is the relevance-theoretic approach. Thus, Hussain and Khuddro (2016a) develop a model based on de Beaugrande and Dressler's approach to help subtitlers with their decision-making process, while Al Harthi (2016) grapples with humour in subtitling by drawing on the general theory of verbal humour. As to Hussain and Khuddro (2016b), they address issues associated with AVT, with specific focus on linguistic or factual errors that might exist in the audiovisual source text (ST) and the type of "mediation" necessary to deal with such errors. They fall short, however, of grounding their study in any theoretical framework.

The discussion below is aimed precisely at addressing the gaps in literature identified above, by exploring the usefulness of Gutt's relevance-theoretic approach to English-Arabic subtitling.

Interlingual subtitling: a more complex translation form

Gottlieb (2012: 37) defines subtitling as a "diamesic translation in polysemiotic media […] *in the form of one or more lines of written text presented on the screen in sync with the original verbal content*" (emphasis in the original). It is, indeed, diamesic insofar as it involves transfer not only from one language to another, but also from one mode to another, i.e. from speech to writing. But it is Gottlieb's (1992: 162) earlier definition that best brings out the complexity of this form of translation: he defined it as an "additive, immediate, synchronous and polymedial" translation. It is additive because instead of replacing the original message, it adds a new verbal visual element to the visual channel of the film, thus creating a tension between what is shown on the screen and the information contained in the subtitle. Subtitles are immediate since viewers cannot control them or re-read previous subtitles. They are also synchronous in that they are presented simultaneously with the original film and dialogue, which calls for synchronization with both the image and the sound. Finally, they are polymedial since they are part of the original message of the film, conveyed through other parallel channels, namely the non-verbal visual and the verbal and non-verbal sound channels. Understanding how these channels, primarily the verbal sound one, i.e. dialogue, contribute to the meaning of the film is a first step towards producing felicitous subtitles.

In his revisited audience design model, Bell (2001) justly maintains that speakers design their speech to accommodate their addressees. Since Bell's model has been conceived based on observations of shifts in news language style in broadcast media, it can account for style shifts in mass communication, in general. Applied to film dialogue, this framework implies that screenwriters design characters' speech in such a way as to cater not so much to the interlocutors on screen as to the target audience. Lending credence to this claim in her seminal

book-length study *Overhearing Film Dialogue*, Kozloff (2000: 15-18) asserts that film dialogue, as an integral part of the narrative, "has been purposely designed for the viewers to overhear", and that all the features of spontaneous speech it contains are deliberate and have an objective in the plot. This implies that translators have to account for every utterance in the dialogue by giving "due respect to the original creator of the text and what he/she intended to convey, even if only through inference or implicature" (Taylor 2000: 7).

Accordingly, subtitling is a translation where "the speech act is in focus; verbal intentions […] are more important than atomized lexical elements" (Gottlieb 1994a: 104). However, as a "diasemiotic translation" (Baker 2003: 245), subtitling makes the task of getting all the "intentions" of screenwriters and speakers on screen across the boundaries of language and mode, a form of tightrope walking. Indeed, subtitles are limited both in time and space by the space on screen, the audience's reading speed, the size of the original utterance, the pace of the dialogue and the specificities of the source and target languages (De Linde & Kay 2016: 6). Subtitles are further constrained by their "additive" feature. Instead of replacing the message, they are superimposed on the verbal visual channel of the film, becoming thus part of the whole message conveyed through four different channels, namely the verbal and non-verbal auditory channels and the verbal and non-verbal visual channels. This necessarily creates a tension between the image and sound on screen, on the one hand, and the subtitle, on the other. It also results in what Gottlieb (1997: 219) identified as an "intersemiotic feedback" from the visual and sound track, which can be positive at times in that it may include redundant elements that make reduction easy, and negative, at other times, in that it may further constrain the margin of manoeuvre left to the translator. Finally, the additive nature of subtitles brings about what Törnqvist (1995: 49) calls the "gossiping effect." Indeed, in subtitling, the source text is constantly and immediately available to the audience alongside the target text, i.e. the subtitles. The latter are therefore open to the scrutiny of those viewers with knowledge of the source language, thus putting additional strain on the translator.

Gutt's relevance-theoretic approach

These constraints become even tighter in the subtitling of films, a form of translation that has been considered by many translation scholars as a subfield within the larger field of literary translation (cf. Snell-Hornby 1995). Indeed, like literary translation, (interlingual) film subtitling requires the translator to move not only from one language to another, but also from one cultural and ideological system to another. The translator is thus constantly engaged in a process of negotiation and balance, and constantly making decisions, not only on how to translate and get cultural and intertextual references across within a very limited physical space, but also on what to translate, how to deal with culture-specific references that need background information for easy processing, and what to leave out precisely to allow for any necessary additions, all without disrupting the viewing experience. Indeed, any overt translation or any translation that requires too much processing against a moving image and source text will interrupt the suspension of disbelief so very necessary for the success of a film and the effectiveness of film dialogue. This specificity of (film) subtitling imposes on the translator what Fawcett (1996: 78) rightly described as "the prescriptivism of translating only what is most relevant." In other words, in the decision-making process that subtitling is, the overriding value is relevance. One theoretical approach that can both provide an account for this specific

translational mode where relevance is crucial, and help practitioners with their decision-making is the relevance-theoretic approach as conceived by Gutt (1991/2014).

In his *Translation and Relevance: Cognition and Context* (1991/2014), Gutt proposes what he calls a "unified account of translation," based on Sperber and Wilson's (1986) Relevance Theory (RT). Central to the latter is the principle of optimal relevance whereby speakers have a communicative intention and what they communicate is expected to be optimally relevant to hearers. Utterances are relevant in a given context to the extent that they have large contextual effects in that context, and that these effects can be recovered with small processing effort (125). In relevance-theoretic terms, the context is "a subset of the hearer's assumptions about the world" (Sperber & Wilson 1986: 15). Also central to RT is the distinction between descriptive and interpretive use of language. A speaker is said to use language descriptively when his/her utterance is (understood as) a true representation of what he/she believes to be true. By contrast, a speaker is said to use language interpretively when his/her utterance is a representation of what someone else said, thought or presented as true.

Drawing on these notions, Gutt (2014: 107) maintains that translation is an instance of interlingual interpretive use where the target text is presumed to interpretively resemble the original in "respects that make it adequately relevant to the receptor language audience". This can only happen if the translation provides "adequate contextual effects" and conveys "the intended interpretation without putting the audience to unnecessary processing effort" (107). According to Gutt, this relevance-theoretic approach to translation is helpful not only for translation theorists, but also for translators in that it provides a clear insight into the relationship that obtains between the original and its translation (107). Giving this claim credence, Smith (2002: 115) asserts that "empowering translators to make right decisions is Gutt's primary contribution." I would add that Gutt's conceptualization of this relationship transcends traditional understandings of fidelity and equivalence by shifting the focus away from the original to the product of the translation and its target context. It also transcends restrictive binary approaches, such as the one proposed by Venuti (1995 and 1998), for instance, by bringing out the decisive role of context in decision-making.

Like all other theoretical approaches to translation, Gutt's received its share of criticism. Talking about Gutt's relevance-theoretic approach as it specifically relates to film translation, Fawcett (1996), for instance, takes Gutt to task for a few inconsistencies. Indeed, Gutt (2014: 129) distinguishes between "translations where the translator is free to elaborate or summarize", which he terms "indirect" translations, and those translations where the translator "has to somehow stick to explicit contents of the original," and which he deems "proper." This distinction would put film translation outside the scope of translation "proper" for which Gutt proposes his relevance-theoretic approach, since film translation often requires considerable summarization or reduction. Fawcett (1996: 79), however, dismisses this distinction as "ironical" insofar as relevance theory "very clearly applies" to film translation, too. More importantly, Gutt (1991/2014: 122) asserts that "the principle of relevance can also be seen behind guidelines given for oral translation (simultaneous interpretation)." If this is the case, Fawcett (1996: 79) rightly points out, then Gutt's theoretical approach "must clearly apply to film translation, since what is said of interpreting is, if anything, even more true of film with its multiple semiotic channels" (Fawcett 1996: 79). In fact, Fawcett aptly argues that the wide range of adaptations necessary in such "indirect translations" as film translation are all dictated by the need to "offer adequate contextual effects," which is one of the main principles of Gutt's approach (79).

Many of the key premises underpinning Gutt's approach can indeed soundly account for subtitling. For instance, the notion of similar contextual effects seems to comply with Gottlieb's (1994a: 256) contention that the ideal in subtitling "would be achieving the same effect on the audience as the one the original audience experienced". Likewise, the notion of processing effort is particularly valid in subtitling where translation is determined by "the balance between the effort required by the viewer to process an item, and its relevance for understanding the film narrative" (Díaz-Cintas and Remael 2014: 113). As a result, several scholars have, over the years, brought to the fore the usefulness of relevance theory to AVT. Thus, Chaume (2008: 134) argues that "[a]mplification and reduction techniques must be monitored by relevance theory (Gutt 1991/2014) and by conventions to which different audiovisual genres are subject in each culture and epoch." Likewise, and speaking more specifically about subtitling, Díaz-Cintas and Remael (2014: 148) maintain that the approach is

> quite useful for analyzing and explaining the logic of subtitling omissions, which cannot simply be put down to linguistic factors. It is the balance between the effort required by the viewer to process an item, and its relevance for the understanding of the film narrative that determines whether or not it is to be included in the translation.

Subtitling *Chicken Run*

The researcher has thus chosen a relevance-theoretic approach to inform the subtitling into Arabic of a 5-minute clip (see appendix) from the British film *Chicken Run* (Lord & Park: 2000). The film is very interesting in that it mixes three genres, namely thriller, action and romance, and is a claymation cartoon. However, *Chicken Run* can fit within a general and broader genre, i.e. the family film since it is "a treat for adults and children alike" (Hawkes 2015). It tells the story of chickens trapped in a poultry farm, as they fight for their freedom. They believe they found help in the character of a rooster, named Rocky, who crash-lands in the farm and is, in all appearances, a flying rooster. Besides, characters' personalities in *Chicken Run* and their relationships are central to the plot. Consequently, the dialogue is full of interpersonal elements and is teeming with humorous utterances and culture-specific references, including allusions to WWII and to other films. Because it is drawing on the action film, the pace of the dialogue can get very fast.

All these features are bound to impinge on the process of subtitling the film. The fast pace, for instance, will necessarily call for substantial reductions, for two main reasons. First, the audience includes very young viewers whose reading pace is not fast. Second, subtitling imposes great spatial limitations on the number of characters allowed on the screen. Díaz-Cintas and Remael (2014) maintain that the number of characters per line of subtitle on screen varies across alphabets, and that while the maximum is 37 characters for English, it ranges from 34 to 36 characters per line in the case of Arabic (85). Al-Junaydi (2012), however, finds issue with such restrictions for Arabic subtitling. Although she (14) concedes that, in practice, there are no norms governing the maximum number of characters per line in Arabic subtitling, with the number varying from as few characters as 26 to as many as 62, she (15) rightly argues that since Arabic texts are generally "more condensed than those in English," then there is no valid reason for Arabic subtitling to be restricted to less than the 37-character limit allowed for English subtitling. It is this limit that will be observed in the present study, which will necessarily call for significant reductions.

On the other hand, the film is anchored in the Anglo-American culture, so its subtitling for a culturally remote audience, namely an Arab audience lacking the necessary background to understand much of the culture-specific references in the movie, presents the subtitler with what Leppihalme (1997: 4) has famously dubbed "culture bumps" that have to be overcome.

Reduction in subtitling

As stated above, reductions constitute a typical feature of subtitling. In fact, according to Antonini (2005: 2013), transfer of a text from speech mode to writing mode reduces it by 40% to 75%. Gottlieb (1992: 166) distinguishes between three types of reduction: a) condensation, which is the concise reformulation of the source text; b) decimation, which is the rendition of the source text through "abridged expression, reduced content"; and c) deletion, which involves the complete omission of verbal content. As to Díaz-Cintas and Remael (2014: 164), they distinguish between two broad types of reduction in subtitling, namely partial and total reductions. Partial reduction would correspond to (a) above, as it is, according to Díaz-Cintas and Remael, achieved through a concise reformulation of the source text. Total reduction would correspond to (c), as it entails the complete elimination of what is deemed of little or no relevance to the comprehension of the source text. They point out, however, that more often than not, and especially in the case of fast speech, subtitling involves the two types of reduction, insofar as an utterance could be "deleted, or reformulated more concisely, or both" in its rendition (147). This would correspond to (b) above, since decimation, according to Gottlieb (1992: 166), is often resorted to when subtitling "fast speech of some importance".

According to the above, the subtitling into Arabic of the clip under study necessarily involved many reductions of different types. The first important instance is 1(a), a long utterance, spoken in a fast pace, thus calling for condensation in its rendering in Arabic.

1(a) Mac: … and sprained the interior tendon connecting your radius to your humerus, I gave her a wee bit of a tweak, Jimmy, and wrapped it up.

والتوى وتر مرفقك
فقمت بتقويمه يا عزيزي.

[and the tendon of your elbow got contorted, so I treated it, my dear]

1(b) Rocky: Was that English?

ماذا قالت؟

[What did she say?]

1(c) Ginger: She said you sprained your wing. She fixed it.

قالت إن جناحك التوى فعالجته.

[She said you sprained your wing so she fixed it]

Utterance 1(a) is a case of idiosyncratic speech. It contains overdetailed information and is unusually fast, which makes it rather confusing to a hearer who is not familiar with the speaker. It also contains an explicature, 'an explicitly communicated assumption' (Sperber & Wilson 1986: 182), namely that Mac 'fixed' Rocky's sprained wing, and several implicatures, i.e. implicitly communicated assumptions (182), about the character of Mac. These include that she is Scottish (the accent), that she is the learned mind of the hen-house (her medical know-

how), and that she wants Rocky to appreciate her by reducing the social distance between them (the use of informal language in 'wee bit', and of the address form 'Jimmy'[1]). All these elements combined make the whole exchange optimally relevant in that the ratio of cost—in terms of processing efforts—to benefit, in terms of contextual effects, is optimal.

For the translation of this exchange, and especially of 1(a), to be relevant, it has to interpretively resemble the source text and maintain the same cost-benefit ratio. However, to transfer all the linguistic and aural clues contained in 1(a) across language and culture barriers and from the spoken to the written mode, is impossible. Schwartz (2002) argues that, in such cases, the translator has to remedy the loss by retaining "as many features as possible" of the idiosyncratic speech. What complicates the reduction in this case is that the source text itself includes a reduction and, indeed, an intralingual translation of 1(a) in 1(c) in that the latter is a condensed and simplified reformulation of 1(a), a reformulation that flattens the character's (Mac) speech.

In relevance-theoretic terms, the suggested translation of 1(a) thus brought out those features that make for optimal relevance in this particular context on the basis of the translator's assumption of what is relevant to the audience. It allowed the latter to have access to the explicature as well as to two of the implicatures, namely Mac's scientific expertise, through the use of the medical jargon, and her eagerness to be appreciated by Rocky, through the use of "يا عزيزي", 'my dear'. The translation of the utterance also compensated for the loss of the idiosyncratic element of Mac's speech, namely speed, with the use of consonance, through the repetition of the consonants "و", "ـــقـــ" and "ف", and assonance, through the repetition of the vowel sound [a]. Indeed, repetition of sounds is known to twist tongues (cf. Nikolic & Bakaric 2002), and is, therefore, as confusing as fast speech. Such rendition in Arabic of 1(a) allowed for a more literal translation of the paraphrase in 1(c), a translation that does not repeat any of the words used in the translation of 1(a). Such repetition would have indeed reduced the relevance of the whole exchange, since it would have required large processing efforts for few contextual effects, on the part of a target text audience trying to make sense of why Rocky needed Ginger's paraphrase in the first place.

Nevertheless, the subtitle lost both the elements of the Scottish accent and slang. Schwartz (2002) asserts that some elements in the spoken language are necessarily lost in subtitling, including regional accents. As to slang, its use is constrained both by the nature of the Arabic language and by the approach adopted. Arabs use different dialects in their everyday speech. As a consequence, audiovisual translation of foreign programs is done in Standard Arabic which is understood by them all regardless of their vernacular, but which does not lend itself to colloquialism (Maluf 2004). Accordingly, to reproduce the regional accent in this clip by one specific dialect would not be optimally relevant for those who do not speak the same dialect since the recovery of the contextual implications in the use of such a dialect would demand large processing efforts from some viewers and may not be possible by others.

2(c) is the second significant case of reduction, more specifically decimation, in the subtitling of this clip:

2(a) Ginger (pointing at the poster): This is our way out of here.
هذا هو سبيلنا للنجاة من هنا!
[This is our way out of here]

[1] Slang, generic names and terms of address are in-group identity markers used under the super-strategy of "positive politeness" (Brown and Levinson 1978: 106-112), whereby the speaker claims common ground with the addressee to gain appreciation and approval (106).

2(b) Baps: We'll make posters?!
سنصنع ملصقات؟!
[We'll make posters?!]

2(c) Ginger: What's on the poster, Baps, what's on the poster. We'll fly out.
لا، بل سنطير خارجا كصاحب الصورة
[No, we will rather fly out like the rooster in the poster]

The narration in this exchange makes use of what Chaume (2004) terms 'semiotic cohesion'. In order for viewers to make complete sense of 2(c), they have to draw not only on the verbal text uttered by the character, but also on the visual elements on the screen, specifically the poster that Ginger found when Rocky crash-landed in the farm. Indeed, the poster features a picture of a rooster with a cap, and a text saying "Rocky, the Flying Rooster." A literal translation would thus make little sense to the target audience since "what is on the poster" may not be readily accessible to those in the audience who do not read and speak English. Besides, although the linguistic content of the poster backs up the utterance, as an element of the visual channel of the clip, it cannot be subtitled since a dialogue is taking place at the same moment the poster is displayed. Consequently, for the subtitle to be a relevant translation of utterance 2(c), it has to interpretively resemble the whole message that 2(c) conveys by reproducing the utterance's propositional content, the interpersonal element of persuasiveness entailed in the repetition, and the linguistic content of the poster. At the same time, the subtitle has to be of a reduced size.

Accordingly, the translation of 2(c) combined both paraphrase and deletion. The paraphrase replaced the implicature in the original by an explicature, as it made explicit what is on the poster; rendered the interpersonal element of emphasis and persuasiveness through the use of the exclamation "لا", no, for negation, together with the conjunction "بل", which means "rather" and is used in Arabic to negate a preceding statement, in this case "we will make posters" and affirm a new one, in this case "we will fly out". On the other hand, the name 'Baps' was deleted. This deletion can be accounted for in terms of the approach adopted. Unlike the address form "Jimmy" above, the only function of this vocative is to enhance the emotive effect of the utterance. This effect is already relayed by the intonation in Ginger's voice. Besides, the visual feedback on the screen already makes it clear that Ginger is addressing Baps. This element has, therefore, no significant contextual effects insofar as it does not change any contextual assumptions.

Another utterance in the source text that called for reduction in the process of subtitling is 3(c), below:

3(a) Ginger: Erm, Mr. Rhodes, is this you?
هل هذا أنت، سيد رودز؟
[is this you, Mr. Rhodes?]

3(b) Rocky: Er, who wants to know?
ولماذا تسألين؟
[and why are you asking?]

3(c) Ginger: A group of rather desperate chickens. If it is you,

then you might be the answer to our prayers.
إذا كان هذا أنت، فقد يكون
الله استجاب لدعواتنا اليائسة
[if this is you, then maybe]
[God has answered our desperate prayers]

In this exchange, we have a case of evasion through hedging. Fraser (2010: 27) defines evasion in language as an instance where "the information you receive from the speaker fails to meet your expectation." Drawing on Partington (2003), he (2010: 28) further maintains that "challenging the questioner or the source" through such hedges as the question "who wants to know?" in answer to another question, is one of the many ways evasion is realized in language (28). In the case of the exchange above, Rocky is clearly evading Ginger's question by challenging her and the other chickens asking him if he could fly. Utterance 3(c), i.e. Ginger's answer, addresses both the illocutionary force of Rocky's utterance 3(b), which is a request of information, and its perlocutionary effect, i.e. that Rocky's question is an attempt to evade giving an answer that meets Ginger's expectations.

For the translation of this exchange to interpretively resemble the source text, all while maintaining the cost/benefit ratio necessary for optimal relevance, it has to recover both the illocutionary force and perlocutionary effect of 3(b). This was done through the Arabic question "ولماذا تسألين؟", "why are you asking?". Changing the propositional content of 3(b) in the process of subtitling allowed for the reduction necessary in the subtitling of 3(c). Indeed, the translation of the latter called for both deletion and paraphrase because of the spatial restrictions. Since the part of this utterance that addresses the illocutionary force of 3(b) is less relevant to the comprehension of the exchange, especially that Rocky and the audience already know who is asking the question, it was the one that was deemed disposable and was, therefore, deleted. This deletion was made possible and unnoticeable by the change in the propositional content of 3(b). The second part of the utterance, which is the most relevant one to the comprehension of the plot, was paraphrased in such a way as to 1) reproduce the propositional meaning of the original utterance, and 2) compensate for the loss of meaning incurred from the deletion of the first part of the utterance through the use of the adjective "يائسة", i.e. desperate, to recover the idea that these chickens are so desperate for help.

Culture-bound problems in subtitling

It is noteworthy that the media-specific constraints of subtitling that impose reduction equally magnify the difficulty of culture-bound problems encountered in all types of interlingual communication. These physical constraints deprive translators of such devices used in translation proper as footnotes and the translator notes (Gottlieb, 1994b: 102). The translator has, thus, to find a rendition that offers the audience a similar balance of contextual effects and mental effort in a different cognitive environment and within the space and time available. The subtitling of the present clip presents a number of culture-bound problems. Utterance 4 is a case in point.

(4) Fowler: and he is a yank!
ثم إنه من رعاة البقر.
[And he is a cowboy]

The literal equivalent of the word “Yank” in Arabic is "أمريكي", meaning “American”. But if the propositional equivalent of “Yank” is, indeed, “American”, the word has negative connotations as it is usually used derogatorily by non-Americans to designate Americans. Besides, the fact that Fowler, an old rooster who served in the Royal Air Force, addresses Rocky as a “Yank” is assumed to be meant as a reminder of WWII. It triggers in the mind of the British viewers especially a world of associations and reminds them of the “friendly” invasion of Britain by the brave but arrogant US soldiers[2]. The translation of “Yank” by "أمريكي", i.e. American, would miss all these associations and would not allow the Arab viewers to infer similar contextual implications. More importantly, it would confuse them when they would later on hear Baps asking about Rocky’s country and the surprise of the chickens at hearing him reply “America”.

The translation of the word by “cowboy” resembles the original in respects assumed to be relevant to the target audience and to yield similar cognitive effect without much mental effort. It explicitly refers to an American and implicitly connotes in the mind of the Arab viewer with bravery, power and arrogance. Besides, the visual feedback in this particular case enhances the effect of the translation since while uttering 4, Fowler pointed at the scarf in Rocky’s neck, which looks very much like the scarves worn by cowboys in Western films, with which the Arab audience is only too familiar. This choice finds further justification in Nord’s functionalist belief (2016: 10) that in cases of connotations and implicitness, the translator has to prioritize function(s) of the target text in the target context over preservation of “meaning or sense in spite of different conditions in source and target communicative situation”.

The second example of cultural problems is a combination of allusion and wordplay:

(5) Fowler: Overpaid, oversexed and over here!

كثرة مال وصحة... وقلة حمد

[**Too much** money and health but **little** contentment]

This utterance was used as a jibe at the American soldiers during the WWII (Hogenboom 2012). Because they were paid much more than the British soldiers, they could afford to entertain women more than the British. Accordingly, the use of this expression in the dialogue is clearly intended by the screenwriters as an allusion to WWII.

Allusions are a form of intertextuality aimed at triggering associations in the mind of the audience (Leppihalme 1997: 7-8). Drawing on Gutt’s view, Leppihalme (8) argues that allusions can be seen as a “message or stimulus which the communicator sends, and it is up to the receiver to find the intended referent”. While the Anglo-American viewer is assumed by screenwriters to be able to “find the intended referent”, the average Arab viewer cannot possibly identify the allusion and draw the same inference since there is no shared cognitive environment between him/her and the Anglo-American viewer. It follows that while utterance (5) is optimally relevant, and thus coherent for Anglo-American viewers, its literal translation would fall short in terms of adequate relevance.

Besides, utterance (5) is a case of wordplay based on homophony. Gottlieb (1997: 223) argues that this type of wordplay suffers the most in the process of subtitling, asserting, however, that the loss incurred can be compensated despite the media-related constraints of

[2] See Reynolds (2000) for an excellent insight into the American-British encounter during the “American Occupation of Britain” in 1942-1945, and the stereotypical perceptions the British had of the American invaders and vice versa.

subtitling. In fact, Delabastita (cited in Schwartz 2002) maintains that wordplays can be reproduced by a rhetoric device that compensates the lost stylistic effect.

Consequently, and to have similar contextual effects which the Arab audience can recover without much processing effort, the translation of utterance (5) reproduced part of the propositional content, avoiding at the same time the overt sexual reference, which the Arab audience would not expect in a family film. It captured the bitter connotation associated with the original and compensated for the stylistic loss by means of antithesis, a rhetoric device common in Arabic.

The translation of utterances (4) and (5) may not reproduce the association with WWII in the mind of the Arab audience. For though average Arab viewers, including adolescents, know about the role of the US in this war, they are not expected to know such culture-bound expressions. Gutt (2014) suggests that it is erroneous to believe that translation can give the target audience access to all the layers of meaning in the original. Besides, the loss at this level does not adversely impair the target audience's understanding of the plot.

Conclusion

While interlingual subtitling shares many commonalities with other types of translation proper, including literary translation, the many physical restrictions under which it is performed do set it distinctly apart. So much apart, in fact, that Díaz-Cintas (2003) has famously dubbed it the "vulnerable translation." In such translation, often calling for drastic interventions, especially when moving between two completely different cultural systems, the translator has to make sure that none of his/her interventions affect the cohesion and coherence of the dialogue. More important, the translator has to ensure none of these interventions alert the audience to the translational act, an act that is, paradoxically, visually foregrounded in subtitling. With its emphasis on the principle of optimal relevance, and the concept of interpretive resemblance rather than equivalence, Gutt's relevance-theoretic approach provides practitioners with the necessary theoretical framework to deal with such stringent conditions and the "prescriptivism" of relevance they impose on translators. As such, the approach is also of particular relevance to practitioners of Arabic audiovisual translation which Gamal (2007: 85) astutely describes as "an industry without a profession," precisely because there is still little academic research to support the profession and guide translators.

References:

Al Harthi, Adel. 2016. Challenges and strategies of subtitling humour. PhD thesis, University of Salford, Salford, UK.

Al-Junaydi, Maysa A. 2012. Towards English/Arabic subtitling standards. MA thesis, Al-Quds University, Jerusalem, Palestine.

Antonini, Rachele. 2005. The perception of subtitled humor in Italy. *International Journal of Humor Research* 18(2). 209-225.

Baker, Mona. 2003. *Routledge encyclopaedia of translation studies*. London, New York: Routledge.

Baker, Mona. 2016. The prefigurative politics of translation in place-based movements of protest. *Translator* 22(1). 1-21. doi:10.1080/13556509.2016.1148438

Bell, Allan. 2001. Back in style: Reworking audience design. In Eckert, Penelope & Rickford, John (eds.), *Style and sociolinguistic variation,* 139-169. Cambridge: Cambridge University.

Brown, Penelope & Levinson, Stephen C. 1978. Universals in language usage: Politeness phenomena. In Goody, Esther N. (ed.), *Questions and politeness: Strategies in social interaction*, 56-311. Cambridge and London: Cambridge University Press.
Chaume, Frederic. 2004. *Cine y traducción*. Madrid: Cátedra.
Chaume, Frederic. 2008. Teaching synchronisation in a dubbing course. In Diaz-Cintas, Jorge (ed.), *The didactics of audiovisual translation*, 129-140. Amsterdam, Philadelphia: John Benjamins Publishing Company.
Delabastita, Dirk. 1989. Translation and mass-communication: film and TV translation as evidence of cultural dynamics. *Babel* 35(4). 193-218.
De Linde, Zoe & Kay, Neil. 2016. *The semiotics of subtitling*. Manchester: St. Jerome.
Díaz-Cintas, Jorge. 2003. *Teoría y práctica de la subtitulación: inglés – español*. Barcelona: Ariel.
Díaz-Cintas, Jorge, Remael, Aline. 2014. *Audiovisual translation: subtitling*. New York: Routledge.
Eldalees, Hani A., Al-Adwan, Amer & Yahiaoui, Rashid. 2017. Fansubbing in the Arab World: modus operandi and prospects. *AWEJ* 1(1). 48-64. DOI: 10.24093/awejtls/vol1no1.4
Fawcett, Peter D. 1996. Translating film. In Harris, G (ed.). *On translating French literature and film*, 65-88. Amsterdam, Atlanta: Rodopi.
Fraser, Bruce. 2010. Pragmatic competence: the case of hedging. In Kaltenböck, Gunther, Mihatosch, Wiltrud, M. & Schneider, Stefan (eds.), *New approaches to hedging*. Bingley: Emerald Group Publishing.
Gamal, Mohammad Y. 2007. Audiovisual translation in the Arab World: A changing scene. *Translation Watch Quarterly* 3(2). 78-95.
Gamal, Mohammad Y. 2008. Egypt's audiovisual translation scene. In *Arab media and society journal*. (www.arabmediasociety.com/?article=675) (Accessed 2020-08-18).
Gamal, Mohammad Y. 2019. Audiovisual translation studies in the Arab world: the road ahead. In Hanna, Sameh, El-Farahaty, Hanem & Khalifa, Abdel-Wahab (eds.), *The Routledge handbook of Arabic translation*, 200-219. London: Routledge.
Gambier, Yves. 2013. The position of audiovisual translation studies. In Bartrina, Francesca & Millan, Carmen (eds.), *The Routledge handbook of translation studies*. New York: Routledge.
Gottlieb, Henrik. 1992. Subtitling – a new university discipline. In Dollerup, Cay & Loddegaard, Annette (eds.), *Teaching translation and interpreting: training, talent and experience*, 161-170. Amsterdam and Philadelphia: John Benjamins Publishing Company.
Gottlieb, Henrik. 1994a. Subtitling: Diagonal translation. *Perspectives: Studies in Translatology* 2(1). 101-121.
Gottlieb, Henrik. 1994b. Subtitling: people translating people. In Dollerup, Cay & Lindergaard, Annette (eds.), *Teaching translation and interpreting 2: Insights, aims and visions*, 261-274. Amsterdam and Philadelphia: John Benjamins Publishing Company.
Gottlieb, Henrik. 1997. On the polysemiotics of subtitling wordplay. In Delabastita, Dirk (ed.), *Traductio: Essays on punning and translation*, 207-232. Manchester: St. Jerome.
Gottlieb, Henrik. 2012. Subtitles – readable dialogue. In Perego, E (ed.), *Eye-tracking in audiovisual translation*, 37-81. Roma: Arcane.
Green, Stuart. 2018. How digital demand is disrupting dubbing. *M&E Journal* (https://www.mesalliance.org/2018/03/15/journal-digital-demand-disrupting-dubbing) (Accessed 2020-08-23).
Gutt, Ernst-August. 1991/2014. *Translation and relevance: cognition and context*, Manchester and Boston: St. Jerome Publishing.
Hawkes, Rebecca. 2015. Chicken run, review: 'Exuberant'. *The Telegraph*. (https://www.telegraph.co.uk/culture/film/filmreviews/11307478/chicken-run-review.html) (Accessed 2020-08-03)
Hogenboom, Melissa. 2012. How the GI influx shaped Britain's view of Americans. *The BBC*. (https://www.bbc.com/news/magazine-20160819) (Accessed 2020-08-21).

Hussain, Alaa E., Khuddro, Ahmad. 2016a. Audiovisual English-Arabic translation: De Beaugrande's Perspective. *IJALEL* 5(3). doi:10.7575/aiac.ijalel.v.5n.3p.212.
Hussain, Alaa E., Khuddro, Ahmad. 2016b. Practical approaches to English-Arabic audiovisual translation. *Global Journal of Arts, Humanities and Social Sciences* 4(2). 15-32.
Khuddro, Ahmad. 2018. *Linguistic issues and quality assessment of English-Arabic audiovisual translation*. Newcastle upon Type: Cambridge Scholars Publishing.
Kovačič, Irena. 1994. Relevance as a factor in subtitling. In Dollerup. Cay, Lindergaard, Annette. *Teaching translation and interpreting 2: Insights, aims and visions*, 245-251. Amsterdam and Philadelphia: John Benjamins Publishing Company.
Kozloff, Sarah. 2000. *Overhearing film dialogue*, Berkeley: University of California Press.
Leppihalme, Ritva. 1997. *Culture bumps: an empirical approach to the translation of allusions*, Clevedon and Philadelphia: Multilingual Matters LTD.
Maluf, Ramez. 2004. Dubbing into Arabic: A Trojan horse at the gates? (http://inhouse.lau.edu.lb/bima/papers/Dubbing.pdf) (Accessed 2020-07-20).
Matielo, Rafael, de Vasconcellos, Maria L., Baldissera, Elaine E. 2015. Subtitling words or omitting worlds? A metafunctionally-oriented analysis. *Revista de estudos da linguagem* 23(2). 363-388.
"Netflix gets 16 million new sign-ups thanks to lockdown"**.** 22 April 2020. *The BBC*. Retrieved July 23, (https://www.bbc.com/news/business-52376022) (Accessed 2020-07-23).
Nikolić, Davor, Bakarić, Davor N. 2016. What makes our tongues twist?: Computational analysis of Croatian tongue-*t*wisters. *Journal of American Folklore* 129(511). 41–52.
Nord, Christiane. 2016. Meaning, sense, function – What is transferred? *Procedia – social and behavioral sciences* 231. 3-10.
Partington, Alan. 2003. *The linguistics of political argument. The spin-doctor and the worlf-pack at the White House*. London; New York: Routledge.
Pérez-Gonzales, Luis P. 2014. *Audiovisual translation: theories, methods and issues*. New York: Routledge.
Reynolds, David. 2000. *Rich relations: The American occupation of Britain, 1942-45*. London: Phoenix Press.
Schwartz, Barbara. 2002. Translation in a confined space. *Translation Journal* 6(4). (https://translationjournal.net/journal/22subtitles.htm) (Accessed 2020-08-21).
Smith, Kevin. 2002. Translation as secondary communication. The relevance theory perspective of Ernst-August Gutt. *Acta Theologica Supplementum* 2. 107-117.
Snell-Hornby, Mary. 1995. *Translation studies: an integrated approach*. Amsterdam and Philadelphia: John Benjamins.
Sperber, Dan, Wilson, Deirdre. 1986. *Relevance: communication and cognition*, Oxford: Basil Blackwell Ltd.
Taylor, Christopher. 2000. The subtitling of film: reaching another community. In Ventola, Eija (ed.), *Discourse and community: Doing functional linguistics* [Conference proceedings], 309-330. Gunter Narr Verlag.
Törnqvist, Egil. 1995. Fixed pictures, changing words. Subtitling and dubbing the film *Babettes Gæstebud*. *TijdSchrift voor Skandinavistick* 16(1). 47-64.
Venuti, Lawrence. 1995. *The translator's invisibility: A history of translation*. London, New York: Routledge.
Venuti, Lawrence. 1998. *The Scandals of translation: Towards an ethics of difference*. London: Routledge.

Sanaa Benmessaoud
University of Sharjah
P.O. Box 27272, UAE
sbenmessaoud@sharjah.ac.ae

In SKASE Journal of Translation and Interpretation [online]. 2022, vol. 15, no. 2 [cit. 2022 12-12]. Available online at http://www.skase.sk/Volumes/JTI23/pdf_doc/01.pdf. ISSN 1336-7811.

Power Relations or Language Interference: Extraposed Linking Constructions in the Polish Translation of *The Economist*

Marta Bołtuć

Abstract

The paper examines how global English influences translations from English into Polish with respect to the so called 'extraposed linking constructions'. The analysis concentrates on different forms of 'extraposed linking constructions', their functions, distribution and their translation equivalents used in the journalistic discourse of The Economist. The use of linking constructions seems to differ in some way in English and in Polish journalistic discourse as Polish word order is freer than the English one. These differences may block English influence on Polish discourse norms via translation. The question that needs to be answered is whether this is the case in this particular discourse type.

Keywords: *globalization, language interference, linking constructions, extraposed absolute linking constructions, extraposed prepositional phrases, syntactic integration.*

Globalization and language interference

Globalization processes have had an impact on the world economy and have influenced the role of translation in the modern world. In the process of globalization, worldwide social relations link distant places and, as a result, what happens locally is influenced by distant events. Globalization is responsible for the rise of translations in many different languages and cultures.

Until recently translators and text producers have shown a tendency to apply a cultural filter. It is the aim of the present paper to investigate the impact of English extraposed linking constructions in the selected articles of *The Economist* on the use of comparable or equivalent Polish linking constructions in the translation process of the articles in question. In other words, do English textual norms in terms of linking constructions 'shine through' the Polish translation of mass media or journalistic discourse? Hence, is there a tendency of cultural and linguistic levelling? Do the usage norms of 'weaker' languages converge with those of the dominant English ones, which may result in universalism in translation? (Under universalism Assmann (2010: 121) understands "the rise of theories, ideas or beliefs with a claim to universal validity".) Or, to the contrary, can one talk about localization in the form of domestication, which means that cultural filter is applied in the translation process?

According to Esselink (2003: 67), localization is about customizing things for a 'local' audience, a process that involves "taking a product and making it linguistically, technically and culturally appropriate to the target locale where it will be used and sold". It can be defined linguistically as adjusting a product to suit the target users with respect to their language expectations. In result, what is culturally and ideologically unacceptable might have to be eliminated in the localization process.

In cultural adaptation studies one can distinguish between domestication versus foreignization (Venuti 1995), depending on whether the target or the source culture elements are more dominant in the translated text. While foreignization ensures adequacy of

translation, domestication determines its accessibility and/or comprehensibility to the local audience. It is noteworthy that according to Boltuc (2019), globalization and localization may be viewed as super-categories, where localization embodies or incorporates the socio-cultural phenomena of foreignization and domestication. Accordingly, four concepts (foreignization versus domestication and globalization versus localization) can be distinguished (Bołtuć 2019).

If a translated text has 'domestic' or 'foreign' elements, they will be viewed as localization processes. If it has more elements of the global culture, these phenomena will be described as globalization processes. Both pairs of these opposing concepts (foreignization versus domestication and globalization versus localization) are not simply the opposites, but they form a continuum (Boltuc 2019). Theoretically, any text could be placed somewhere inside the triangle below, not necessarily on its perimeter. The more the target text is altered (in comparison to the original or the start text), the more domesticated or localized it is. *Fig. 1* shows a tentative scheme for analyzing domestication versus foreignization, and localization versus globalization.

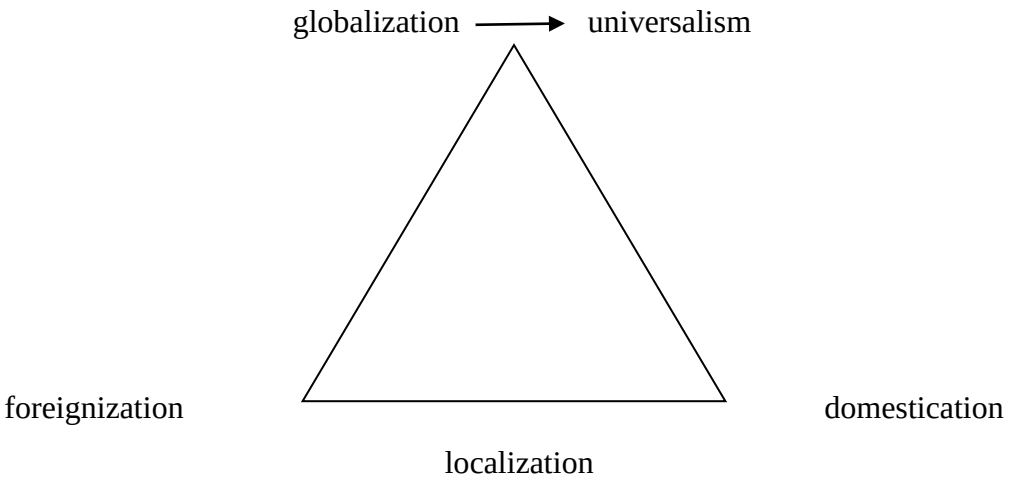

Figure 1 A tentative scheme for analyzing foreignization/domestication and globalization/localization.

It is possible for mass media or global texts to be classified somewhere in the area between or inside the three extremes: 1. foreignized and localized, 2. domesticated and localized, and 3. globalized or rather universal. It should also be mentioned that universalism seems to be an intellectual and spiritual phenomenon. Globalization, in turn, is a political, economic and civilizational process that is connected with material culture (Assmann 2010).

While it is evident that the translation of mass media texts is the effect of globalization, the relation between globalization and language interference needs to be clarified. Language interference (also known as linguistic interference or cross meaning) refers to the situation in which speakers or writers apply knowledge from their native language to a second language. Dulay et al (1982) define interference as the automatic transfer, due to habit, of the surface structure of the first language onto the surface structure of the target language. This kind of interference can also be treated as errors in the learner's use of the foreign language that are generated under the influence of the mother tongue. Ellis

(1997: 51) in turn defines interference as transfer, which he says is "the influence that the learner's L1 exerts over the acquisition of the L2."

It is noteworthy that we can talk about language inference not only in connection with the process of acquiring a foreign language, but also in connection with the process of translating from the source language into the target language and it does not seem to matter whether the translator is translating form or into his or her mother tongue. The interference of the English language, however, would more likely be referred to as a borrowing process. In most cases, either the foreign or the domestic elements in terms of word choices or even some grammatical constructions are more predominant in the translated texts.

Register and linking constructions

Register can be defined as a variety of language made use of for a particular purpose or in a particular situation, for example, degree of formality. Change of register is largely perceived as culture-dependent. It can also be claimed that there is always some ideology behind style (defined as a specific manner of writing, characteristic of a given writer, historical period or genre, less or more literary, for example). Thus, the notion of style, contrary to register, seems to be more ideology-dependent.

Van Dijk (1988: 73), however, defines style as "the total set of characteristic, variable structural features of discourse that are an indication of the personal and social context of the speaker, given a semantic, pragmatic, or situational invariant*".* In this definition style encompasses register. Van Dijk (1988: 27) is right claiming that style is an indicator of context in a text as it may unveil some personal or social factors of the communicative context. As a result, it seems obvious that style and register are somehow interrelated.

One can also find the trace of style and/or register used by an author in the linking words or constructions used. Linking words and constructions are single, or multiple-word, lexico-grammatical patterns used to indicate a relationship between some part of a prior and/or following discourse. We can distinguish 'unembedded/extraposed linking constructions', unattached to the syntax of the clause they introduce (Haegeman 2009). They appear on the left periphery of a sentence (*After all, In addition,* for example). Linking constructions negotiate information between a writer and a reader, identify, reinforce or foreground theme, exemplify, introduce, re-introduce, add and contrast information or a referent and indicate temporal sequencing. According to Prince (1985), they facilitate the processing of discourse-new entities.

Bührig and House (2007) identified two types of 'extraposed linking constructions': 1. extraposed absolute linking constructions (*Given this result, Simply put, Viewed differently*) and 2. extraposed prepositional phrases (*After all, In particular, On the other hand, In fact, In short, In contrast, In addition*). These linking constructions can behave differently in translation; they can be preserved, substituted, omitted or they can undergo syntactic integration with some other part of a sentence. According to House (2017: 52), extraposed absolute linking constructions are infrequent in the popular-science corpus.

Data and methodology

I will investigate the behavior of English extraposed linking words or constructions, occupying sentence initial position, in the translation into Polish in global or mass media discourse. The data analyzed consists of 11 articles published in *The Economist* (special

edition – *The World in 2011*) and their Polish translations. The limited scope of the material is justified by the fact that *The Economist* is normally not translated into Polish, only its special editions can sometimes be translated. All full-length articles published in the special edition of *The Economist – The World in 2011* have been chosen for the analysis (about 12 000 words altogether).

There are a few hypotheses underlying the analysis of occurrence, variation, omission and change of linking constructions discussed in the previous section:

(1) Global English as dominant lingua franca influences communicative preferences and discourse norms in Polish through language contact in translation.
(2) The cultural filter applied in cultural adaptation – domestication – is no longer applied because of English influence on Polish translated texts – foreignization.
(3) Anglophone influence is particularly visible in mass media, journalistic, economic and scientific discourse.

Following House (2017), who draws the same conclusions about German popular-science discourse, I assume that Anglophone and Polish mass media discourse preferences can vary along the following dimensions: directness versus indirectness, orientation towards content versus orientation towards persons and explicitness versus implicitness. As a result, there can be a more detached communicative style in Polish and German, and a more interactional and involved style in English mass media, journalistic or popular science discourse.

In the 18th and 19th centuries, in the continental tradition (including Poland) popular-science journals were lighter scientific journals addressed to educated audience such as engineers or high school teachers, whereas in the mid or late 19th century in the Anglo-American tradition popular-science magazines addressing a broader audience emerged. For example, *Scientific American, National Geographic* and *The Economist* were all originally published in the 19th century. Consequently, it seems that a journalistic or popular science genre as such has a longer tradition in the English-speaking countries than in Poland. The popular science genre is less established in Polish culture (translations of these magazines or their Polish versions appeared towards the end of the 20th century).

As a result, Polish and German popular science texts tend to be less 'popular' and closer to scientific texts (Bołtuć 2016, House 2017). Similar conclusions can be drawn in relation to other most widely used European languages (French and Spanish, for example) (Kranich and González Diaz 2010, Küppers 2008, Probst 2001). For instance, House (2017: 51) writes that "the German popular science genre is generally less interpersonally oriented, less addressee- and more content-oriented as well as more written than spoken". According to Bołtuć (2016), the same can be said about Polish journalistic and especially popular-science discourse.

The analysis undertaken in the project consisted of the following steps:

(1) Extraction of all occurrences of English 'extraposed linking constructions' and their translational occurrence or variation in translated texts.
(2) Frequency counts in the comparative corpora.
(3) Are equivalent items used for the same communicative purpose in different corpora?
(4) Interpretation of findings.

The methodology employed in this project is a combination of qualitative and quantitative methods. It consisted in categorizing corresponding linking devices and frequency count in the two languages to see if they were susceptible to variation and/or

change under the influence of the English norms. Manual annotation was used to find different co-occurrences of various kinds.

Analysis of extraposed linking constructions

What struck me most in the material analyzed is the frequency of the use of contrasting conjunction *but* in this kind of journalistic discourse. Originally it appeared 29 times in sentence initial position in English. When it appeared in the mid position of a sentence it was usually translated in a literal way as *ale* or *jednak*. In sentence initial position it was usually translated as *Ale* – there were 9 such instances. At this point it needs to be stressed that it is uncommon or even ungrammatical in Polish to begin sentences with '*ale*' and it can certainly be considered a colloquial style. This is where the influence of English is very visible as translators exhibited the tendency to translate this conjunction in a literal way and substituted it with the closest equivalent '*ale*', even in sentence initial position. This conjunction, however, is equally frequently substituted in Polish by other contrasting conjunctions such as *Jednak, Niemniej* or *Natomiast* (literally, 'yet', 'but', 'however',) – 11 such instances have been found. It is more common in Polish to begin sentences with *Jednak, Niemniej* (which is a short form of *niemniej jednak* ('nevertheless')) than with *Ale* which sounds more colloquial. There are also 4 instances where the conjunction *but* was omitted in translation when it was in sentence initial position in the original. And there are 5 instances where English initial *But* was substituted by Polish *jednak, niemniej jednak, natomiast* (synonymous conjunctions to English *but*) and was embedded in mid-sentence position. Consider the examples below:

(1) *But* the savagery could also be the start of something much more beneficial.
Ale zdecydowane cięcia mogłyby też doprowadzić do czegoś o wiele bardziej pozytywnego.
'*But* decisive cuts could also lead to something much more positive'.
(there were 9 instances with Polish initial *ale*)

(2) *But* that something should not be austerity at any price.
Jednak tym czymś nie może być oszczędzanie za wszelką cenę.
'*However,* that something cannot be saving at any cost'.
(there were 11 instances with Polish initial *jednak, niemniej*)

(3) *But* no numerological thrill-seeker need feel short-changed in 2011.
W 2011r. żaden numerologiczny poszukiwacz nie powinien się czuć zawiedziony.
'In 2011 no numerological seeker should feel disappointed'.
(there were 4 instances of omission of initial *but*)

(4) *But* a more balanced German economy will not be enough to rescue some euro-zone countries from a difficult year.
Bardziej stabilna gospodarka Niemiec nie wystarczy *jednak*, by uchronić niektóre kraje strefy euro przed trudnościami w nadchodzącym roku.
'A more stable German economy is not enough, *however*, to save some euro zone countries from the difficulties in the coming year'.
(there were 5 instances where English initial *But* was incorporated/embeded in mid-sentence position in Polish translation).

It is a different story with English initial conjunction *And* ('*I*' in Polish). It appeared in sentence initial position only 5 times in English. It was substituted by Polish *A* (literally,

'but', which usually introduces some contrasting information in Polish) twice and once by *Ponadto* (literally, 'moreover'). There were also two instances where English initial *And* was omitted in Polish translation.
Consider the following sentences:

(5) *And* the blame for most of this lies with California's voters.
A największą winę za ten stan rzeczy ponoszą kalifornijscy wyborcy.
'*And* the Californian voters bear the greatest guilt for this state of affairs'.
(there were 2 instances with Polish initial *A* and one with *Ponadto*)
(6) *And* that's a new year's resolution that can start painlessly – on 1:1:11.
Takie nowoczesne postanowienie byłoby pierwszego dnia bezbolesne: 01-01-11.
'This new year's resolution would be painless on the first day – 01-01-11'.
(there were 2 instances where English initial *And* was omitted in Polish translation)

English *So* in turn appeared only 3 times; it was translated as *Dlatego też* (literally, 'also, for this reason') once. It was translated as *więc* (literally *'*so', and was embedded in the Polish sentence structure in mid-position) once and was also omitted one time in Polish translation. Consider the following instances:

(7) *So* the private sector will need to take the strain, *and* the signs here are not good.
Dlatego też powstanie potrzeba, by ciężar ten został przejęty przez sektor prywatny, *a* tu znaki nie wróżą dobrze.' (one instance)
'*Also, for this reason*, a need will arise for the private sector to take over the strain, and here the signs are not good'.
(8) *So* the fashion may not last.
Moda może *więc* nie potrwać długo.' (one instance)
'*So* the fashion may not last long'.
(9) *So* Mr Weber's characteristic bluntness may count against him.
Obcesowość Webera może działać na jego niekorzyść. (one instance of omission)
'Mr Weber's bluntness may work to his disadvantage'.

It is noteworthy that it is quite uncommon or not very grammatical in Polish to begin sentences with *Dlatego też*, so the sentence number 1 above, being the evidence of the English influence, is not very grammatical. In a similar vein another Polish sentence that start with *Ponieważ* (literally, 'because or since') is also rather ungrammatical:

(10) *Since* human beings have ten fingers…
Ponieważ istoty ludzkie mają dziesięć palców…
'*Since* human beings have ten fingers'

There are also two instances with *Yet*, at the beginning of English sentences. This English conjunction is translated by means of *Niemniej* once, which is more or less a literal translation and it is omitted once in translation. Other extraposed absolute linking constructions or linking words used in sentence initial position by English journalists and their Polish translations are as follows:

(11) *Indeed – W rzeczy samej* (literal translation by means of an extraposed prepositional linking construction in Polish)
(12) *First – Po pierwsze* (literal translation by means of an extraposed prepositional linking construction in Polish)
(13) *Second – Po drugie* (literal translation by means of an extraposed prepositional linking construction in Polish)
(14) *Strangely – Dziwne to, ale* (literal translation with a deictic *to,* 'this', in Polish translation and conjunction *ale*, 'but', which makes the initial conjunction incorporated in the sentence that follows)
(15) *Even so – Mimo to* (literal translation with a deictic *to*, 'this', in Polish translation)
(16) *Unfortunately – Niestety* (literal translation in Polish)
(17) *No doubt – Nie ma wątpliwości* (literal translation in Polish)
(18) *Even – Jednak* (literal translation in Polish)
(19) *Hence – Stąd* (literal translation in Polish)
(20) *Most fundamentally – Co najważniejsze* (literal translation in Polish)
(21) *Despite all this – Mimo to* (literal translation, but here English *all* is omitted in translation)
(22) *One way or another – Tak więc* (synonymous expression and *Tak czy inaczej* would be a more literal translation here)

The examples above illustrate the tendency to translate English extraposed absolute linking words or constructions in a literal way into Polish, preserving their sentence initial position.

There are also some extraposed prepositional linking constructions in the English texts analyzed. They are usually translated in a literal or synonymous way; there are however some rare instances when they are omitted or embedded in the body of the sentence in Polish translation. Consider the following examples:

(23) *In general*, making government smaller is a good idea…
Zmniejszenie rządu to dobry pomysł… (linking construction *In general* was omitted in Polish translation)
'Making the government smaller is a good idea…'
(24) *After all – Koniec końców* (synonymous translation in Polish, 'in the end')
(25) *Far from – Co gorsza* (synonymous translation, 'what is worse')
(26) *In this respect – Pod tym względem* (literal translation)
(27) *In short – Krótko mówiąc* (literal translation)
(28) *For a start – Na początek* (literal translation)
(29) *In its aftermath*, the differences between the developed and the emerging economies look stark, and in 2011 they will be strikingly clear.
Kryzys wyraźnie ujawnił różnice między gospodarkami rozwiniętymi i wschodzącymi; w 2011 różnice te będą się rzucały w oczy. (linking construction *In its aftermath* was omitted here in translation)
'The crisis clearly revealed the differences between the developed and the developing economies; in 2011 these differences will be conspicuous'.
(30) *In essence*, the multinational company of the past was a series of national businesses co-ordinated by a single global headquarters.

W przeszłości firma wielonarodowa była *w gruncie rzeczy* zbiorem firm narodowych koordynowanych przez globalną centralę. (here the linking construction was translated literally and embedded or incorporated into the translated sentence, taking a mid-sentence position)
'In the past the multinational company was, *in essence,* a collection of national firms coordinated by a global headquarters'.

(31) *By contrast*, the very wealthy in fast-growing emerging markets such as China and India will feel less heat.
Na szybko rosnących rynkach wschodzących, jak Chiny czy Indie, *będzie inaczej* – bardzo bogaci odczują mniejszy ogień krytyki. (here the linking construction is incorporated in the sentence in question, meaning literally 'will be different')
'It will be different in fast-growing emerging markets such as China and India, the very wealthy will feel less heat'.

(32) *Among other things – Między innymi* (literal translation)
(33) *Compared with – W porównaniu* (literal translation)
(34) *By contrast – W przeciwieństwie* (literal translation)

It can be said that extraposed prepositional linking constructions are in most cases translated in a literal or synonymous way. They can also sometimes be omitted in translation, but if translated, their initial position is usually preserved in translation into Polish.

Interpretation of findings and conclusions

It can be said that English extraposed absolute linking constructions or words as well as English extraposed prepositional phrases are more often than not translated in Polish articles of *The Economist.* In most cases, they seem to be translated literally which may result in their unnecessary and often unjustified influence on the Polish texts in question, resulting in deliberate or nondeliberate interference of the English language. As a result, Polish sentences begin with words such as *Ale* ('but'), *Dlatego też* ('also for this reason') or *Ponieważ* ('because'), which is quite uncommon or not very grammatical in Polish and sounds colloquial.

On the other hand, there can be instances where extraposed linking constructions were omitted in translation or incorporated in the middle of the sentences in Polish. In a few examples, deictic expressions were used together with linking constructions in Polish – this need may be generated by the structure of the Polish language, whose word order is less strict than it is in English. In this respect the number of Polish extraposed linking constructions in translation is surprisingly high and can possibly be treated as a trace of the English language influence on Polish mass media discourse norms. Generally, extraposed linking constructions are more efficient and reader-oriented as they seem to be beneficial for readers' processing efforts, requiring less processing time than embedded linking constructions, which are frequently used in the Polish language. In other words, one of the main reasons why Anglo-American expressions (including terminology, but also linking constructions which are often translated in a literal way) are, in some sense, absorbed into other national languages, including Polish, is that they can be more efficient and easier to understand for a broader audience.

The function of extraposed linking constructions is to set the theme/scene of the clause that follows "as it were the peg on which the message is hung" (Halliday 1970:161)

and the theme setting acts like a frame (Goffman 1972, Gumperz 1982). This kind of framing in discourse has a strong cohesive effect, but it can also enhance readers' comprehension and clarity of the message.

It seems that while Polish texts sound a bit less interactional or are less involved, more content oriented than the English ones, there seems to be some influence of the English language on the Polish texts in translation. Translations from English seem to be more colloquial in register or style in comparison to Polish mass media parallel texts. This tendency is the result of foreignization or even globalization processes or universalism, the hegemonic position of English language and not quite equal power relations between languages, which is especially visible in mass media language. As has been mentioned, cultural filtering can often be suppressed in translation, which results in linguistic interference or English textual norms to 'shine through' (Bołtuć 2016:187, House 2009:82). Consequently, all three hypotheses mentioned in the paper above in section 3. seem to have been confirmed by this preliminary analysis.

Finally, it might be said that the distribution/position and meaning or function of the extraposed linking devices, both the absolute ones and the prepositional ones, is quite similar in the original and translated Polish texts of *The Economist*. It might also be interesting to investigate the impact of English textual norms on Polish mass media or journalistic parallel texts/parallel mass media discourse.

References:

ASSAMANN, Jan. 2010. Globalization, universalism and the erosion of cultural memory. In ASSMANN A. and CONRAD S. *Memory in a Global Age*. New York: Palgrave Macmillan, 2010, pp. 121-137.

BOŁTUĆ, Marta Izabela. 2016. *Lost in translation: linguistic creativity in popular-science text as illustrated by National Geographic heads*. Rzeszów: Rzeszów University Publishing House, 2016.

BOŁTUĆ, Marta. 2019. Globalization, localization and national identity in translation. Peter Lang [forthcoming], 2019.

BÜHRING, Kristin, HOUSE, Juliane. 2007. 'So, given this common theme': linking constructions in discourse across languages. In REHBEIN J., HOHENSTEIN Ch., and PIETSCH L. *Connectivity in grammar and discourse*. Amsterdam: John Benjamins, 2007, pp. 345-366.

DULAY, Heidi, BURT, Marina, KRASHEN Stephen. 1982. *Language two.* New York: Oxford University Press, 1982.

ELLIS, Rod. 1997. *Second language acquisition.* Oxford: Oxford University Press, 1997.

ESSELINK, Bert. 2003. The Evolution of localization. available at: www.alchemysoftware.ie/index.html. 01/01/2019.

GOFFMANN, Erving. 1974. *Frame analysis*. New York: Harper & Row, 1974.

GUMPERZ, John. 1982. *Discourse strategies*. Cambridge: Cambridge University Press, 1982.

HAEGEMAN, Liliane. 2009. Parenthetical adverbials: the radical orphanage approach. In SHAER, B., COOK, P., FREY, W., MAIENBORN, C. *Dislocated elements in discourse*. London: Routledge, 2009, pp. 348-365.

HALLIDAY, Michael A.K. 1970. Language structure and language function. In LYONS, J. *New horizons in linguistics*. Harmondsworth: Penguin, 1970, pp. 140-160.

HOUSE, Juliane. 2009. *Translation*. Oxford: Oxford University Press, 2009.

HOUSE, Juliane. 2017. Quality in translation studies. In MILLÁN, C. and BARTRINA, F. *The Routledge handbook of translation studies*. London and New York: Routledge, 2017, pp. 534-560.
KRANICH, Scenja, GONZALEZ DIAZ, Victoria. 2010. Good, great or remarkable: evaluation in English, German and Spanish letter to shareholders. Paper given at the conference on *New challenges for multilingualism in Europe*. Dubrovnik, 10-15 April 2010.
KÜPPERS, Anne. 2008. *Sprecherdeiktika in deutschen und französischen Aktionärsbriefen.* Arbeiten zur Mehrsprachigkeit 87. Hamburg: Universität Hamburg, 2008.
PRINCE, Ellen. 1985. Fancy syntax and 'shared knowledge. In *Journal of pragmatics*, 1985, vol. 9, no. 1, pp. 65-81.
PROBST, Julia. 2001. Der kulturelle Filter in Wirtschaftstexten multilingualer Unternehmen: Englisch-Deutsch-Französisch. In CIGADA S., GILARDONI S. and MATHEY M. *Comunicare in ambiente professionale plurilingue. Kommunikation in der mehrsprachigen Arbeitswelt. Communiquer en milieu professionel plurilingue. Communicating in professional multilingual environments*. Lugano: Proceedings of the VALS/ASLA Congress, 14-16 September 2000, pp. 263-275.
VAN DIJK, Teun. 1988. *News as discourse*. Hillsdale, N.J.: L. Erlbaum, 1988.
VENUTI, Lawrence. 1995. *Translator's invisibility: a history of translation.* London: Routledge, 1995.

Source material:

The Economist: The World in 2011:
Tremble Leviathan; A year of palindromes; Much less stimulating; Avoidable errors, Multinationimble; Cooling the Earth; Another year, another billion; The filthy rich; Curl up with a good screen; Powerhouse Deutschland; The emerald no longer shines.
The Economist: Świat w 2011:
Drżyj Lewiatanie; Rok palindromów; Dużo mniej stymulacji; Błędy do uniknięcia; Rącze i żwawe korporacje globalne; Schładzanie ziemi; Kolejny rok, kolejny miliard; Obrzydliwie bogaci; Ekran do poduszki; Potęga Niemiec; Szmaragd już nie błyszczy.

Marta Bołtuć
the University of Rzeszów
Department of Translation Studies
al. Rejtana 16c
Rzeszów 35-310
e-mail: mibotuc@gmail.com

In SKASE Journal of Translation and Interpretation [online]. 2022, vol. 15, no. 2 [cit. 2022 12-12]. Available online at http://www.skase.sk/Volumes/JTI23/pdf_doc/02.pdf. ISSN 1336-7811.

Multiword Expressions as Discourse Markers in Hebrew and Lithuanian

Giedrė Valūnaitė-Oleškevičienė, Chaya Liebeskind

Abstract

Multiword expressions are of key importance in language generation and processing and could also operate as discourse markers. We combined the alignment model of the phrase-based statistical machine translation and manual treatment of the data in order to examine English multiword discourse markers and their equivalents in Lithuanian and Hebrew, by researching their changes in translation. We focused on the two most frequent: 'I think' and 'you know' aiming to research if they demonstrate their functional stability as discourse markers in translation and what changes they undergo in Lithuanian and Hebrew translation.

Keywords: *multilingual corpus; multiword expression; discourse relation; discourse marker; translation.*

Research on multiword expressions has identified that language is not produced just word by word but it usually involves generating certain chunks using a lot of formulaic constructions (Barlow 2011). Native speakers have a multitude of memorized sequences to perform various functions within language, for example, organizing discourse (Nattinger and DeCarrico 1992), or processing language by the speaker and the hearer (Siyanova-Chanturia, Conklin, and Van Heuven 2011). Formulaic language includes idioms and proverbs, various clichés and collocations, lexical bundles, and phrasal verbs. Biber et al. (2004) observed that lexical bundles constitute a high percentage of the produced language and the authors identified that one function of lexical bundles is to organize discourse by providing an example of such bundles, for example, *I think*, which relates to the research on discourse markers. Phrases such as *you know* and *I think* have also been classified as discourse markers that perform certain discourse organising functions. However, Maschler & Schiffrin (2015) observe that there is no a priori theoretical classification of discourse markers and the analysis of function in the data is necessary. Research on discourse markers as tools of discourse management prove that they carry several functions, including signposting, signalling, and rephrasing. Furthermore, there are ongoing attempts to investigate the importance of discourse layers in language production, communication, second language learning, and translation. Additionally, Dobrovoljc (2017) has recently attempted to research multiword expressions as discourse markers in a corpus of spoken Slovene, identifying structurally fixed discourse marking multiword expressions.

The purpose of the current research is to examine multiword expressions used as discourse markers in TED talk English transcripts focusing on 'I think' and 'you know' and compare them with their counterparts in Lithuanian and Hebrew by following Maschler & Schiffrin (2015) observation on the necessity of closer investigation on their function as discourse markers. To achieve the aim of the research, the set objectives were to create a parallel research corpus to identify multiword expressions used as discourse markers and to analyse their translations in Lithuanian and Hebrew to determine if they function as discourse markers and are also multiword expressions or one word translations, or if they acquire any other linguistic forms. An additional benefit of the study was extending the available resources and providing linguistic processing for several languages by creating a multilingual

parallel corpus (including English, Lithuanian, and Hebrew) based on social media texts; the created corpus is shared and interlinked via CLARIN open language resources.

Theoretical background

The literature overview briefly takes into account the research languages, studies related to multiword expressions and their use as discourse markers, the importance of discourse markers for discourse management, and certain insights into discourse marker translation.

Cultural heritage and languages of the research

First, it is necessary to briefly discuss the cultural heritage of the languages of the research, which, in a way, guided the choice of languages for our study. According to Bieliauskienė (2012), Jewish and Lithuanian cultures coexisted on the same territory from the first half of the 14th century. The author stressed that from 19th century onwards, in the Republic of Lithuania, Vilnius was called Lithuania's Jerusalem, attracting knowledgeable people in the field of education and inspiring a flourishing high culture, for example, in theatre, art, and literature. In fact, both languages, Lithuanian and Hebrew, formed the cultural heritage of the region. In this study, we research the Lithuanian and Hebrew corpus in parallel with pivotal English.

Lithuanian is an old surviving Baltic language, retaining forms related to Sanskrit and Latin and preserving the most phonological and morphological aspects of the Proto-Indo-European language. Thus, it has gained importance in Indo-European language studies and has been researched by many scientists so far, including Ferdinand de Saussure, who considered Lithuanian "the Galapagos of linguistic evolution" (Joseph 2009). Lithuanian is rich in declensions and cases inside the declensions and the oldest layer of the Lithuanian language vocabulary is related to the Indo-European language, which is dated to be approximately over 5000 years old.

Hebrew is a very old, northwest Semitic language belonging to the group of Canaanite languages; the first examples of Paleo-Hebrew date back to the 10th century. It is a successful example of a revived dead language. It survived in the medieval period as the language of religious scriptures, being revived, in the 19th century, into a spoken and literary language (Joslyn-Siemiatkoski 2007). Hebrew is an important language for researchers specializing in Middle East civilizations and Christian theology studies.

Multiword expressions as discourse markers

The research areas of natural language processing (NLP), linguistics, and translation are closely related to discourse research, focusing on discourse relations between clauses or sentences. NLP research focuses more and in depth on multiple language-related areas, such as semantic phenomena, dialogue exchange structure, and discourse textual structure (Webber and Joshi 2012). NLP recognizes that language is not just placing words in the right order but getting the meaning and deeper textual relations as well as organizing ideas into a logical textual flow. According to researchers (Barlow 2011; Sinclair 1991), language is not just generated word by word; it is also formulaic. Speakers possess multiple learnt formulaic

sequences, which, according to Siyanova-Chanturia et al. (2011), are important in organizing discourse and help the language producer and recipient to manage language processing. However, formulaic language is not easy to manage and categorize for NLP research, as it may seem at first sight, since the sequences that could be considered formulaic vary in length, meaning, fixedness, etc., and the finalized definition of formulaic language has not fully crystallized. It could be considered as an umbrella term embracing idioms, proverbs, clichés, phrasal verbs, collocations, and lexical bundles (Wray 2012). According to Wei & Li (2013), formulaic language covers approximately 60% of written texts in their researched corpus of English academic language. According to Biber et al. (1994; 1999), lexical bundles are groups of words that show a statistical tendency to co-occur and could be considered as extended collocations, for example, *I think*. Biber et al. (2004) identify that lexical bundles have functional purposes, such as organizing discourse, expressing stance, and referential meaning. Based on the evidence of the formulaic nature of language for communication, research has turned to investigating multiword expressions used as discourse markers (Dobrovoljc 2017), identifying structurally fixed discourse marking multiword expressions.

Another important issue in NLP is discourse management, which is related to discourse relations, connecting ideas between sentences and bigger parts of the text. Discourse relations may remain implicit or be expressed explicitly through discourse markers, which help textual coherence and discourse management, and are used for making coherent speech appropriately segmented to enable textual understanding. Discourse markers perform important functions, such as signposting, signalling, and rephrasing, by facilitating discourse organization. They are mainly drawn from syntactic classes of conjunctions, adverbials, and prepositional phrases (Fraser 2009), as well as expressions such as *you know*, *you see*, and *I mean* (Schiffrin 2001; Hasselgren 2002; Maschler and Schiffrin 2015). Hasselgren (2002) advocated that better discourse marker signal fluency contributes to interaction and even makes the speaker sound more 'native-like'. Recently, discourse relations and discourse maker research has gained certain impetus with corpora annotation for exploring discourse structure in texts, for example, the Penn Discourse Tree Bank (PDTB); (Webber et al. 2016). Furthermore, there was a rise in annotated multilingual corpora for researching different means of expressing discourse relations and managing discourse (Stede et al. 2016; Zufferey and Degand 2017; Oleskeviciene et al. 2018; Zeyrek et al. 2019).

Language, especially spoken, is characterised by discourse marker use; however, some of them (e.g., *you know*, *I think*, *well*) are sometimes referred to in a critical manner, as indicating a lack of fluency (O'Donnell and Todd 2013). Still, discourse markers are abundantly used and, according to Crystal (1988), they enhance communication if used appropriately and should not be considered unnecessary or undesirable. As Biber (2006) observed, discourse markers, such as *you know*, or *well*, are very rare in written language. However, they are quite common in spoken discourse and should not be treated as just fancy words since they serve the function of organizing discourse by signalling, rephrasing, marking, or relating ideas. Svartvik (1980) observed that, if a foreign language learner makes a mistake (e.g., *he goed*), it can be easily identified and redeemed by the native speaker; however, if a learner misses words such as *you know*, or *well*, the native speaker cannot identify any error and the speech might sound impolite or even dogmatic. The same idea is also supported by Hasselgren (2002), who observed that discourse markers enhance interaction. Furthermore, it has also been researched using learner corpora to demonstrate the importance of discourse level knowledge, especially at more advanced levels of language learning (Granger 2015; Cobb and Boulton 2015).

Translation issues of discourse markers

Discourse markers are used in both written texts and spoken discourse to connect ideas and guide the reader or the listener through expression by ensuring that the ideas are grasped correctly. Discourse markers have been researched by applying various theoretical approaches, such as Rhetorical Structure Theory (Mann and Thompson 1988), Segmented Discourse Representation Theory (Asher, Asher, and Lascarides 2003), and PDTB (Prasad et al. 2008), first focusing on the monolingual approach, which resulted in multilingual studies focusing on translation (Degand and Pander Maat 2003; Pit 2007; Dixon 2009; Zufferey and Cartoni 2012). As Zufferey & Cartoni (2012) observed, multilingual studies are more complicated as languages differ in the use of discourse markers and their expression. The authors also added that often discourse markers are poly-semic, which means that a single expression of a discourse marker may perform in expressing various discourse relations. They provided an example of the English *since*, which could express temporal or causal discourse relations depending on the surrounding contexts.

Recently, much research has gained interest in using parallel translated corpora. For example, Dupont & Zufferey (2017) focused on the investigation of translation corpora to study if the effect of register, translation direction, or translator's expertise could influence the shifts of meaning and omissions of English and French markers of concession. Hoek et al. (2017) investigated a parallel corpus on English parliamentary debates translated into Dutch, German, French, and Spanish, searching what types of discourse connectives might have a higher tendency to be more frequently omitted in translation. Baker (2018), in her extensive studies on translation, observed that discourse markers could be used to signal different relations and these relations could be expressed by a variety of means. The author provided the example that, in English, the expression of causality could be realized through content verbs, such as *cause* or *lead*, or more simply, through a discourse marker signalling the causality relation. Further, different languages demonstrate different tendencies – some languages prefer using simpler structures connected by a variety of discourse markers, while other languages favour complex structures, sparsely using explicit discourse markers. The author analysed the example of an evident difference between English and Arabic, identifying that, while English prefers signalling discourse relation through discourse markers, Arabic prefers grouping the information into bigger grammatical chunks and using fewer discourse markers. The finding is supported by (Al-Saif & Markert (2010), who observed that, in Arabic, many discourse relations are expressed via prepositions with nominalizations. Therefore, translation poses a challenge in adapting various preferences of the source and target languages. Translators face various choices of inserting discourse markers to make the flow of the ideas smoother in the target text, however, they risk making the translation sound foreign or transposing the grammatical syntactic structure, ending up using different means of expressing discourse markers or simply omitting them. It appears that it is not always possible to use the word for word technique and natural changes in translation are sometimes inevitable. According to Baker (2018), grammatical changes in translation involve certain techniques, such as substitution, transposition, omission, and supplementation. Substitution is the change of the grammatical category of the source unit in translation. For example, active voice is more common in Lithuanian; therefore, English passive voice units could be changed into active units:

(1) He was told the news. – jam pranešė naujienas

Similarly, in the following example, the verb in the source language is changed into a noun in Hebrew translation.

(2) We should have broken ten minutes before. – היינו צריכים לצאת להפסקה לפני 10 דקות

Transposition represents a change of position in the order of elements of the source textual unit or changing the part of speech in translation, which implies the change in the order of the elements in the translated text. In Lithuanian translation, we observe a change in the order of the elements in the sentence.

(3) After he had left – Jam išėjus.

In the case of Hebrew translation, the change of the order of the elements could be observed in the following example.

(4) Classical music – מוזיקה קלאסית

Omission occurs when some elements of the original text could be considered excessive or redundant in translation. In the Lithuanian translation example, the whole phrase *I thought* is omitted.

(5) I thought you said you were alright. – Bet tu sakei, kad viskas gerai.

In the following example in Hebrew, the translation of *are* is omitted.

(6) We still are – אנחנו עדיין

Supplementation involves changes when new elements, which are non-existent in the source text, appear in the translated text in order to ensure structural adequacy of the latter. Such modifications are usually considered structurally or contextually motivated. For example, due to the elliptical nature of the English language, the Lithuanian translation should use supplementation to make the translation understandable.

(7) Soap star – muilo operos žvaigždė (although the word opera is omitted in English due to ellipsis, it should be added in Lithuanian translation to make it contextually coherent).

The same technique should be applied in the Hebrew translation.

(8) Soap star – כוכב אופרת סבון

As shown above, translation is not a mere process of transposing words from one language into another but requires certain motivated changes. Thus, translation involves grammatical transformations, as a result of the process of looking for approximate correspondences in the translated texts.

Research data resources

It should be stressed that parallel data resources are not extensive, and researchers still need to work on creating parallel corpora for their research, especially if they would like to cover the variety of languages and areas. One of the most prized parallel multilingual resources is Europarl (Koehn 2005). It comprises the translations of the European Parliament proceedings (at most 50 million words) in most European languages; however, it covers just one specific domain of parliamentary proceedings.

TED talks subtitles to their videos seem to be a growing resource of parallel linguistic material, covering a multitude of languages. In addition, being an open and a developing

resource, TED talks attract attention of researchers and their subtitles cover a wide variety of knowledge fields (Cettolo, Girardi, and Federico 2012), which makes the data of the talks widely applicable. However, researchers should keep in mind that the talks are translated by volunteers although with administratively managed quality checks, and the translation is mostly unidirectional from source English subtitles to other target languages. Furthermore, Dupont & Zufferey (2017) identified that such talks contain features of both spoken and written language, as they are semi-prepared speeches by nature. Additionally, (Lefer & Grabar (2015) observed that subtitle translation bears certain specificity in itself. Even by taking into account the features of TED talks discussed by researchers, TED talks are extensively useful as they are an open resource and could provide large amounts of parallel data for research. Besides, parallel corpora are employed as a pool of data for statistical machine translation systems and TED talks is one of the most frequent data resources referred to explore multilingual Neural MT (NMT) (Aharoni, Johnson, and Firat 2019; Chu, Dabre, and Kurohashi 2017; Hoang et al. 2018; Khayrallah et al. 2018; Tan et al. 2018; Xiong et al. 2019; Zhang, Meng, and Liu 2019). NMT, as currently the newest technique of MT, stems from the model of the functioning of the human brain neural networks, which place information into different layers for processing it before generating the outcome. With the technological advancements, NMT gained impetus, as it used to be, resource and computation wise, too costly to outdo phrase-based MT, which operates on the basis of translating entire sequences of words. Now, the neural approach of NMT started challenging the long-lasting prevalence of phrase-based MT techniques. However, in the current research, phrase-based MT was applied relying on two main reasons: NMT techniques do not allow extensive processing of phrases and NMT procedures are not as explicit as phrase-based MT processes. The current study does not involve the full set of phrase-based MT systematic procedures, as it is used just for a phrase table construction, which is a single step of the phrase-based MT paradigm. The detailed description of the research procedures is provided in the research methodology section.

Research methodology

The research aim comprised examining multiword expressions used as discourse markers in TED talk English transcripts and comparing them with their counterparts in Lithuanian and Hebrew. Thus, there was a need to achieve the double objectives of creating the parallel corpus for the research data and carrying out the research on multiword expressions used as discourse markers in the studied languages. Unlike working on one language and using statistical methods we used parallel corpus knowledge alignment algorithm. Initially, the list of multiword and one word expressions that could potentially be used as discourse markers was generated relying on theoretical insights by Schiffrin (1987) and the classification provided by Fraser (2009). Fraser's extensive classification was taken as a basis, and Huang's (2011) theoretical analysis of discourse marker characteristics for spoken discourse, for example, *you know, you see, I mean, I think,* was also included.

Parallel Corpus creation

First, a parallel corpus meeting the research aim needed to be created. We decided to use TED Talk transcripts, as they are publicly available and provide appropriate material for

parallel data. In order to create a substantial parallel corpus containing data in English, Lithuanian, and Hebrew, the talks were extracted automatically using a special code, which ensured that English sentences with the candidate discourse markers from the theoretically based list were extracted and matched with their Lithuanian and Hebrew counterparts. The process of creating the parallel corpus could be viewed as an innovative achievement as it allows parallelizing the data of any researched languages. While building the corpus, the parallel texts in English, Lithuanian, and Hebrew were extracted from TED talk transcripts. Then, the sentences were aligned to make a parallel corpus for further research. The corpus contains 87.230 aligned sentences (published in LINDAT/CLARIN-LT repository http://hdl.handle.net/20.500.11821/34).

Multiword discourse marker extraction

Another stage of the research focuses on multiword expressions that are used as discourse markers to ensure textual cohesion and, according to Fraser (2009), to relate separate discourse messages. For example, phrases such as *you know*, *I mean*, *of course*, are characteristic of spoken language (Maschler and Schiffrin 2015; Furkó and Abuczki 2014; Huang 2011). Thus, 3.314 aligned sentences containing the earlier mentioned multiword expressions were extracted and manually annotated, spotting the cases in which the expressions were used as discourse markers. One-word discourse marker identification did not represent much challenge; however, turning to multiword expressions, they certainly caused challenges. For example, to identify if the expression *you know* is used as a connective, the context in which it occurs should be examined by identifying if the expression serves as a discourse marker. As such, two situations arise: (1) the multiword expression *you know* is used to introduce a new discourse message, or (2) they are content words fully integrated into the sentence.

(1) You know, this is really an infinite thing.
(2) You know exactly what you want to do from one moment to the other.

After that, the variations of the translations of discourse markers into Lithuanian and Hebrew were extracted automatically for a comparative study, determining the variations in translation. We ran an NLP word-alignment algorithm to extract a phrase table of all the possible translations of the researched discourse markers, using our parallel corpus (in our case, source = English, target = Lithuanian/Hebrew). The extraction of the translation variations was dependent on the phrase-based statistical machine translation model introduced by Koehn et al. (2003). The model could be visually represented in the research languages by the figures below.

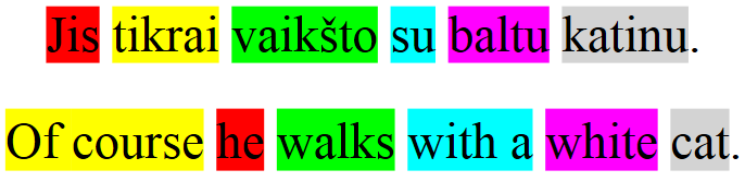

Figure 1. Lithuanian – English phrase alignment

Figure 1 visualizes Lithuanian – English corresponding phrases marked in respective colours.

Figure 2. English – Hebrew phrase alignment

Figure 2 shows English – Hebrew respective phrase alignment, with a note for the reader that Hebrew text should be read from right to left.

The model applies the segmentation of the input into sequences of words, which are called phrases, and then each phrase is translated into English phrases that could later be reordered in the output. Such a model ensures the correspondence between the units of phrases. After being extracted, all the possible translations were manually filtered to reject the wrong translation variants and prepare the data for the machine analysis stage. This helped us extract sentences with translations of the researched discourse markers from the target language corpus and analyse their use.

While analysing the data, we noticed that there was a small amount of data left which did not fit the variations of possible translations. The first supposition was that it might represent the cases of omissions; however, we decided to analyse it closely to verify. We checked manually the extracted non-attached data and established that most of the analysed cases involved omission with some minor grammatical transformation cases, incorrect translations, and some phrases not included in the possible translations by the machine.

Research findings

Multiword discourse markers in the corpus

The most frequent multiword expressions used in the study corpus have been extracted and are presented in the table below.

Table 1. Most frequent multiword expressions in the corpus

Multiword expression	Frequency
I think	580
You know	573
That is	370
Of course	312
You see	287
In fact	256
I mean	199

For example	161

It could be seen in Table 1 that the two most frequent multiword expressions in the corpus are *I think* and *you know*.

As mentioned earlier, multiword expressions needed to be manually annotated, spotting the cases when the expressions were used as discourse markers. The manual annotation revealed that some multiword expressions are used as discourse markers more frequently while others more often used as content words fully integrated into sentences.

Table 2. Most frequent multiword expressions used as discourse markers

Multiword expression	Used as discourse marker	Content word
I think	473	107
You know	380	193
That is	29	341
Of course	233	79
You see	47	240
In fact	217	39
I mean	168	31
For example	117	44

It is visible in Table 2 that multiword expressions *That is* and *You see* although identified as discourse markers by the theoretical literature, in this study, they demonstrate a weak tendency to be used as discourse markers and are mainly used as content words in the current corpus, while multiword expressions *I think* and *you know* demonstrate a high tendency of being used as discourse markers and the stability of remaining discourse markers in Lithuanian and Hebrew translation.

The translations of discourse marker "I think"

Further, following our research aim, we present a detailed analysis of the translations of the two most frequent multiword expressions used as discourse markers – *I think* and *you know*. The alignment approach allowed extracting direct output of the translations together with the figures of the translation frequency. First, we explore the translations of the most frequent multiword discourse marker, *I think*.

Table 3. Translations of discourse marker *I think*

Lithuanian			
Discourse marker	Translation variants		Number of cases used
I think	Mano manymu	In my opinion	17
	Man atrodo	It seems to me	7

	Man rodos	It seems to me (different derivation)	6
	Mano nuomone	In my opinion (different derivation)	20
	Mano galva	In my head	2
	Aš galvoju	I think	8
	Aš susimąstau	I reflect	1
	Aš tikiu	I believe	1
	Manau	I think (different derivation)	350
	Tikiu	I believe (different derivation)	3
	Atrodo	It seems	4
	Galvoju	I consider	8
	Manyčiau	I would think	1
	Prisimenu	I remember	1
	Omission	48	48
	Grammatical transformation	3	3
Hebrew			
Discourse marker	Translation variants		Number of cases used
I think	אני חושב	I think (male)	215
	אני חושב ש	I think that	4
	אני חושבת	I think (female)	51
	ואני חושב	And I think	70
	אני מאמינה	I believe (female)	1
	אני משוכנע	I am convinced (male)	1
	אני משערת	I assume (female)	1
	אני סבור	I think (male)	17
	אני סבורה	I think (female)	4
	כך אני סבור	So I think (male)	1
	שאני סבור	As I think	1
	דעתי	In my opinion	55
	כמדומני	It seems to me	2
	לטעמי	to one's taste	1
	נדמה	It seems	2

	אבל נראה לי	But it seems to me	2
	נראה לי	It seems to me	13
	Omissions		23
	Grammatical transformation		1
	Missing derivations		6
	Missing phrases		2

The most frequent multiword expression in the researched corpus, *I think*, has a number of translation variants in both researched languages, Hebrew and Lithuanian. The most frequent one in Lithuanian is a one-word expression – an inflected verb, *manau*, which, due to Lithuanian being a highly inflected language (Zinkevičius, Daudaravičius, and Rimkutė 2005), fully represents the verb-pronoun cases. Other one-verb variants and multiword expressions do not demonstrate high. A separate case is represented by omission, which comprises 48 situations, showing that such a technique is also chosen by the translators.

Referring to Hebrew, the most frequent translation is **אני חושב**, which refers to a male derivative, while the female derivate, **אני חושבת**, comprises only 51 cases. The prevalence of male derivatives could be explained by the nature of the Hebrew language, which has the feature that male derivatives are used while addressing purely male and mixed audiences (Tobin 2001). However, Hebrew translation variant choices differ from the Lithuanian ones, as they mostly remain multiword expressions in translation. Another interesting observation in Hebrew is that a number of 70 cases include the additionally integrated connective *and* into the derivative **ואני חושב**. It reveals that sometimes translators prefer inserting additional information into the translation, which could be related not to the direct semantic meaning of addition of *and* but more to the pragmatic inferences drawn by the translators form the surrounding contexts, which relates to the observations of (Blakemore & Carston (1999), and Moeschler (1989). Hebrew demonstrates less omission cases than Lithuanian for the discourse marker *I think* as the number of omissions in Hebrew is 23, almost half of the Lithuanian omission number**.**

The translations of discourse marker 'you know'

Another commonly used multiword discourse marker, *you know*, demonstrates far more variable translations.

Table 4. Translations of discourse marker *you know*

Lithuanian			
Discourse marker	Translation variants		Number of cases used
You know	Na jūs žinot	Just you know	2

	Jūs žinot/e	You know	7
	Kaip žinote	As you know	8
	Jūs suprantat	You understand	2
	Ar ne	Isn't it	3
	Ar žinot	Do you know	2
	Norėtumėte žinoti	You would like to know	1
	Na suprantate	you just understand	2
	Kaip matote	As you see	1
	Bet žinote	But you know	7
	Žinote	You know (different derivation)	116
	Žinot	You know (different derivation)	16
	Na	Particle (just)	71
	išties	right	2
	Žinai	You know (different derivation)	26
	Žinoma	It's known	1
	Matote	You see	3
	Greičiausiai	Probably	1
	juk	Particle (yeah)	5
	žinokite	Just know	1
	suprantama	It's understandable	1
	suprantat	You understand (different derivation)	2
	omission	31	31
	Grammatical transformation	8	8
	Missing derivation	1	1
Hebrew			
Discourse marker	Translation variants		Number of cases used
You know	אתם יודעים	You know (plural, male)	191
	אתן יודעות	You know (plural, female)	2
	אתה יודע	You know (singular, male)	26
	את יודעת	You know (singular, female)	17

	אתם מבינים	You understand (plural, male)	2
	אתם מכירים	You know (plural, male)	1
	כידוע	As you know	1
	Omissions		113
	Grammatical transformation		21
	Missing derivations		5
	Missing phrases		0
	Typo		1

A closer investigation into the translations of discourse marker *you know* reveals that the most common ones in Lithuanian are also one-word verbs žinote/ žinai/ žinot, which represent verb-pronoun cases. Another quite frequent translator choice is the single particle *na*. Although not numerous, very interesting cases of multiword expressions with particles could be found, such as *na jūs žinote* or *na suprantate*, or a single particle *juk*. Even a single particle is used as discourse marker, which is characteristic of the Lithuanian language. There are also cases of multiword expressions involving a connective and inflected verb phrases, for example, *kaip žinote*, *bet žinote*. The translator's choice to additionally use particles or connectives is obviously related not to the translation of semantic meaning but more to the pragmatic meaning inferred by them from the surrounding context. It connotes with the deep observation made by (Nau & Ostrowski (2010) that Lithuanian particles contain the component of subjectivity and inter-subjectivity, and their meaning is mostly coloured by the surrounding context.

In Hebrew, the translation variants for the discourse marker *you know* are not as variable. The most frequent ones, again, are the variants referring to the male gender, including both plural (191) **אתם יודעים** and singular (26) **אתה יודע**, which by far exceeds the number of female derivatives in plural (2) **אתן יודעות** and singular (17) **את יודעת**. In Hebrew, this discourse marker is much prone to omission, as the number of omissions amounts to 113 cases, which are a bit less than the number of the translated cases. Again, multiword expressions remain multiword expressions with just one case of one-word choice in translation.

The translation choices for the multiword expression serving as a discourse marker *you know* are more versatile than those of *I think* and certain cases of grammatical transformation could be observed in the case of the former in Table 5.

Table 5. Grammatical changes in translation of the multiword discourse marker *you know*

Lithuanian			
Discourse marker	Translation variants with grammatical change		Number of cases used
You know	t.y.	That is	1

	Kaip sakiau	As I said	1
	taigi	so	2
	įsivaizduokit	You imagine	1
	laikoma	It is thought	1
	Iš tiesų	really	1
	gerai	okay	1
Hebrew			
Discourse marker	Translation variants with grammatical change		Number of cases used
You know	טוב נו,	colloquial in Hebrew okay, well,	1
	ואז כמובן,	Then of course	1
	לדוגמא,	For example	2
	וכמובן	And of course	1
	הרי	Indeed, therefore	3
	אם יודעים	If you know (plural, male)	1
	כאילו	As if	2
	ואנו יודעים	And we know (plural, male)	1
	נחשו מה,	guess what	1
	נוטים להיות	Tend to be	1
	למעשה	In fact	1
	איך לומר	How to say	1
	נו	well	1
	ברור	clearly	1
	תראו	look	1
	לידיעתכם	For your information	1
	ככה	This way	1

In Lithuanian, eight cases of grammatical changes were found and, even amongst those, one-word discourse connectives prevail. The multiword discourse marker *you know* is translated also into a conjunction, *taigi* (so), and adverbs *gerai* (okay) and *iš tiesų* (really). However, such translator choices are absolutely rare, considering the size of the dataset.

The grammatical transformation cases are more numerous, comprising of 21 occurrences, and much more versatile in Hebrew. The most interesting cases include: **טוב נו,** (okay), which is a usual colloquial saying in Hebrew, **נחשו מה,** (guess what), and two conjunctions used successively, **כאילו** (as if). There are also some cases when a conjunction is just added as in the following example**, ואז כמובן** (then of course), which could be done by the translator simply to stress the discourse management role of the discourse marker used or possibly attaches a rhetorical function to the integrated conjunction. Even among the limited cases of grammatical transformation, multiword expressions as discourse markers prevail in

Hebrew. What is similar to Lithuanian is that there are also adverbs used in the Hebrew translation: **הרי** (indeed), **נו** (well), **ברור** (clearly). Reflecting why different discourse markers demonstrate different translation choices could be based on the nature of the target language into which the texts are translated; for example, Lithuanian is rich in particles and, as the analysis has demonstrated, translators choose to additionally integrate particles into discourse markers to add supplementary discourse expressions**.**
In Hebrew, the male gender prevails in translation, and translators automatically give preference to male derivatives as in English; the gender is not expressed and the choice of the gender of the derivative is completely the translator's choice. Another observation regarding Hebrew is that multiword discourse markers remain multiword because of the translator choice to relay more on word for word translation, while in Lithuanian there is a tendency to omit the pronoun by using just an inflected verb, and this way, multiword discourse markers turn into one-word discourse markers.

Conclusions

The study results showed that English multiword expressions 'I think' and 'you know', identified as discourse markers according to Maschler and Schiffrin (2015) function-based approach, remain discourse markers in Lithuanian and Hebrew translation but they demonstrate variability in Lithuanian and Hebrew translations: they are either translated into multiword expressions or one inflected word, or they are completely omitted. In Hebrew, the translation of multiword discourse markers prevail, and there is a clear tendency for translators to give preference to male over female derivatives, which is due to the nature of the Hebrew language (Tobin 2001). However, it should be stressed that, in Lithuanian, there is a clear tendency observed for one-word discourse markers in translation. One-word translations mainly include verbs, for example, *žinote; suprantate, įsivaizduojate*, which, due to Lithuanian being a highly inflected language (Zinkevičius, Daudaravičius, and Rimkutė 2005), fully represent the verb-pronoun cases. It should be noted that Lithuanian translations of pronoun-verb multiword expressions and one-word verb cases could be considered almost word-for-word translations.
More interesting cases include translator choices of particle-verb or connective-verb multiword expressions, which, due to the use of additionally integrated particles and conjunctions, also carry out certain additional discourse meaning. For example, in Lithuanian, the multiword expression discourse marker *you know* splits into a number of multiword expressions and one-word translations. Multiword expressions could be classified into cases representing pronoun-verb phrases – *jūs žinote, jūs suprantate, jūs įsivaizduojate, jūs esate girdėję* – (which do not have additional colouring), particle-verb phrases – (*na/juk/ir*) *žinote, suprantate* – or connective-verb phrases – (*kaip, kad*) *žinote, matote* – in which connectives could be used in a pre- or post-position relative to the verb (which carry additional discourse meaning due to the integrated particle or connective). In addition, in Hebrew translations, the connective *and* is integrated into the derivate in quite a significant number of occurrences, and there are cases of integration of other connectives. The integration of particles for Lithuanian and connectives for both languages evidently carries the pragmatic meaning that could have been inferred from the surrounding contexts by the translators (Nau and Ostrowski 2010; Blakemore and Carston 1999; Moeschler 1989). Concerning discourse layer, based on the results of the current study revealing the cases

where translators chose to insert particles in Lithuanian and connectives in Hebrew, both of which carrying a certain additional discourse meaning in the translation, it seems that translator choices might be also guided by the inner discourse managing system of the target language.

Referring to omissions, they are moderate in number except for surprisingly high occurrences of *you know* omissions in the Hebrew translation, which could be explained by the fact that such a discourse connective is not naturally used in Hebrew. Consequently, translators choose either omission or grammatical transformation, which is also a bit higher in number in this case.

Future research

The translator's choice to insert particles and connectives needs closer investigation and might be studied in future research. Furthermore, keeping in mind that each language is a unique system with unique features, research could be carried out without English as a pivotal language, which means furthering the current research and using linguistically linked open data (LLOD) and thus accessing related linguistic data directly and comparing the languages. This has already been done for related languages; for example, Snyder et al (2010) analysed Ugaritic (an ancient Semitic language spoken in the second millennium BCE) through resources originally developed for Hebrew. However, linked data provide a sound basis and potential for interoperable resources relating across various languages and enable research across languages and areas**.**

References:

Aharoni, Roee, Melvin Johnson, and Orhan Firat. 2019. "Massively Multilingual Neural Machine Translation." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 3874–84.

Al-Saif, Amal, and Katja Markert. 2010. "The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic." In *LREC*, 2046–53.

Asher, Nicholas, Nicholas Michael Asher, and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Baker, Mona. 2018. *In Other Words: A Coursebook on Translation*. Routledge.

Barlow, Michael. 2011. "Corpus Linguistics and Theoretical Linguistics." *International Journal of Corpus Linguistics* 16 (1): 3–44.

Biber, Douglas. 2006. "Stance in Spoken and Written University Registers." *Journal of English for Academic Purposes* 5 (2): 97–116. https://doi.org/10.1016/j.jeap.2006.05.001.

Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. “If You Look At…: Lexical Bundles in University Teaching and Textbooks.” *Applied Linguistics* 25 (3): 371–405.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1994. “Corpus-Based Approaches to Issues in Applied Linguistics.” *Applied Linguistics* 15 (2): 169–89.

Biber, Douglas, Stig Johansson, Geoffrey Leech, S. Conrad, Eclwarcl Finegan, and Randolph Quirk. 1999. “Longman.” *Grammar of Spoken and Written English*.

Bieliauskienė, Roza. 2012. “Vilnius–Jidiš Kalbos Jeruzalė.” *Krantai*, no. 4: 56–61.

Blakemore, Diane, and Robyn Carston. 1999. “The Interpretation of And-Conjunctions.” *Iten, C. &*.

Cettolo, Mauro, Christian Girardi, and Marcello Federico. 2012. “Wit3: Web Inventory of Transcribed and Translated Talks.” In *Conference of European Association for Machine Translation*, 261–68.

Chu, Chenhui, Raj Dabre, and Sadao Kurohashi. 2017. “An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation.” In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 385–91.

Cobb, Tom, and Alex Boulton. 2015. *Classroom Applications of Corpus Analysis*.

Crystal, David. 1988. “Another Look at, Well, You Know….” *English Today* 4 (1): 47–49.

Degand, Liesbeth, and Henk Pander Maat. 2003. “A Contrastive Study of Dutch and French Causal Connectives on the Speaker Involvement Scale.” *LOT Occasional Series* 1: 175–99.

Dixon, Robert MW. 2009. “The Semantics of Clause Linking in Typological Perspective.” *The Semantics of Clause Linking: A Cross-Linguistic Typology*, 1–55.

Dobrovoljc, Kaja. 2017. “Multi-Word Discourse Markers and Their Corpus-Driven Identification: The Case of MWDM Extraction from the Reference Corpus of Spoken Slovene.” *International Journal of Corpus Linguistics* 22 (4): 551–82.

Dupont, Maïté, and Sandrine Zufferey. 2017. “Methodological Issues in the Use of Directional Parallel Corpora: A Case Study of English and French Concessive Connectives.” *International Journal of Corpus Linguistics* 22 (2): 270–97.

Fraser, Bruce. 2009. “An Account of Discourse Markers.” *International Review of Pragmatics* 1 (2): 293–320.

Furkó, Péter, and Ágnes Abuczki. 2014. “English Discourse Markers in Mediatised Political Interviews.”

Granger, Sylviane. 2015. “Contrastive Interlanguage Analysis: A Reappraisal.” *International Journal of Learner Corpus Research* 1 (1): 7–24.

Hasselgren, Angela. 2002. “Learner Corpora and Language Testing: Smallwords as Markers of Learner Fluency.” *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 143–74.

Hoang, Vu Cong Duy, Philipp Koehn, Gholamreza Haffari, and Trevor Cohn. 2018. “Iterative Back-Translation for Neural Machine Translation.” In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, 18–24.

Hoek, Jet, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted JM Sanders. 2017. “Cognitive Complexity and the Linguistic Marking of Coherence Relations: A Parallel Corpus Study.” *Journal of Pragmatics* 121: 113–31.

Huang, Lan Fen. 2011. “Discourse Markers in Spoken English: A Corpus Study of Native Speakers and Chinese Non-Native Speakers.” PhD Thesis, University of Birmingham.

Joseph, John E. 2009. “Why Lithuanian Accentuation Mattered to Saussure.” *Language & History* 52 (2): 182–98.

Joslyn-Siemiatkoski, Daniel. 2007. “The Cambridge History of Judaism: The Late Roman-Rabbinic Period.” *Theological Studies* 68 (4): 924.

Khayrallah, Huda, Brian Thompson, Kevin Duh, and Philipp Koehn. 2018. “Regularized Training Objective for Continued Training for Domain Adaptation in Neural Machine Translation.” In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, 36–44

.

Koehn, Philipp. 2005. “Europarl: A Parallel Corpus for Statistical Machine Translation.” In *MT Summit*, 5:79–86. Citeseer.

Koehn, Philipp, Franz J. Och, and Daniel Marcu. 2003. “Statistical Phrase-Based Translation.” University of Southern California Marina Del Rey Information Science Inst.

Lefer, Marie-Aude, and Natalia Grabar. 2015. “Super-Creative and over-Bureaucratic: A Cross-Genre Corpus-Based Study on the Use and Translation of Evaluative Prefixation in TED Talks and EU Parliamentary Debates.” *Across Languages and Cultures* 16 (2): 187–208.

Mann, William C., and Sandra A. Thompson. 1988. “Rhetorical Structure Theory: Toward a Functional Theory of Text Organization.” *Text* 8 (3): 243–81.

Maschler, Yael, and Deborah Schiffrin. 2015. “Discourse Markers: Language, Meaning, and Context.” *The Handbook of Discourse Analysis* 2: 189–221.

Moeschler, Jacques. 1989. “Pragmatic Connectives, Argumentative Coherence and Relevance.” *Argumentation* 3 (3): 321–39.

Nattinger, James R., and Jeanette S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford University Press.

Nau, Nicole, and Norbert Ostrowski. 2010. “Background and Perspectives for the Study of Particles and Connectives in Baltic Languages.” *Particles and Connectives in Baltic*, 1–37.

O’Donnell, William R., and Loreto Todd. 2013. *Variety in Contemporary English*. Routledge.

Oleskeviciene, Giedre Valunaite, Deniz Zeyrek, Viktorija Mazeikiene, and Murathan Kurfalı. 2018. “Observations on the Annotation of Discourse Relational Devices in TED Talk Transcripts in Lithuanian.” In *Proceedings of the Workshop on Annotation in Digital Humanities Co-Located with ESSLLI*, 2155:53–58.

Pit, Mirna. 2007. “Cross-Linguistic Analyses of Backward Causal Connectives in Dutch, German and French.” *Languages in Contrast* 7 (1): 53–82.

Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi, and Bonnie L. Webber. 2008. “The Penn Discourse TreeBank 2.0.” In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*. Marrakech, Morocco: European Language Resources Association (ELRA).

Schiffrin, Deborah. 1987. *Discourse Markers*. 5. Cambridge University Press.
———. 2001. “Discourse Markers: Language, Meaning, and Context.” *The Handbook of Discourse Analysis* 1: 54–75.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford University Press.

Siyanova-Chanturia, Anna, Kathy Conklin, and Walter JB Van Heuven. 2011. “Seeing a Phrase ‘Time and Again’ Matters: The Role of Phrasal Frequency in the Processing of Multiword Sequences.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37 (3): 776.

Snyder, Benjamin, Regina Barzilay, and Kevin Knight. 2010. “A Statistical Model for Lost Language Decipherment.” In . Association for Computational Linguistics.

Stede, Manfred, Stergos Afantenos, Andreas Peldzsus, Nicholas Asher, and Jérémy Perret. 2016. “Parallel Discourse Annotations on a Corpus of Short Texts.” In *10th International Conference on Language Resources and Evaluation (LREC 2016)*, 1051–58.

Svartvik, Jan. 1980. “Well in Conversation.” *Studies in English Linguistics for Randolph Quirk* 5: 167–77.

Tan, Xu, Yi Ren, Di He, Tao Qin, Zhou Zhao, and Tie-Yan Liu. 2018. “Multilingual Neural Machine Translation with Knowledge Distillation.” In *International Conference on Learning Representations*.

Tobin, Yishai. 2001. “Gender Switch in Modern Hebrew.” Gender across Languages: The Linguistic Representation of Women and Men 1: 177–98.

Webber, Bonnie, and Aravind Joshi. 2012. “Discourse Structure and Computation: Past, Present and Future.” In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*, 42–54.

Webber, Bonnie, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2016. “A Discourse-Annotated Corpus of Conjoined VPs.” In *Proceedings of the 10th Linguistic Annotation Workshop Held in Conjunction with ACL 2016 (LAW-X 2016)*, 22–31.

Wei, Naixing, and Jingjie Li. 2013. “A New Computing Method for Extracting Contiguous Phraseological Sequences from Academic Text Corpora.” *International Journal of Corpus Linguistics* 18 (4): 506–35.

Wray, Alison. 2012. “What Do We (Think We) Know about Formulaic Language? An Evaluation of the Current State of Play.” *Annual Review of Applied Linguistics* 32 (1): 231–54.

Xiong, Hao, Zhongjun He, Hua Wu, and Haifeng Wang. 2019. “Modeling Coherence for Discourse Neural Machine Translation.” In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:7338–45.

Zeyrek, Deniz, Amália Mendes, Yulia Grishina, Murathan Kurfalı, Samuel Gibbon, and Maciej Ogrodniczuk. 2019. “TED Multilingual Discourse Bank (TED-MDB): A Parallel Corpus Annotated in the PDTB Style.” *Language Resources and Evaluation*, 1–27.

Zhang, Yuqi, Kui Meng, and Gongshen Liu. 2019. “Paragraph-Level Hierarchical Neural Machine Translation.” In *International Conference on Neural Information Processing*, 328–39. Springer.

Zinkevičius, Vytautas, Vidas Daudaravičius, and Erika Rimkutė. 2005. “The Morphologically Annotated Lithuanian Corpus.” In *Proceedings of The Second Baltic Conference on Human Language Technologies*, 365–70.

Zufferey, Sandrine, and Bruno Cartoni. 2012. “English and French Causal Connectives in Contrast.” *Languages in Contrast* 12 (2): 232–50.

Zufferey, Sandrine, and Liesbeth Degand. 2017. “Annotating the Meaning of Discourse Connectives in Multilingual Corpora.” *Corpus Linguistics and Linguistic Theory* 13 (2): 399–422.

Giedrė Valūnaitė-Oleškevičienė
Mykolas Romeris University
Ateities 20
LT-08303, Vilnius, Lietuva
e-mail: gvalunaite@mruni.eu

Chaya Liebeskind
Jerusalem College of Technology
21 Havaad Haleumi str.
9116001, Jerusalem, Israel
e-mail: liebchaya@gmail.com

In SKASE Journal of Translation and Interpretation [online]. 2022, vol. 15, no. 2 [cit. 2022 12-12]. Available online at http://www.skase.sk/Volumes/JTI23/pdf_doc/03.pdf. ISSN 1336-7811.