USAGE OF FILLER AND "EMPTY" WORDS IN ENGLISH AND UKRAINIAN

Karaban Viacheslav
Doctor of Philological Sciences,
Full professor at the department of theory and practice of translation from English,
Institute of Philology,
Taras Shevchenko National University of Kyiv
60, Volodymyrska Str., Kyiv, Ukraine

Verba Lidiia
Candidate of Philological Sciences,
Full professor at the department of English philology and intercultural communication,
Institute of Philology,
Taras Shevchenko National University of Kyiv
60, Volodymyrska Str., Kyiv, Ukraine

Karaban Anna
Candidate of Philological Sciences,
Associate professor at the department of English philology and intercultural communication,
Institute of Philology,
Taras Shevchenko National University of Kyiv
60, Volodymyrska Str., Kyiv, Ukraine

This paper looks into correspondences of filler use in English and Ukrainian, their semantics, and the problem of their translatability. The study was carried out under the working hypothesis that such speech disfluencies as filler and empty words are paid little, if any, attention by translators either for lack of language competence or in an attempt to produce a more fluent target text. The assumption for this study was that the English language is more lax and forgiving of speech disfluencies as well as less restrictive in terms of filler use. The research attempts to explain the discrepancies in filler and empty word use and methods of their translation employing a corpus-based approach. The findings show such weak areas in translating filler and empty words as monocomponent fillers enjoying a higher rate of translatedness as opposed to multicomponent elements, with onomatopoeic fillers remaining the biggest challenge; at the same time, the lexical density and variety of fillers in target languages experiences a decrease while inaccuracy of the equivalents increases. To amend the aforementioned practices the paper uses its conclusions to provide the basis for the following changes to be made in this area, namely: compiling larger corpora of a non-normative language (parallel and otherwise), adding an emphasis in translator training as to limiting their role in normalization and sanitation of the target text. The problem of overtranslation and disregard of translating pragmatic meaning as opposed to purely lexical were observed in the material studied and are viewed as potential avenues for future research.
Key words: *speech disfluency, translatability, filler words, corpus-based approach, semantics, pragmatic meaning, lexical density and variety, English-Ukrainian translation.*

Introduction. Language competence comes with some strings attached, recognizing disfluencies and interpreting them correctly, to name just a few. Filler and empty words have become somewhat of languages' dark secrets that everybody knows about, but nobody acknowledges or addresses. They make speech disfluent, they bring down the level of conversation and act as shortcuts in stereotyping. In spoken language, they are just unavoidable, even in fiction, movies, and TV, they are just there, hanging around to create an illusion of natural speech. So how do we go about translating them? How do we bring the element of spontaneity and normalcy into the target text? Or even should we? This research will hopefully shed some light on the filler use and translation as well as their translatability altogether.

Theoretical framework. The study of the filler word phenomenon hails dates back to the time of Leonard Bloomfield's *Language* [1, 186] where he introduced them as "hesitation forms", viewing fillers from the perspective of syntax, and positing that «when a speaker hesitates, English and some other languages offer special parenthetic hesitation-forms».

The recent years both in the English-speaking and non-English speaking countries (Ukraine included) have seen a significant upsurge in the use of fillers, otherwise referred to as 'pause fillers', 'empty words', 'hesitation forms' or, as they are popularly called in Ukraine, 'parasite words'. We regard filler words as semantically eroded (to a varying degree) words and phrases that 'fill' pauses in speech arising from hesitation, nervousness or the speaker's lack of linguistic competence.

The problem of filler equivalence across languages comes with a set of questions which this study tries to sort out: does their use across languages carry the same connotation, say, that of a register change, or that of a marker of a social group, or simply relaying the lack of control people experience under the emotional influence; how does that affect their translatability and to what extent equivalence can be achieved?

The pragmatic nature of fillers or discourse particles, as suggested by Siegel [7, 5] is such of "higher order speech acts, mediators between mentality and the real world, part of a participation framework, tacit performatives or evincives". While in English discourse particles are commonly accepted to have a dual semantic and pragmatic role and their use, although frowned upon, is still prevalent and overshadowed by the importance of the role they play in speech, in Ukrainian usage fillers are far less accepted, and although they serve similar semantic and pragmatic functions, speakers are more likely to sacrifice those for the sake of avoiding disfluencies. Ukrainian linguists go as far as hypothesizing that the filler infestation in Ukrainian is to a major degree a result of the intervention of other cultures and languages, the primary sources being, at a rough estimate, Russian and English. Masko [3,35] maintains that "a strong intervention by foreign cultures (mainly Russian and Anglo-American) leads to an unreasonably intensive use of borrowed … fillers and exclamations of foreign origin." In a recent contrastive study, Pradana [4, 468-476] states the prospective interest of filler production to translation studies, in particular, in the realm of interpretation.

The basic premise of complete equivalence for fillers in English in Ukrainian is impossible, based solely on the difference in idiosyncrasies of languages. While they belong to the same discursive category, their role and function have rather different implications across the languages. Thus, this foray will focus on where these discrepancies stem from as well the prospective semantic and pragmatic equivalents for translation of fillers from English into Ukrainian and vice versa.

The aim and methods of research. Since this is preliminary research, and building substantial parallel corpora is a long-term undertaking even for a team of scholars, not to mention individuals, for the purposes of this study small, sample versions of such prospective parallel corpora were used (both Ukrainian-English and English-Ukrainian corpora are SL to TL fiction translation ones). To come up with the solution to the problem of filler translatability and translation practice, we have forayed into a contrastive study of parallel English-Ukrainian corpus to establish the most commonly used correspondences for such discursive particles, as well as the adequacy or necessity of such counterparts.

To observe the frequencies of filler use in English and Ukrainian, we turned to corpus methodology. The first stage included registering the filler use in comparable English and Ukrainian corpora. The next step required the collection of sample English-Ukrainian and Ukrainian-English corpora to compare the ratio of these frequencies. The final stage of the research involved the analysis of similarities/dissimilarities and factors influencing those.

The hypothesis for this research stage was that the English language is characterized by more frequent uses of fillers and lesser restrictions on their use. To verify this premise, we used the fiction sections of the British National Corpus (1) and the Corpus of the Ukrainian Language (2), obtaining the following results (the most representative, i.e. frequently used fillers, are provided below):

Table 1
English fillers in The British National Corpus

Filler	Frequency	Density
well	23379	0.1375
you know	12139	0.0714
i mean	5319	0.0312
like	4176	0.0245
er	868	0.0052
um (uhm)	646	0.0038
huh	534	0.0031

Table 2
Ukrainian fillers in The Corpus of the Ukrainian language

Filler	Frequency	Density
ну	5992	0.0221
знаєш (знаєте)	2455	0.0090
е (ем)	259	0.0038
хм	112	0.0004

The results obtained made it possible to arrive at the following conclusions: English fillers

are far more frequently used and, hence, more acceptable in fiction; the use of English fillers is more balanced, while in Ukrainian there is an observable preference for the use of the "all-purpose" filler word *"ну"*; English fillers are much more varied, with words and phrases favored over sound representations, meanwhile, the Ukrainian language leans towards mono- and disyllabic words and sound-representations.

While translation scholars view the elimination of fillers as part of translation "cleanup" or, as Kenny [2, 2-4] puts it, "sanitation", the distribution of frequencies is more revealing in the nature of fillers, and their use might better account for the translator's reluctance to use fillers as more of a language use competence, rather than a subjective drive to normalize the target text.

The next stage in dealing with filler frequencies was comparing the above results with frequency percentages from sample parallel corpora. A sample English-Ukrainian parallel corpus was based on fiction translations from English to Ukrainian and comprises the source text portion (347,231 words) and the target text portion (28,2905 words), while a sample Ukrainian-English parallel corpus was based on fiction translations from Ukrainian into English, and comprises the source text portion (281,109 words) and the target text portion (246,687 words). Therefore, before even looking into whether the elimination of fillers is as rampant a phenomenon as it is hypothesized to be, the "shrinking" of the target text is quite significant – in the English-Ukrainian corpus it is almost 19 %, and in the Ukrainian-English, roughly 12.5 %, so while language use norms might be driving the "sanitation" of the target text (elimination not exclusive to just fillers, but profanities, etc.), the reduction in the target text size is a bigger trend overall.

To look into whether trends in use frequencies correlate not only between sample corpora but established corpora versus sample parallel ones, the data is presented and analyzed separately.

Table 3

Filler ratios in the sample English-Ukrainian parallel corpus

English filler	Fre-quency	Density	Ukrainian filler	Fre-quency	Density
well	364	0.1048	ну	218	0.0770
you know	54	0.0155	знаєш (знаєте)	46	0.0162
i mean	48	0.0013	е (ем)	9	0.0003
like	21	0.0060			
er	27	0.0077			
um(uhm)/ huh	2	0.0004			

The frequencies for the top three English fillers are in line with the data from the BNC (1), although slightly deviating to a lower frequency of the filler use, while sound-representing fillers account for a comparatively larger number. The target text data is more revealing in a sense that the frequencies of the filler use are more similar to those of the source text rather than the data from the Corpus of the Ukrainian language (2), which in turn could testify to the fact that, although the elimination does occur (the numbers are lower in the target text by 15 to 41%), the frequency of the filler use in the target text is higher than in original Ukrainian texts. That is why moving forward we should take into account the premise that these could be the signs of translators' overcorrecting, and rather than working within the constraints and limitations of the target language filler use, they opt for overtranslation and favor equivalence on the lexical, rather than pragmatic, level.

Table 4

Filler ratios from the sample Ukrainian-English parallel corpus

Ukrainian filler	Fre-quency	Density	English filler	Fre-quency	Density
ну	194	0.0690	well	225	0.0912
знаєш (знаєте)	45	0.0160	you know	63	0.0255
			i mean	3	0.0001

In the Ukrainian-English parallel corpus, trends are more difficult to reveal, as seen from the corpus data above. The Ukrainian language (in fiction, in particular) is more resistant to the use of fillers on the whole, although it should be noted that the frequencies are more reminiscent of those in the English-Ukrainian sample corpus, rather than the accepted use of fillers (which is still similar, but lower). What is notable here is that the process of the target text "sanitation" is much more evident, while, generally, English is characterized by a higher frequency of the filler use, for the English target text showed a drastic decrease in sound-representing filler-words which were eliminated completely from the target text. This could mean that while translators still undertranslate, on the opposite end, we can observe that the frequencies of the filler use in the target text are higher than in the source text by 14-29%. This is a significant upswing, considering that "sanitation" would have meant quite the opposite. Thus, we do not regard filler words as subject to a "cleanup" in the target language; contrarily, they are rather more in line with one of the translation universals, normalization, as defined by Scott [5, 112; 6] – "the translator's

sometimes conscious, sometimes unconscious rendering of idiosyncratic text features in such a way as to make them conform to the typical textual characteristics of the target language".

Hence, our assumption would be that translation "sanitizes" the target text exclusively in sound-representing fillers, not for the target text to fit in with the characteristics of target language use, but, rather, for the text to fit in with the characteristics of written English, rather than English oral speech, thus normalizing writing, which is not necessarily required and could be one of the areas for improvement, i.e., actually translating or introducing a sound-representing filler when the target language use would normally prompt their use.

Research findings. If we delve into practicalities of translation, we can observe general strategies for the filler and empty word translation from English to Ukrainian (eliminating up to 40-50% of fillers, without further introducing extra fillers into the target text) and from Ukrainian to English (while some fillers can be removed, the majority of those should be translated, while in other contexts that normally trigger the filler use in English, up to 30-40% could be added in the target text).

Returning to semantic equivalents, it should not go unnoticed that within those general strategies more precise tactics are still to be put forward, particularly regarding the contexts and frequencies where fillers could be eliminated/added. In this aspect, we turned to contextual dictionary Glosbe (3) based on translation memories, i.e., parallel corpora, to try to establish traditionally acceptable correspondences for filler and empty words in translation from English to Ukrainian:

1) Based on the results, the following conclusions were drawn - in roughly 50% of the instances, *"well"* is in fact translated as *"ну"*, although with a register restriction, i.e., the target text is open for fillers only if it itself is in low/colloquial register; in 40% of the instances, there is no lexical equivalent, but in roughly 24%, translation acquires an emphatic tone due to syntax or lexicon change; the rest 10% account for varied contextual equivalents *"що ж"*, *"так"*, *"вочевидь"* and arise rather out of the need for an equivalent than the context of the target text prompting the use of the filler.

2) *"I mean"* is translated in 88% of the instances of its use predominantly via word-for-word translation of the phrase *"Я маю на увазі"* and *"Я розумію"* (which is rather unfortunate, as Ukrainian correspondences are not necessarily pragmatic or discursive equivalents, for they are not discourse particles in Ukrainian, but rather direct semantic equivalents); in this context, we

believe the introduction of a true discourse particle (filler) like *"тобто"*, *"ну"*, *"направду"* is more appropriate (which is true for 12% of translated instances), while, in our opinion, elimination thereof is still preferable.

3) As it is evident from the parallel corpora, the equivalence of *"you know"* and *"знаєш (знаєте)"* is confirmed by the margin of translatability being around 94%; though superficially it is similar to the correspondence above (resembling word-for-word translation), the Ukrainian counterparts have already undergone semantic erosion, i.e., in Ukrainian, they function as fillers as well, probably due to the influence/interference of English (language contamination through translation).

4) English *"like"* does not have a direct translation correspondence in Ukrainian, but this could be easily amended by introducing any Ukrainian filler of similar frequency. Audibly filled pauses like *"uhm"*, *"um"*, *"huh"*, *"er"*, etc., on the whole, remain untranslated, in roughly 17% they are rendered as Ukrainian *"хмм"*, *"хм"*, *"е"*, *"а"*, which is within the idiosyncrasy of Ukrainian, without paying as much attention to registering phonological phenomena in text.

The results of Ukrainian-English correspondences can be summed up as follows:

1) The highest translatability is observed with *"ну"* at approximately 92%, its equivalent being *well* in 89% of the instances of use, and *"so"*, *"now"*, *"you know"* and *"like"* accounting for the rest.

2) The situation with "*знаєш (знаєте)*" is quite similar to *"I mean"*/*"Я маю на увазі"* in a sense that the Ukrainian filler is rendered into the target language almost without a fault, but not always through an equivalent English filler (*"you know"* indeed accounts for the majority of correspondences), but rather a direct semantic equivalent (*"know"* - not as in *"you know"*, *"realize"*) in more than a quarter of the instances.

3) Sound-representing *"ем"*/*"хмм"* are translated in roughly half of the instances via equivalent sound-representing fillers *"uh"*/*"hmm"* (singular use of *"eh"*), which is a place for improvement in doing translations, as this seems to be the weakest point for both English-Ukrainian and Ukrainian-English pairs.

Conclusions and directions for future research. Overall observations of the aforementioned correspondences provide insight into areas where there is room for improvement of filler and empty word translation, both on a higher level (establishing the norm of the filler use and then adhering to it, in line with the universal of normalization) and on a lower one (tracking the prevalent correspondences in filler translation

and assessing their fit within the target language idiosyncrasy, thus suggesting adjustments be made on a case-to-case basis).

As this paper is merely a preliminary inquiry into a corpus approach to translatability of filler and empty words, we argue that there are many avenues of prospective investigation in this field: 1) building larger parallel corpora to cross-check the findings and verifying if the hypothesis still holds true when testing on more representative corpora; 2) foraying into spoken-language filler use and translation in English and Ukrainian could provide a better toolkit for recognizing and dealing with speech disfluencies in machine translation with the abovementioned language pair; and, finally, 3) tracking filler use and the corresponding translation patterns could be of value in machine translation for a currently fuzzy area of a non-normative language.

BIBLIOGRAPHY
1. Bloomfield L. Language. Chicago: University of Chicago Press, 1933. 564 p.
2. Kenny D. Creatures of Habit? What Translators Usually Do with Words. *Meta*. 2002. Vol. 43, no. 4. P. 515–523. URL: https://doi.org/10.7202/003302ar (date of access: 01.10.2021).
3. Масько Д. Український сленг та міжмовна інтерференція. *Культура слова*. 2003. Т.62. С. 34–39. URL: http://kulturamovy.univ.kiev.ua/KM/pdfs/Magazine62-8.pdf (дата звернення: 01.10.2021).
4. Pradana F. Linguistic Fillers in English and Javanese: A Contrastive Analysis. *KnE Social Sciences*. 2019. Vol. 3, no. 10. P. 468. URL: https://doi.org/10.18502/kss.v3i10.3935 (date of access: 01.10.2021).
5. Scott N. M. Investigating normalization in literary translation. 1996. *paper presented at "Looking at language into the millenium" seminar*. Dept. of the English Language, University of Glasgow.
6. Scott M.N. Normalisation and readers' expectations : a study of literary translation with reference to Lispector's A Hora da Estrela : Electronic Thesis or Dissertation. 1998. URL: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.366489 (date of access: 01.10.2021).
7. Siegel M.E.A. Like: The Discourse Particle and Semantics. *Journal of Semantics*. 2002. Vol. 19, no. 1. P. 35–71. URL: https://doi.org/10.1093/jos/19.1.35 (date of access: 01.10.2021).

ILLUSTRATIVE MATERIAL
1. *British National Corpus (BNC)*. English corpora. Retrieved October 1, 2021, from https://www.english-corpora.org/bnc
2. *The Corpus of the Ukrainian language.* Mova.Info. Retrieved October 1, 2021, from http://www.mova.info/corpus.aspx?l1=209
3. *Glosbe Dictionary*. Glosbe. Retrieved October 1, 2021, from https://glosbe.com/

REFERENCES
1. Bloomfield, L. (1933). *Language.* University of Chicago Press. [in English].
2. Kenny, D. (2002). Creatures of Habit? What Translators Usually Do with Words. *Meta*, 43(4), 515–523. https://doi.org/10.7202/003302ar [in English].
3. Masko, D. (2003). Ukrainskyi sleng ta mizhmovna interferentsiya. [Ukrainian slang and cross-cultural interference]. *Kultura slova*, *62*, 34–39. http://kulturamovy.univ.kiev.ua/KM/pdfs/Magazine62-8.pdf. [in Ukrainian].
4. Pradana, F. I. (2019). Linguistic Fillers in English and Javanese: A Contrastive Analysis. *KnE Social Sciences*, *3* (10), 468. https://doi.org/10.18502/kss.v3i10.3935. [in English].
5. Scott, M. N. (1996). *Investigating normalization in literary translation*, paper presented at "Looking at language into the millenium" seminar, Dept. of the English Language, University of Glasgow. [in English].
6. Scott, M. N. (1998). *Normalisation and readers' expectations: a study of literary translation with reference to Lispector's A Hora da Estrela* (Doctoral dissertation, University of Liverpool). https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.366489. [in English].
7. Siegel, M. E. A. (2002). Like: The Discourse Particle and Semantics. *Journal of Semantics*, 19(1), 35–71. https://doi.org/10.1093/jos/19.1.35. [in English].

ВИКОРИСТАННЯ СЛІВ-ПАРАЗИТІВ ТА «ПОРОЖНІХ» СЛІВ У АНГЛІЙСЬКІЙ ТА УКРАЇНСЬКІЙ МОВАХ

Карабан В'ячеслав
доктор філологічних наук,
професор кафедри теорії та практики перекладу з англійської мови
Інституту філології
Київського національного університету імені Тараса Шевченка
вул. Володимирська, 60, Київ, Україна

Верба Лідія
кандидат філологічних наук,
професор кафедри англійської філології та міжкультурної комунікації
Інституту філології
Київського національного університету імені Тараса Шевченка
вул. Володимирська, 60, Київ, Україна

Карабан Анна
кандидат філологічних наук,
доцент кафедри англійської філології та міжкультурної комунікації
Інституту філології
Київського національного університету імені Тараса Шевченка
вул. Володимирська, 60, Київ, Україна

У статті досліджується відповідність використання слів-паразитів у англійській та українській мовах, їх семантика та проблема їх перекладу. Дослідження проводилося на основі робочої гіпотези, що таким збоям мовлення, як слова-паразити та порожні слова, перекладачі приділяють мало уваги або через брак мовної компетенції, або через намагання створити більш плавний цільовий текст. Припущення цього дослідження полягало в тому, що англійська мова є менш нормативною та допускає мовні збої, а також менш обмежує мовця у вживанні слів-паразитів. У статті робиться спроба пояснити розбіжності у використанні слів-паразитів та порожніх слів в англійському та українському мовлення та методах їх перекладу, використовуючи корпусний підхід. Висновки підкреслюють такі слабкі місця у перекладі слів-паразитів та порожніх слів: монокомпонентні слова-паразити мають більшу ймовірність бути перекладеним на відміну від багатокомпонентних елементів, тоді як ономато-пеїчні слова-паразити залишаються найбільшою проблемою; в той же час, щільність та різноманітність таких слів у мовах перекладу зменшується, а натомість неточність еквівалентів зростає. Для того, щоб змінити вищезгадану практику, на основі отриманих висновків пропонуються наступні зміни, які необхідно внести у цій галузі дослідження, а саме: укладання більших корпусів ненормативної мови (паралельних та не тільки) і збільшення уваги у навчанні перекладачів щодо обмеження їх ролі у нормалізації та санації цільового тексту. Аналіз проблем надмірного перекладу та ігнорування передавання прагматичного значення у перекладі з англійської українською мовою на відміну від суто лексичного значення розглядаються як потенційні напрями майбутніх досліджень.

Ключові слова: хезитація, перекладність, слова-паразити, корпусний підхід, семантика, прагматичне значення, лексична щільність і різноманітність, англо-український переклад.