Discourse markers and connectives in interpreted Hungarian discourse: A corpus-based investigation of discourse properties and their interdependence

Andrea Götz[1]

[1] *Károli Gáspár University*

Abstract

This paper investigates the frequency of discourse markers and connectives in interpreted Hungarian discourse. Even though the relationship between these items and other linguistic phenomena, such as hesition, is well known, no study to date has set out to explore it in relation to Hungarian interpreted discourse. This study examines the link between the frequency of these items, delivery speed, and filled pauses in a corpus of European Parliamentary speeches interpreted from English to Hungarian. According to the results, discourse markers and connectives are more frequent in interpreted than original discourse, and their frequency positively and significantly correlates with delivery speed, while filled pauses do not show such a straightforward relationship.

1. Introduction

Discourse markers and connective items (DMCs) have been essential to the investigation of translated discourse due to their role in discourse cohesion. They have also been at the centre of attention in the rapidly developing field of corpus-based interpreting studies. Corpus-based interpreting studies has experienced a growth spurt in recent years, going from a "cottage industry" to a "booming research field" (Bendazzoli, 2018; Bendazzoli et al., 2018), which resulted in numerous studies on a wide variety of topics. A number of findings have emerged with regard to DMCs in simultaneous interpreting:

- DMC frequency increases in interpreting relative to source speeches (Defrancq et al., 2015),

Email address: gotz.andrea@kre.hu (Andrea Götz)

- DMC frequency is higher in interpreted than in translated texts (Defrancq et al., 2015),
- DMC frequency can be higher in interpreted than in non-interpreted speeches (Defrancq, 2018),
- DMCs can be used differently in interpreted than in non-interpreted texts (Defrancq, 2016).

To explain these results, research mostly points to the unique conditions (e.g. time pressure) under which simultaneous interpreters work that inevitably shape interpreted discourse output. DMC frequency differences between interpreted and non-interpreted discourse are understood to come about because interpreters "drastically reshape the discourse structure of the source text" (Defrancq et al., 2015). However, as the properties of spoken and interpreted discourse alike are the result of the complex interplay of many factors, it stands to reason that DMCs should be studied in relation to other discourse factors too.

In fact, there is a growing trend in the study of interpreted discourse to investigate the interrelationship of discourse properties (e.g. Collard & Defrancq, 2017; Defrancq & Plevoets, 2018; Plevoets & Defrancq, 2018; Collard & Defrancq, 2019, 2020).

Filled pauses, for example, are known to respond to several factors in interpreted discourse. Filled pause frequency of interpreters increases with factors boosting information load, such as higher delivery speed (Plevoets & Defrancq, 2016), and decreases with factors lowering cognitive load, such as formulaicity, which seems to "free up" cognitive bandwidth (Plevoets & Defrancq, 2018).

As interpreted discourse has been found to contain more hesitation than original discourse in a number of languages (e.g. Plevoets & Defrancq, 2016; Götz, 2018, 2019b; Collard & Defrancq, 2020), and given that DMCs are interlinked with hesitation in structuring spoken discourse (Crible, 2018), forming functionally distinct patterns (Crible, 2017) that perform specific discourse functions (e.g. transition, giving floor, etc.) (Crible et al., 2017), it would be all the more important to understand how filled pauses and DMCs interact in interpreted speech.

Delivery speeds have also been linked to DMCs (Götz, 2019a; Magnifico & Defrancq, 2020). In an exhaustive study of multiple factors (e.g. target language, the speaker's and the interpreter's gender), only delivery speed was found to have a statistically significant impact on DMC frequency (Magnifico & Defrancq, 2020). This result is somewhat surprising since a number of "striking gender differences" have been pinpointed in interpreted discourse (Magnifico & Defrancq, 2020, 6.) that vary with the particular target language.

Gender seems to influence, for example, politeness (Magnifico & Defrancq, 2016) and hedge use (Magnifico & Defrancq, 2017). On the other hand, target language also remains profoundly influential even when gender differences emerge (Magnifico & Defrancq, 2017). Nevertheless, gender has been also clearly ruled out as a decisive factor in an extensive study on hesitation in interpreted speech, finding only limited gender differences (Collard & Defrancq, 2020).

On the whole, however, men are usually observed to hesitate more (Collard & Defrancq, 2017; Götz, 2018), DMC frequency is variably higher among female (Götz, 2019a) or male interpreters (Magnifico & Defrancq, 2020), with delivery speed showing similar variation (cf. Russo, 2018; Götz, 2019a).

With all this complexity, one aspect of interpreted discourse has so far escaped attention: individual differences. It is well understood in Hungarian discourse marker research that age and gender can influence discourse marker choice (Markó & Dér, 2011; Vukov Raffai, 2016; Schirm, 2019), but it is also clear that individual-specific patterns exist in DMC use (e.g. Dér & Markó, 2010; Vukov Raffai, 2016; Schirm, 2019), as well as hesitation marker use (Horváth, 2014), and these can override group-level tendencies. This means that both groups and individuals need to be examined.

This study represents a preliminary investigation into DMC frequency, delivery speed, and filled pauses, including the interrelationships of these properties in English to Hungarian simultaneously interpreted discourse. In addition, the properties of individual discourse output is contrasted with that of groups.

The remainder of the paper is structured as follows. First, the paper presents a brief overview of DMCs in interpreted discourse, proceeding with the data and methods of this study, which is followed by the results and the conclusion. According to the results, DMCs are more frequent in interpreted than in original discourse, and while DMC frequency correlates positively with delivery speed, filled pauses and DMC frequency do not show such a clear relationship.

2. Discourse markers and connectives in corpus-based interpreting studies

A number of misconceptions persist about the role of discourse markers and connectives in interpreted discourse. As DMCs do not contribute to propositional meaning, they are often seen as non-essential, and thus "vulnerable in the interpretation process" (Defrancq et al., 2015, 198.).

Empirical research, however, disproved this received wisdom, finding a high frequency of these items in interpreted discourse (Defrancq et al., 2015), indicating that they play a profound role in re-creating cohesion in interpreting. But beyond linking segments of discourse, DMCs have various role, and therefore can cause considerable problems for interpreters in the "pragmatic aspects of discourse" (Hale, 1999, 57.). DMCs, for example, can influence how speakers are perceived and judged, which can have far-reaching consequences in a legal or political context.

As a rule, discourse markers in interpreted speech are attributed to the speaker and not to the interpreter (Blakemore & Gallai, 2014). Since DMCs in utterance comprehension serve as clues to the cognitive processes of speakers, guiding the interpretation processes of the hearer (Blakemore, 2002), in interpreting, they are perceived to reflect thought processes of the speaker, and not those of the interpreter (Blakemore & Gallai, 2014).

This means that in case DMCs interpreters use are stigmatized, these items can damage the image of the speaker, even if the particular DMCs do not originate from the speaker. Accordingly, "polished" versions of interpreted testimonies, meaning that they had been edited to remove hedges and discourse markers, are evaluated as significantly more competent, credible, and intelligent than unedited versions which contain discourse markers (Hale, 2010).

Another important aspect of DMC use is the particular institutional context in which the discourse itself is created. Most interpreting corpora and studies on EU languages use European Parliamentary (EP) speeches (e.g. European Parliament Translation and Interpreting Corpus (EPTIC) in Bernardini et al. (2016); European Parliament Interpreting Corpus Ghent (EPICG) in Plevoets & Defrancq (2018)). Since EP data looms so large, at least in European corpus-based interpreting studies, it is important to consider the impact of the EP's institutional context, as well as the limitations of these data sets.

The discourse of EP interpreters could converge in some aspects. Case in point, on the basis of DMC frequencies (*well, now, so*) in French, Spanish, and Italian to English interpreting, EP interpreters could be described as forming a discourse community, while EP interpreters and MEPs together could not (Defrancq, 2018). Such patterns could be caused by interpreters adhering to certain institutional norms of discourse (cf. Magnifico & Defrancq, 2020).

In summary, the frequency of DMCs has been observed to increase in interpreting compared to source texts (Defrancq et al., 2015), some DMCs have also been found to be more frequent in interpreted than in non-interpreted, original discourse (Defrancq, 2018), and while gender differences can appear, these are statistically not significant in the studies so far available (cf. Magnifico & Defrancq, 2020). Their frequency, however, positively and significantly correlates with delivery speed (Magnifico & Defrancq, 2020).

3. Research design

3.1. Research goals and hypotheses

The purpose of this study is to examine the frequency of DMCs in interpreted Hungarian discourse and test whether this frequency is related to other factors, such as delivery speeds or the frequency and duration of filled pauses. The study consequently investigates the following hypotheses:

1. DMCs are more frequent in interpreted than in original Hungarian discourse.
2. DMCs are more frequent in Hungarian target speeches than in their English source speeches.
3. Interpreted discourse contains more filled pauses than non-interpreted.
4. DMC frequency is positively correlated with delivery speed.
5. Both filled pause duration negatively correlated with DMC frequency.
6. The discourse output of female and male interpreters shows significant differences.

3.2. Corpora

This study uses three corpora: the Interpreted Hungarian Corpus (IHC), the Original Hungarian Corpus (OHC), and corpus of English source speeches (ESC). The texts of these corpora are sourced from the Hungarian European Parliamentary intermodal corpus (HEPIC) (Götz, 2017). The HEPIC is composed of EP speeches delivered between 2008 and 2012 on plenary sitting days, presently comprising about 230,000 words.

Table 1 shows the properties of the corpora, while Table 2 presents them broken down according to the sex of the speakers. Both corpora contain over one hour's worth of speeches produced by five female and five male speakers. Table 3 displays the properties of the discourse of the individual speakers. For the full data set of the corpora see Appendix 2.

3.3. Methods

This study investigates the frequency of discourse markers (raw frequency, normalized frequency per minute), delivery speeds (number of words per minute), and the frequency and duration of filled pauses. Statistical significance is probed using t-tests, correlation by Pearson's correlation coefficient, sex differences are

Table 1: Properties of the interpreted and the original corpora

	IHC	**OHC**
Speech time	1 hour 18 min 40 sec	1 hour 15 min 56 sec
English speech time	1 hour 16 min 58 sec	
No. of speeches	50	40
No. of words	8064	9174
No. of English words	12,183	
No. of speaker	5 female, 5 male	5 female, 5 male

Table 2: Properties of the interpreted and original corpora by the sex of the speakers

	IHC	**OHC**
Speech time (min) (f)	41 min 30 sec	36 min 27 sec
Speech time (min) (m)	37 min 10 sec	39 min 29 sec
No. of speeches (f-m)	27-23	20-20
No. of words (f)	4240	4469
No. of words (m)	3824	4705

explored with the Mann-Whitney test, and outliers are identified with Grubb's test.

The data of this study are derived from two sources: individual speeches, and the discourse output of individual interpreters. The former is utilized in the descriptive analysis of DMC frequency, rate of delivery speeds, and the frequency and length of filled pauses, while the two are combined to probe correlations.

The frequency DMCs set is examined which contain both traditional conjunctions and discourse markers. These sets are based on the most frequent items and are matched between Hungarian and English. Both DMC sets can found in Appendix 1.

Filled pauses are defined here as vocalisations not contributing to the propositional. Although lengthening vowels and consonants could be classified as filled pauses, this study opts for a more restricted definition by only including schwa-

Table 3: Properties of the interpreted and original by individual speakers

IHC	Speech time	No. of speeches	No. of words
Female interpreter #1	4 min 24 sec	3	488
Female interpreter #2	1 min 52 sec	2	192
Female interpreter #3	6 min 6 sec	5	703
Female interpreter #4	13 min 16 sec	7	1304
Female interpreter #5	15 min 49 sec	10	1553
Male interpreter #1	3 min 44 sec	3	427
Male interpreter #2	20 min 47 sec	13	2083
Male interpreter #3	2 min 9 sec	2	216
Male interpreter #4	6 min 55 sec	3	740
Male interpreter #5	3 min 32 sec	2	358
OHC	**Speech time**	**No. of speeches**	**No. of words**
Female speaker #1	6 min 50 sec	5	879
Female speaker #2	7 min 34 sec	4	899
Female speaker #3	10 min 42 sec	5	1150
Female speaker #4	6 min 13 sec	3	685
Female speaker #5	5 min 6 sec	3	856
Male speaker #1	8 min 7 sec	4	904
Male speaker #2	6 min 14 sec	3	776
Male speaker #3	13 min 2 sec	7	1612
Male speaker #4	5 min 40 sec	3	720
Male speaker #5	6 min 24 sec	3	693

like neutral vowel resembling [ø] or [ə] and its combinations with [m], and [ɒ] type filled pauses. For the annotation of filled pauses Exmaralda Partitur Editor was used.

4. Results and discussion

4.1. Frequency of discourse markers and connective items

Table 4 shows the raw and normalized frequency of DMCs, breaking the data down by the sex of the speakers as well. In the case of the English source corpus (ESC), "female" and "male" refer to the input that Hungarian female and male interpreters were exposed to, and not the sex of the English speakers.

Table 4: The frequency of DMCs

DMCs	IHC	OHC	ESC
Raw frequency	1355	1069	1093
Raw frequency (f)	735	545	573
Raw frequency (m)	620	524	520
Frequency /1 minute	17.2	14.1	14.2
Frequency /1 minute (f)	17.7	15.0	14.1
Frequency /1 minute (m)	16.7	13.3	14.3

Frequency of DMCs is significantly higher in the interpreted Hungarian speeches than in their English sources speeches ($t = -2.379$, $p = 0.01$), or in non-interpreted Hungarian speeches ($t = 3.122$, $p = 0.001$).

Higher frequency in comparison to source speeches can, of course, be influenced by cross-linguistic tendencies, therefore it should not be interpreted as an increase purely due to the effect of interpreting. However, the fact that DMC frequencies are similar across the original Hungarian and English source speeches, but higher in interpreted discourse, might indicate that interpreted discourse makes use of these items differently.

Although female speakers in interpreted and original Hungarian discourse use more DMCs than men, this variation is not statistically significant (interpreters: $z = -0.954$, $p = 0.342$, original speakers: $z = 0.122$, $p = 0.904$).

4.2. Delivery speed

Table 5 shows the delivery speeds of Hungarian and English original speakers and interpreters. Interpreted and original Hungarian delivery rates differ significantly overall (t = -4.793, $p =< 0.000$). From the three groups, interpreters have the lowest delivery speeds. Of course, simultaneous interpreting cannot be expected to produce similar delivery speeds to original speakers who read out loud their speeches.

Table 5: Delivery speeds

	IHC	**OHC**	**ESC**
Words/min	102.5	120.8	158.3
Words/min (f)	102.2	122.6	156.9
Words/min (m)	102.9	119.2	159.9

Interpreters produce approximately 18 words fewer per minute than original speakers. Female and male interpreters show very similar delivery speeds with no statistically significant variation ($z = 0.058$, $p = 0.476$). Female interpreters were exposed to an overall slightly slower English delivery speed than male interpreters.

4.3. Filled pause duration

Table 6 shows the normalised duration (seconds per minute) of filled pauses in interpreted and original Hungarian discourse.

The normalised duration of filled pauses is significantly longer in the IHC than in the OHC ($t = 8.603$, $p =< 0.000$). The duration of filled pauses differs significantly between female and male interpreters ($z = -2.141$, $p = 0.032$).

4.4. Filled pause frequency

Table 7 shows the absolute and normalized frequency of filled pauses in the corpora.

The normalised frequency of filled pauses is significantly higher in the interpreted corpus ($t = 9.233$, $p =< 0.000$) than in the OHC. Unlike duration,

Table 6: The duration of filled pauses

	IHC	**OHC**
Absolute duration of filled pauses (sec)	231.4	30.2
Absolute duration of filled pauses (sec) (f)	110.4	10.1
Absolute duration of filled pauses (sec) (m)	121.0	20.1
Normalised duration of filled pauses (sec)	2.9	0.4
Normalised duration of filled pauses (sec) (f)	2.7	0.3
Normalised duration of filled pauses (sec) (m)	3.3	0.5

Table 7: Frequency of filled pauses

	IHC	**OHC**
Absolute frequency of filled pauses	723	99
Absolute frequency of filled pauses (f)	377	36
Absolute frequency of filled pauses (m)	346	63
Normalised frequency of filled pauses	9.2	1.3
Normalised frequency of filled pauses (f)	9.1	1.0
Normalised frequency of filled pauses (m)	9.3	1.6

frequency is slightly higher for female interpreters than for males, though not significantly ($z = -0.954$, $p = 0.342$).

4.5. Correlation

4.5.1. DMC frequency and delivery speed

Figure 1 presents the correlation between DMC frequency and delivery speed in each speech interpreted by female and male interpreters. DMC frequency and delivery speed show a statistically significant, positive moderate correlation. This suggests that there is a tendency for higher delivery speeds to correlate with higher DMC frequencies. Table 8 presents the results of the correlation tests.

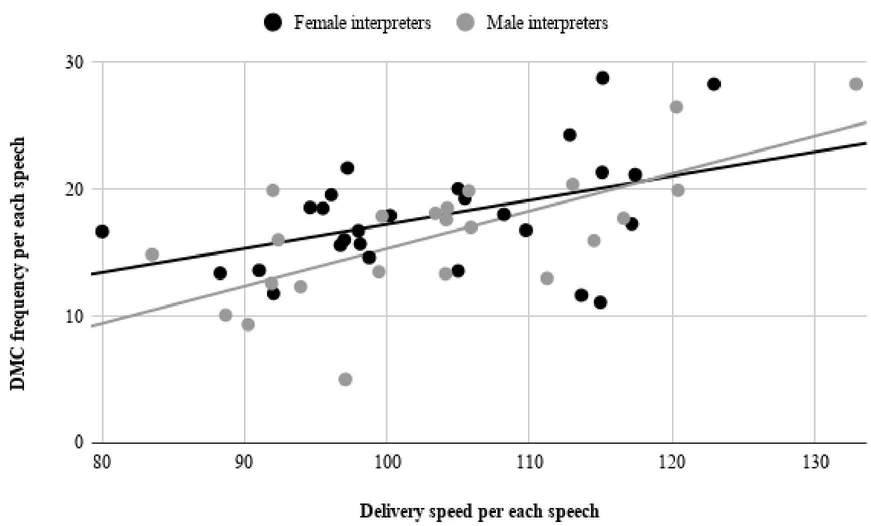

Figure 1: The correlation between delivery speed and DMC frequency in each speech interpreted by female (black) and male (grey) interpreters

Table 8: Pearson's correlation test of DMC frequency and delivery speed

	r	**p**
IHC	0.6	0.0
OHC	0.6	0.0
ESC	0.6	0.0

However, since all corpora show this relationship with very similar, statistically significant r values, this tendency does not exclusively characterize interpreted speech but rather represents a more universal tendency.

For female interpreters, the relationship is weaker ($r = 0.448$, $p = 0.019$) but significant. Among male interpreters, the connection is stronger ($r = 0.691$, $p = 0.000$) and similarly significant.

Since DMCs in English source speeches could have an effect on DMC frequency in interpreted Hungarian discourse, it is necessary to provide a comparison. Figure 2 demonstrates the correlation between the frequency of English DMCs in the source speeches and their corresponding Hungarian values.

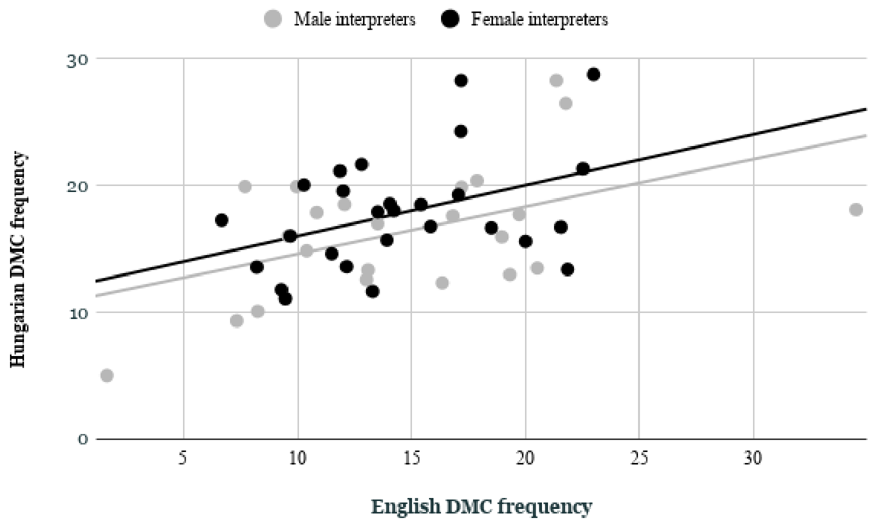

Figure 2: Correlation between English DMC frequency and Hungarian DMC frequency in each speech interpreted by female (black) and male (grey) interpreters

English and Hungarian DMC frequency correlate positively, and though the relationship is weak, it is statistically significant ($r = 0.444$, $p = 0.001$). This relationship is weaker among women ($r = 0.407$, $p = 0.035$) than men ($r = 0.491$, $p = 0.017$) but it is significant in both groups.

Figure 3 shows this correlation in the discourse output of each individual interpreter calculated from their total interpreting output.

When it comes to individual interpreters, there is a positive weak, statistically not significant correlation between DMC frequency and delivery speed ($r = 0.085$, $p = 0.815$), which is stronger for women ($r = 0.087$, $p = 0.889$) and weaker for men ($r = 0.053$, $p = 0.933$). Grubb's test detected no significant outliers in the group, either in terms of delivery speed or DMC frequency.

This might indicate that while there is an overall tendency for higher delivery speeds to correlate with higher DMC frequencies, there is variation and

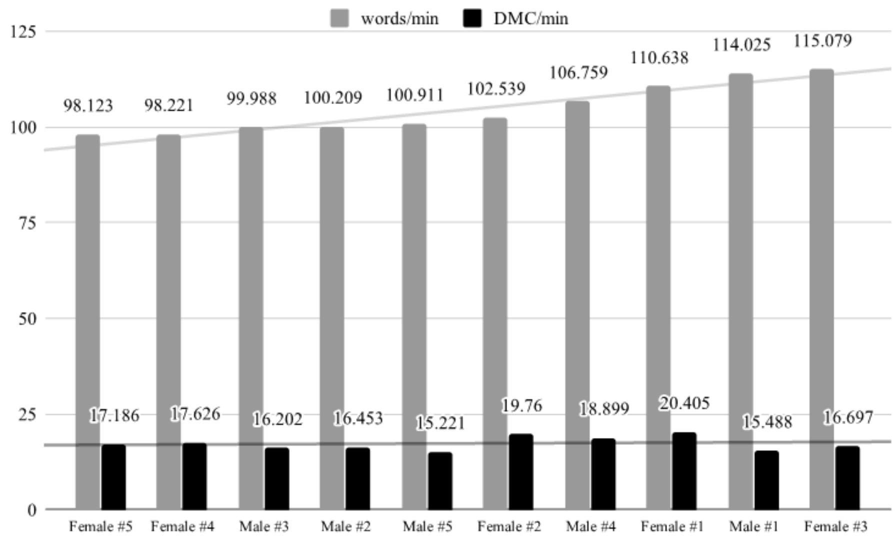

Figure 3: DMC frequency (black) and delivery speed (grey) in the discourse of individual interpreters

divergence from this tendency within the interpreting output of individual interpreters as the performance of interpreters can vary from occasion to occasion.

4.5.2. DMC frequency and filled pause duration

Figure 4 demonstrates the relationship between normalized filled pause duration and DMC frequency. DMC frequency and normalized filled pause duration show a weak, negative, statistically not significant correlation. Table 9 shows the results of the correlation tests.

Table 9: Pearson's correlation of DMC frequency and filled pause duration

	r	**p**
IHC	-0.2	0.2
OHC	0.1	0.6

Filled pause duration and DMC frequency do not correlate significantly or strongly either for either group, but the correlation is positive among women

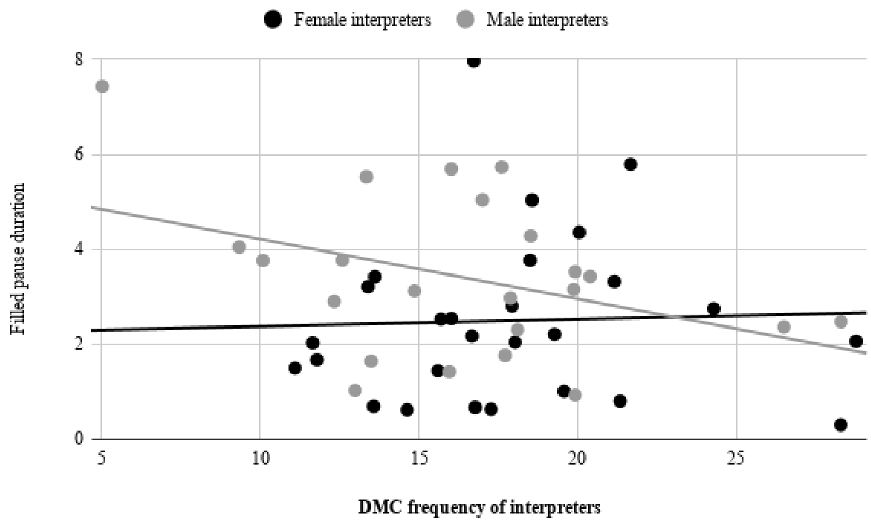

Figure 4: The correlation between DMC frequency and normalized filled pause duration in each speech interpreted by female (black) and male (grey) interpreters

($r = 0.038$, $p = 0.850$) and negative among men ($r = -0.393$, $p = 0.863$). Figure 5 shows the correlation of DMC frequency and normalized filled pause duration in the discourse of individual interpreters.

Looking at the discourse production of individual interpreters, the correlation between DMC frequency and pause duration is weak and positive ($r = 0.134$, $p = 0.712$), both for women ($r = 0.182$, $p = 0.770$) and men ($r = 0.380$, $p = 0.528$). No significant outliers were detected.

4.5.3. DMC frequency and filled pause frequency

Figure 6 shows the correlation between the normalized frequency of DMCs and filled pauses. Filled pause frequency and DMC frequency correlate negatively and weakly, not forming a statistically significant relationship. Table 10 presents the results correlation tests.

For female interpreters, the relationship is positive and weak ($r = 0.099$, $p = 0.623$), while for men, the correlation is negative and weak ($r = -0.227$,

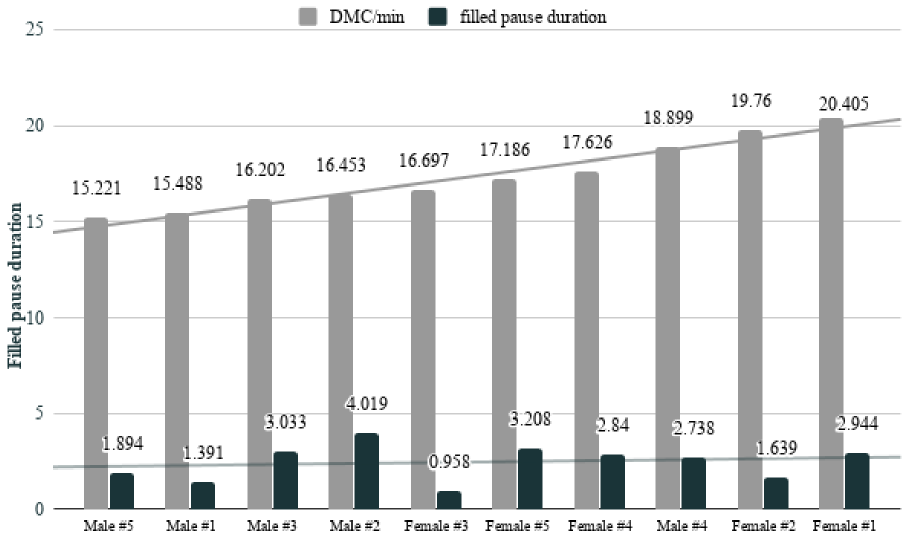

Figure 5: DMC frequency (grey) and normalized filled pause duration (black) in the discourse of individual interpreters

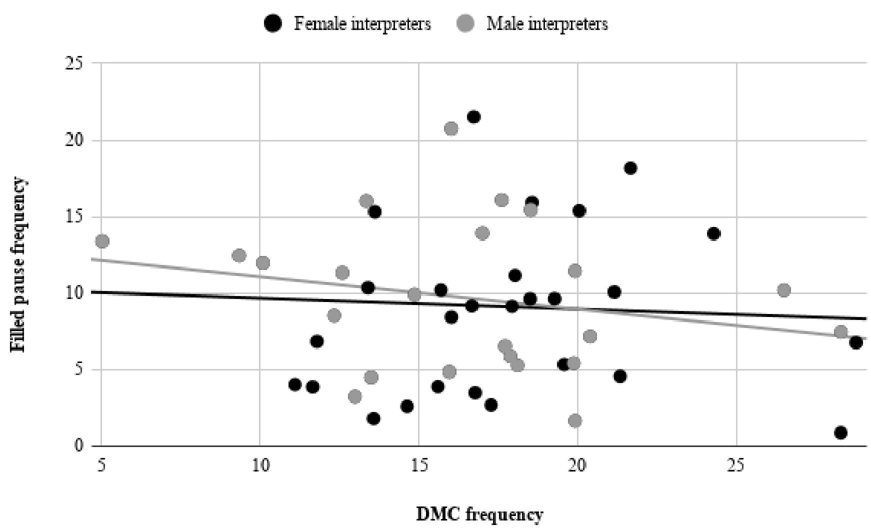

Figure 6: The correlation between DMC frequency and normalized filled pause frequency in each speech interpreted by female (black) and male (grey) interpreters

Table 10: Pearson's correlation of DMC frequency and filled pause duration

	r	**p**
IHC	-0.1	0.6
OHC	0.1	0.5

$p = 0.298$). The contrast between these trends underlines the need for an individual investigation.

Figure 7 presents the correlation between DMC frequency and filled pause frequency in the discourse of the individual interpreters. Pause frequency correlates positively with DMC frequency ($r = 0.214$, $p = 0.553$), more strongly for women ($r = 0.233$, $p = 0.706$) than men ($r = 0.061$, $p = 0.922$). Despite the noticeable variation in filled pause frequency, Grubb's test did not identify significant outliers.

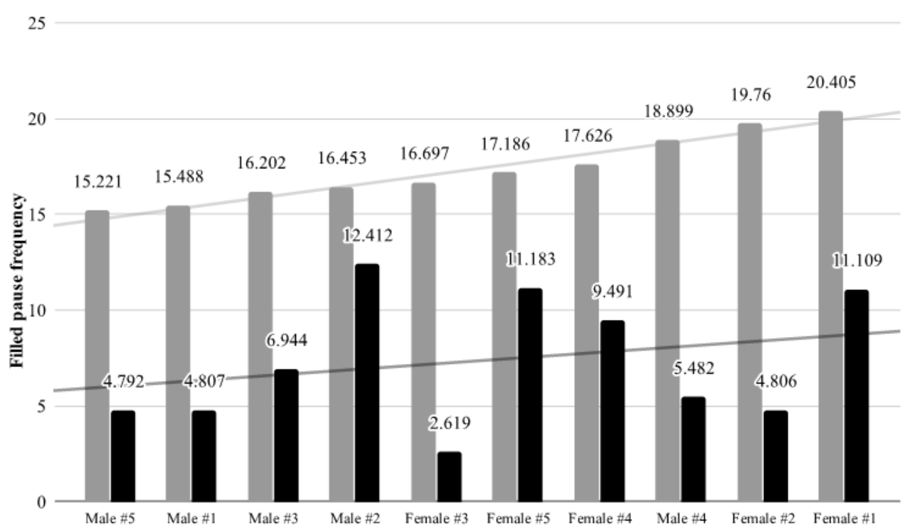

Figure 7: DMC frequency (grey) and normalized filled pause frequency (black) in the discourse of individual interpreters

Filled pause frequency, as opposed to duration, seems to show greater variation among individuals. The difference between the group-based and individual results therefore could stem from the individual speech styles and varying performance of the particular interpreters. However, since they are unevenly represented in the corpus, these results can only be taken as preliminary, outlining directions for future research.

5. Conclusion

This study investigated the frequency of DMCs in interpreted Hungarian EP speeches, probing the relationship between DMC frequency, delivery speed, filled pause duration and frequency.

First, it examined whether DMCs are more frequent in interpreted than in original discourse. Since DMC frequency is highest in Hungarian interpreted discourse compared to both English source speeches and Hungarian original speeches, hypotheses 1 and 2 are both confirmed.

Despite these positive findings, a number of important caveats must be pointed out. It should be stressed that language mediation could have a range of effects. Interpreters can add items but can also omit shorter and longer sections from the original speeches. It is therefore not obvious whether higher DMC frequencies indicate a general tendency in interpreted discourse to use more DMCs, or a higher frequency is simply the result of interpreters omitting sections, while maintaining cohesion. This would mean that interpreters omit sections, conveying only essential information, but without compromising discourse cohesion between the omitted and the interpreted sections, thus transferring and inserting DMCs into abridged texts, creating shorter but equally or more cohesive texts. This could explain why interpreted texts are shorter but contain more DMCs.

Then this paper investigated if interpreted discourse contains more filled pauses. Both in terms of duration and frequency, filled pauses have been found to be more prevalent in interpreted than in original Hungarian discourse. As filled pauses are both over seven times as long and as frequent in interpreted than in original Hungarian discourse, this hypothesis is confirmed. In accordance with other studies, male interpreters hesitate more than women (cf. Collard & Defrancq, 2017).

This paper also sought to establish whether certain discourse properties form interdependent relationships, namely DMC frequency with delivery speed, and DMC frequency with filled pauses. This study has found a positive moderate, statistically significant correlation between delivery speed and DMC frequency in interpreted Hungarian discourse, lending support to hypothesis 4. However, the fact that a very similar level of correlation is revealed in original Hungarian, as well as English speeches, indicates that this relationship is not exclusive to interpreted discourse.

By contrast, the relationship between filled pauses and DMC frequency is not nearly as straightforward. It differs between female and male interpreters, although overall both filled pause duration and frequency correlate weekly and negatively in the IHC. Despite these findings supporting hypothesis 5, weak correlation and considerable individual variation in the data caution against confirming this relationship.

Finally, this paper also tested potential gender differences among interpreters. The only statistically significant variation is found in filled pause duration: male interpreters produce longer filled pauses. In this study, female interpreters used DMCs slightly more frequently, while the differences on the other measures were negligible. Correlation tests exposed more divergence between the sexes. While DMC frequency and delivery speeds correlated significantly in both groups, filled pause duration and frequency correlated positively with DMC frequency for women, and negatively for men. As these trends are weak and not significant, they are most likely caused by individual variation. On the basis of these results, hypothesis 6 is rejected.

As a final note, the role of individual variation deserves more intense research attention. Due to the size of the corpora used here, no broad generalisations can be reached. However, the results of this study do underline the need to account for individual differences when examining interpreted discourse.

Acknowledgements

Supported by the ÚNKP-17-3 New National Excellence Program of the Ministry of Human Capacities.

References

Bendazzoli, C. (2018). Corpus-Based Interpreting Studies: Past, Present and Future Developments of a (wired) Cottage Industry. In M. Russo, C. Bendazzoli, & C. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (pp. 1–19). Singapore: Springer.

Bendazzoli, C., Russo, M., & Defrancq, B. (2018). Corpus-based interpreting studies : a booming research field. *InTRAlinea Online Translation Journal*, (pp. 1:1–1:2).

Bernardini, S., Ferraresi, A., & Miličević, M. (2016). From EPIC to EPTIC — Exploring simplification in interpreting and translation from an intermodal perspective. *Target*, *28*, 61–86. doi:10.1075/target.28.1.03ber.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press.

Blakemore, D., & Gallai, F. (2014). Discourse markers in free indirect style and interpreting. *Journal of Pragmatics*, *60*, 106–120.

Collard, C., & Defrancq, B. (2017). *Sex Differences in Simultaneous Interpreting: A Corpus-Based Study*. [Poster]. Conférence Internationale permanente d'Instituts Universitaires de Traducteurs et Interprètes (CIUTI)'s Forum, Geneva. http://hdl.handle.net/1854/LU-8518872.

Collard, C., & Defrancq, B. (2019). Predictors of ear-voice span, a corpus-based study with special reference to sex. *Perspectives*, *27*, 431–454. doi:10.1080/0907676X.2018.1553199.

Collard, C., & Defrancq, B. (2020). Disfluencies in simultaneous interpreting, a corpus-based study with special reference to sex. In L. Vandevoorde, J. Daems, & B. Defrancq (Eds.), *New Empirical Perspectives on Translation and Interpreting* (p. 264–300). Routledge. URL: https://doi.org/10.4324/9780429030376-12. doi:10.4324/9780429030376-12.

Crible, L. (2017). Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics*, *22*, 242–269. doi:10.1075/ijcl.22.2.04cri.

Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. John Benjamins Publishing Company.

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency. *Languages in Contrast*, *17*, 69–95. doi:10.1075/lic.17.1.04cri.

Defrancq, B. (2016). Well, interpreters... a corpus-based study of a pragmatic particle used by simultaneous interpreters. In G. Pastor, & M. Seghiri (Eds.), *Corpus-based Approaches to Translation and Interpreting* (p. 105–128). Peter Lang D. URL: https://doi.org/10.3726/b10354. doi:10.3726/b10354.

Defrancq, B. (2018). The European Parliament as a discourse community: Its role in comparable analyses of data drawn from parallel interpreting corpora. *The Interpreters' Newsletter*, *23*, 115–132. URL: https://doi.org/10.13137/2421-714x/22401. doi:10.13137/2421-714x/22401.

Defrancq, B., & Plevoets, K. (2018). Over-uh-Load, Filled Pauses in compounds as a Signal of cognitive Load. In M. Russo, C. Bendazzoli, & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (pp. 43–64). Springer Singapore. doi:10.1007/978-981-10-6199-8_7.

Defrancq, B., Plevoets, K., & Magnifico, C. (2015). Connective Items in Interpreting and Translation: Where Do They Come From? In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015* (p. 195–222).

Springer International Publishing volume 3. doi:10.1007/978-3-319-17948-3_9.

Dér, Cs. I., & Markó, A. (2010). A Pilot Study of Hungarian Discourse Markers. *Language and Speech*, *53*, 135–180. doi:10.1177/0023830909357162.

Götz, A. (2017). Az első magyar intermodális korpusz bemutatása. Kutatási lehetőségek. *Argumentum*, *13*, 126–139.

Götz, A. (2018). Sex differences in interpreted hungarian discourse [poster]. In *2nd International Conference on Sociolinguistics (ICS.2), Budapest*. URL: http://rgdoi.net/10.13140/RG.2.2.28928.48645.

Götz, A. (2019a). A corpus-based, individualised investigation of Hungarian of interpreted discourse.

Götz, A. (2019b). Diskurzusjelölők és kötőelemek gyakorisága írott és beszélt, mediált és nem mediált diskurzusokban. In K. Laczkó, & S. Tátrai (Eds.), *Kontextualizáció és metapragmatikai tudatosság* (p. 291–313). ELTE Eötvös József Collegium.

Hale, S. (1999). Interpreters' treatment of discourse markers in courtroom questions. *Forensic Linguistics*, *6*, 57–82. doi:10.1558/sll.1999.6.1.57.

Hale, S. B. (2010). *The discourse of court interpreting: Discourse practices of the law, the witness, and the interpreter*. John Benjamins Pub. (Pbk. ed., with corrections).

Horváth, V. (2014). *Hezitációs jelenségek a magyar beszédben*. ELTE Eötvös Kiadó.

Magnifico, C., & Defrancq, B. (2016). Impoliteness in interpreting: A question of gender? *Translation and Interpreting*, *8*, 26–45. doi:10.12807/ti.108202.2016.a0.

Magnifico, C., & Defrancq, B. (2017). Hedges in conference interpreting: The role of gender. *Interpreting*, *19*, 21–46. doi:10.1075/intp.19.1.02mag.

Magnifico, C., & Defrancq, B. (2020). Norms and gender in simultaneous interpreting: A study of connective markers. *Translation & Interpreting*, *12*, 1–17. doi:10.12807/ti.112201.2020.a01.

Markó, A., & Dér, Cs. I. (2011). Diskurzusjelölők használatának életkori sajátosságai. In J. Navracsics, & Zs. Lengyel (Eds.), *Lexikai folyamatok egy- és kétnyelvű közegben. Pszicholingvisztikai tanulmányok II* (p. 49–61). Tinta Könyvkiadó. URL: http://real.mtak.hu/26155/.

Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies*, *11*, 202–224. URL: https://doi.org/10.1075/tis.11.2.04ple. doi:10.1075/tis.11.2.04ple.

Plevoets, K., & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh(m). *Interpreting*, *20*, 1–28. URL: https://doi.org/10.1075/intp.00001.ple. doi:10.1075/intp.00001.ple.

Russo, M. (2018). Speaking patterns and gender in the european parliament interpreting corpus: A quantitative study as a premise for qualitative investigations. In M. Russo, C. Bendazzoli, & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (p. 115–131). Springer Singapore. doi:10.1007/978-981-10-6199-8_7.

Schirm, A. (2019). A diskurzusjelölő-használat életkori sajátosságai a nyelvi interjú szövegtípusában. *Beszédkutatás*, . URL: https://doi.org/10.15775/beszkut.2019.187-205. doi:10.15775/beszkut.2019.187-205.

Vukov Raffai, E. (2016). A diskurzusjelölő-választások életkori sajátosságai az így, ilyen, hát, mondjuk, ugye esetében. *Magyar Nyelvőr*, *140*, 483–497.

Appendix 1

Hungarian DMCs: *ahogy* 'as, like', *akkor* 'then', *ám* 'although', *azért* 'because of that', *azonban* 'however', *aztán* 'then', *bár* 'although', *csak* 'just, only', *de* 'but', *-e* 'if, whether', *egyébként* 'by the way', *éppen* 'just', *és* 'and', *ezért* 'because of this', *ha* 'if', *hanem* 'but', *hát* 'well', *hiszen* 'since', *hogy* 'that conj.', *hogyha* 'if', *így* 'so', *illetve* 'and', *is* 'too, also', *itt* 'here', *már* 'already', *mármint* 'meaning', *még* 'yet', *mégis* 'still, yet, nevertheless', *mert* 'because', *most* 'now', *nemtom* 'dunno', *noha* 'although, while', *nos* 'well', *pedig* 'yet', *például* 'for example', *s* 'and', *sőt* 'what is more', *talán* 'maybe', *tehát* 'because', *tényleg* 'really', *tudniillik* 'namely', *úgy* 'so', *ugyan* 'although', *ugyanakkor* 'at the same time, nonetheless', *ugyanis* 'since', *ugye* 'is it not?, right?', *úgyhogy* 'so', *vagy* 'or', *vagyis* 'namely', *vajon* 'I wonder', *valamint* 'as well', *viszont* 'but'

English DMCs: *actually, after, albeit, already, also, although, and, anyway, because, before, but, considering, either, even, finally, here, however, if, indeed, instead, maybe, meanwhile, nevertheless, now, oh, okay, once, only, or, otherwise, secondly, since, then, therefore, though, till, too, unless, until, well, when, whenever, where, whereas, while, whilst, yeah, yet*

Appendix 2

Interpreted Hungarian Corpus

Speaker and text	Speech time	Word count	DMCs	Filled p. (sec)	Filled p. no.
F (1)	119.08	233	32	6.6	20
F (1)	76.43	125	16	3.22	13
F (1)	69.14	130	13	3.17	16
F (2)	35.44	68	12	1.22	4
F (2)	76.91	124	15	1.85	5
F (3)	59.5	114	11	1.49	4
F (3)	66.3	116	9	0.77	2
F (3)	67.84	139	22	0.34	1
F (3)	111.12	217	22	1.17	5
F (3)	61.78	117	9	2.09	4
F (4)	170.61	285	38	7.98	26
F (4)	98.52	145	17	5.28	17
F (4)	71.13	115	11	3.01	10
F (4)	188.21	305	48	18.16	57
F (4)	85.84	157	19	0.96	5
F (4)	114.83	189	19	1.19	5
F (4)	67.43	108	17	1.14	6
F (5)	105.82	186	23	3.9	17
F (5)	70.52	107	11	4.02	18
F (5)	162.1	258	39	10.19	26
F (5)	89.73	157	22	6.52	23
F (5)	75.3	123	18	10	27
F (5)	78.72	151	23	1.05	6
F (5)	69.87	126	17	2.38	13
F (5)	157.75	242	30	4.41	18
F (5)	67.85	107	20	5.69	18
F (5)	71.97	96	11	2.61	11
Subtotal (f)	2489.74	4240	544	110.41	377

Speaker and text	**Speech time**	**Word count**	**DMCs**	**Filled p. (sec)**	**Filled p. no.**
M (1)	86.46	165	13	2.05	7
M (1)	73.9	137	12	1.27	4
M (1)	64.33	125	17	1.89	7
M (2)	84.76	118	12	4.41	14
M (2)	57.83	87	7	3.9	12
M (2)	152.61	338	46	6.3	19
M (2)	63.65	98	12	6.04	22
M (2)	78.35	136	18	7.48	21
M (2)	67.43	117	7	6.21	18
M (2)	58.86	118	18	2.32	10
M (2)	77.72	135	14	5.55	20
M (2)	95.31	146	18	5.99	18
M (2)	35.84	58	2	4.44	8
M (2)	77.62	137	13	6.52	18
M (2)	255.7	378	26	16.05	51
M (2)	141.52	217	29	8.32	27
M (3)	66.39	117	20	3.49	6
M (3)	63.23	99	10	3.06	9
M (4)	72.26	145	19	1.12	2
M (4)	234.81	390	47	11.64	23
M (4)	108.82	205	26	6.22	13
M (5)	133.36	221	25	3.66	10
M (5)	79.5	137	21	3.06	7
Subtotal (m)	2230.26	3824	432	120.99	346