Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

When "uhm", "and" and "yeah" sound the same—prosodic aspects of discourse pragmatic markers in American English

Marlene Böttcher[1], *Margaret Zellers*[1]

[1]Kiel University, Germany

mboettcher@isfas.uni-kiel.de, mzellers@isfas.uni-kiel.de

Abstract

This study explores the distributional and prosodic similarities and differences of different particles in discourse. Both lexicalized discourse particles (e.g. *yeah, so, well*) and filler particles (e.g. *uh, uhm*) share structuring discourse functions in English. This study looks at different types of particles and investigates their occurrence at the boundary between broader units of discourse, where they are likely to be used in a structuring function to indicate so called frame shifts. The data analysis focuses on elements occurring at discourse boundaries in 40 formal and informal English narrations by mono- and bilingual speakers from the RUEG corpus. Since prosody is also an important means in discourse organization, the analysis also includes prosodic aspects of the discourse particles. A variety of elements and their combinations was found at discourse boundaries, including filler particles, discourse markers, tongue clicks and connectors. While filler particles, alone or in combination, are more frequent in formal narrations, discourse markers and connectors are more frequently found in informal narrations. Prosodically, the boundary elements were produced similarly in pitch and duration, but different types also showed finer phonetic differences. Both the choice of particle and their prosodic realization are influenced by the formality of the situation.

Index Terms: discourse particles, filler particles, discourse structure, spontaneous speech, prosody, English, corpus data

1. Introduction

Both lexicalized discourse particles like *yeah, so, well* and filler particles like *uh, uhm* are frequently produced in speech and are typical for unplanned, spontaneous discourse [1]. They have been described as indicators of (dis)fluency [2, 3]. Yet, they also fulfill discourse-pragmatic functions beyond hesitations [4, 5] and share a functional overlap in turn-holding and discourse structuring [6, 7]. Both discourse markers and filler particles occur at (major) discourse boundaries, indicating a shift on the discourse level [8, 9]. This exploratory study takes a functional perspective and focuses on discourse elements occurring at discourse boundaries, including discourse markers and filler particles, and additionally analyses their prosodic form.

Discourse markers (DM) are a sub-set of pragmatic markers [10]. DMs are often homophonous with adverbs or conjunctions, yet characterized by semantic bleaching and a high pragmatic differentiation [11]. DMs are poly-functional and can, among other functions, be used to indicate hesitation, correction or repair, initializing a narrative or a thematic switch. DMs are not obligatory and do not contribute to the proposition and/or truth value of an utterance but contribute on discourse level. Syntactically DMs are not integrated and prosodically they are frequently phrased separately and set apart from the following proposition by a prosodic break [11]. Additionally, DMs are phonetically reduced and unstressed compared to their homophonous counterparts [12, 13]. Discourse connectors (CON) are a subset of DMs predominantly contributing to textual cohesion and with a stronger functional link to their homophonous counterparts. They are used to introduce additions (e.g. *and*) or contrasts (e.g. *but*) on discourse level rather than syntactic phrases [14].

Filler particles (FPs) have for a long time been disregarded in linguistics as errors or disfluencies [15]. More recent research has shown them to contribute to fluent conversations [1, 8] and to be more than mere hesitations. FPs are lexically underspecified, syntactically unconstrained and semantically empty [16]. The occurrence of FPs is related to discourse structure as they frequently precede new paragraphs or new topics [5, 17] and they are highly multifunctional, similar to DMs [16, 18, 19]. FPs are constituted by a central(ized) vowel followed by an optional nasal [16, 17, 19]. In English they are commonly transcribed as *uh/er* and *uhm/erm*. Prosodically, FPs are produced with lower fundamental frequency compared to the immediate context [16] and with weak pitch movement [20], and they frequently occur inter- or postpausally [16, 17].

Next to these discourse structuring vocalizations, non-phonemic tongue clicks can also be used in discourse organisation. Tongue clicks (TCs) are ingressive stop sounds and while phonemic in some languages (e.g. Khoisan languages) they are used with pragmatic functions in languages like English, German and Swedish [21, 22, 23]. Similar to lexicalized particles (DMs, CONs) and non-lexicalized particles (FPs) they are polyfunctional and can indicate a new topic, word search, turn uptake, back channeling, stance marking or initiate repairs [23].

While DMs and CONs are assumed to be lexicalized and therefore part of language grammar, the status of filler particles and tongue clicks is less clear [24]. Yet, the different types of particles can all be used to fulfill discourse-pragmatic functions beyond hesitations [17, 5]. Due to their functional overlap and the occurrence at discourse boundaries, this study analyses CONs, DMs, FPs and TCs together.

Spontaneous discourse can be segmented into stretches of speech with the same topic or discourse function. Between these chunks of speech discourse boundaries can be identified [25]. An important distinction in narratives is the one between openings and closings [26]. These can be defined as speech that, respectively, precedes and follows the core of the narration. Example 1 [1] shows an example from the data analysed in this study. The opening consists of a greeting, addressing the person and an initial contextualisation of the scene. The closing evaluates the situation and relates the discourse back to the interlocutor (for more details see 2). Between the opening and the

[1]The example is produced by a Greek bilingual speaker living in the U.S. as part of the RUEG corpus. (It can be accessed as file "USbi02FG_isE" via the link to the corpus: https://korpling.german.hu-berlin.de/annis/#_c=UlVFRy1FTl8xLjAtU05BUFNIT1Q.

10.21437/SpeechProsody.2024-183

main narration the speaker produces a silent pause (indicated by (-)) and the FP *um* and at the boundary between the main narration and the closing the speaker produces a silent pause and the DM *so*. These boundaries are the discourse boundaries we investigate in the current study.

1. **opening**
   [tongueclicking] hey costa uh as i was (-) walking i saw this car accident
   main narration
   (-) um there was a ball that rolled out into the street (-) and the dog running after it (-) and there was two cars driving and the first car stopped and the second car (-) ended up (-) rearending (-) the other guy but everyone=s (-) okay no one=s hurt
   closing
   (-) so just make sure you=re careful always when you=re driving (-)

Following [27] this study looks at discourse boundaries within narrations and the particles that are produced in this context as well as their prosodic form. Since prosody, too, is an important aspect of discourse organization [28] and FPs and DMs share discourse functions, this study also analyses the prosodic shape of particles at discourse boundaries. Additionally, the association of the particles under discussion with more unplanned and informal speech prompts the investigation of the distribution and form across different speech registers. In more detail the research questions are: i. Which elements occur at major discourse boundaries in spontaneous English narrations? ii. How are they produced prosodically in terms of duration, fundamental frequency (F0), F0 difference to the context and their pitch contour? And iii. is there a difference in the distribution or prosodic form related to speech register?

2. Method

For this study spoken English data of mono- and bilingually raised speakers from the RUEG corpus [29] was used. The corpus contains spontaneous written and spoken elicitations of a car accident prompted by a video narrated by participants in two scenarios: a formal situation addressing the police and an informal scenario addressing a friend (order of elicitation was balanced within the corpus). The corpus is currently being annotated for the discourse structure considering openings (OPEN), main narration (MAIN) of the event and closing (CLOSE). Utterances that directly depict the events are considered as part of the main narration as illustrated in example (1) and can be identified easily due to the elicitation set-up with a video prompt. A subset of the data for which the discourse structure annotations were complete was made available for use in this study.

Table 1: *Overview of categories and items at discourse boundaries found in the data. Items in the respective categories are ordered with descending frequency*

Category	Items	n
SP	(-)	79
DM	so, yeah, basically, like, okay, yo	40
FP	um, uh	28
CON	and, but	19
PHRASE	you know, and then, I mean	3
TC	(tongue click)	3

The analysis in this study considers English data of speakers from the U.S. with different or no heritage language background (Greek, Turkish, German, Russian 4 speakers each and 4 monolinguals). To reduce further socio-demographic variation, data from 20 female adolescents were chosen and analysed. The elements that occurred at discourse boundaries were categorized into one of the following: silent pause (SP), DM, FP, CON, TC or phrase (PHRASE; e.g. *I mean*). Table 1 presents an exhaustive overview of these items. The prosodic analysis looks at the vocalized elements at boundaries, i.e., CONs, DMs and FPs and their combinations up to two elements (this excludes few instances of combinations of three elements and PHRASEs). Elements were analysed in Praat [30] with a focus on the prosodic form of elements: The duration measured for each element was normalized by the speech rate calculated for each narration to compare across different speakers. The mean fundamental frequency (F0) was extracted in semitones (re: 1 Hz) for each element as well as the speaker's overall mean F0. $F0_{norm}$ represents the normalized F0 value calculated as the measured value minus speaker mean. Following [31] $F0_{norm}$ values above 0.6 ST are considered above a just noticeable difference (JND) for English listeners. Following [32], the contour was extracted within the mid 80 % of the element. Excursion sizes larger than 1 semitone (ST) were classified as falling or rising depending on the order of their minimum and maximum F0. Given the small data set the results will be descriptive and no inferential statistics are applied.

3. Results

Not all narrations contained an explicit opening and/or closing, and discourse particles did not arise at every annotated discourse boundary. This analysis, thus, focuses on 95 annotated discourse boundaries which included at least one of the elements presented in Table 1 (preceding the opening: 35; between opening and main narration: 37; between main narration and closing: 23).

3.1. Particles at discourse boundaries

Preceding (_OPEN) and following OPEN (OPEN_MAIN) the discourse boundaries are marked by elements in the majority of narrations and with a similar distribution (_OPEN formal: n = 19 , 95%; informal: n = 16, 80%; OPEN_MAIN formal: n = 19, 95%; informal: n = 18, 90%). There is a tendency for fewer boundary markings at the end of a narration in formal compared to informal situations (MAIN_CLOSE formal: n = 9, 45%; informal: n = 14, 70%). While DMs are equally frequent at all types of boundaries (_OPEN: n = 10 , 25%; OPEN_MAIN formal: n = 15, 38%; MAIN_CLOSE: n = 15, 38%), CONs are more frequently introducing MAIN (OPEN_MAIN: n = 12, 60%) and FPs are more frequently used at the very beginning of a narration (_OPEN: n = 13, 46%; see also Table 2).

In formal narrations the most frequent element was a FP, either single (n = 15) or in combination (n = 5). In informal narrations the most frequent elements were CONs and DMs, either single (CON: n = 11, DM: n = 8) or in combination (CON: n = 4, DM: n = 9). Figure 1 presents an overview of the different elements and their combinations across formalities.

The most frequent boundary marking was a single element optionally accompanied by a SP (n = 48) while many boundaries were marked by a SP alone (n = 25). Combinations of elements at discourse boundaries included either two (n = 16) or three elements (n = 4) as well as PHRASEs (n = 3).

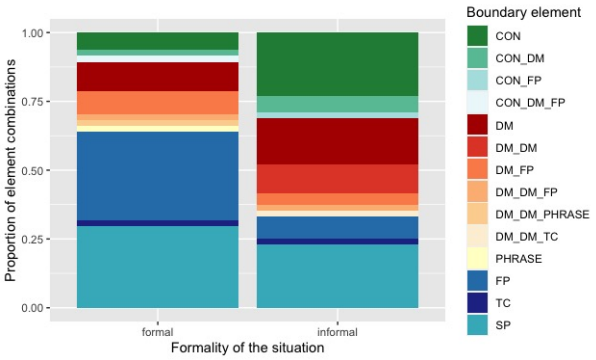

Figure 1: *Overview of discourse particles and their combinations across the two formalities; green shades = CONs and their combinations, red shades = DMs and their combinations, yellow = PHRASES, blue shades = the non lexical elements FPs,TCs and SPs. The combinations are presented irrespective of the order of elements in the narration (i.e., DM_FP also contains FP_DM.*

Of the single elements marking a discourse boundary the most frequent ones are FPs (n = 19, 40% of single elements) followed by CONs (n = 14, 29%) and DMs (n = 13, 27%) and two TCs. The most frequent combination of two elements were DMs in combination with FPs (n = 6, 37% of pairs) followed by two DMs (n = 5, 31%) and CONs in combination with DMs (n = 4, 25%). There was also one instance of a combination of CON and FP. While DMs, CONs and FPs occured in the first position, the second position of pairs was predominantly filled by DMs (see Table 5 for further details). There were few combinations of three elements and only one of them occurred twice: two DMs and a FP. The other combinations were CON, DM and FP, and two DMs and a TC.

3.2. Prosody at discourse boundaries

Overall the particles' mean normalized F0 is higher in the beginning compared to later parts of the narration (_OPEN: $\overline{x}$ = 1.5 ST; OPEN_MAIN: $\overline{x}$ = 0.66 ST; MAIN_CLOSE:$\overline{x}$ = 0.08 ST). The F0 in DMs is similar across different boundaries, while the difference in CONs and DMs is above a JND. Table 2 shows

Table 2: *Prosodic aspects of particles at different discourse boundaries: normalized mean duration related to the speech rate (% syll), the mean F0 difference to the speakers' mean ($F0_{norm}$) in ST, and the percentages of falling ($fall_{\%}$), rising ($rise_{\%}$) and level contours ($level_{\%}$).*

_OPEN	**n**	**syll$_{\%}$**	**F0$_{norm}$**	**fall$_{\%}$**	**rise$_{\%}$**	**level$_{\%}$**
CON	0	-	-	-	-	-
DM	10	1.07	2.04	60	30	10
FP	13	0.88	1.12	54	31	15
OPEN_MAIN						
CON	12	1.04	-0.62	58	25	17
DM	15	0.96	1.95	53	-	47
FP	12	0.94	0.48	67	8	25
MAIN_CLOSE						
CON	8	0.84	-4.79	62	12	25
DM	15	0.77	2.66	40	33	27
FP	3	0.65	-1.72	67	-	33

Table 3: *Prosodic aspects of discourse particles across different formalities: normalized mean duration related to the speech rate (% syll), the mean F0 difference to the speakers' mean ($F0_{norm}$) in ST, and the percentages of falling ($fall_{\%}$), rising ($rise_{\%}$) and level contours ($level_{\%}$).*

formal	**n**	**syll$_{\%}$**	**F0$_{norm}$**	**fall$_{\%}$**	**rise$_{\%}$**	**level$_{\%}$**
CON	4	0.98	-2.09	100	-	-
DM	8	0.94	2.86	38	25	38
FP	11	0.82	0.21	55	18	27
informal						
CON	13	0.95	-1.79	54	31	15
DM	10	0.90	2.87	50	40	10
FP	5	1.02	0.23	80	20	-

a more detailed overview of prosodic aspects across the different boundaries. Elements at all discourse boundaries predominantly occur with falling contour. However, rising contours are produced frequently with FPs and DMs in the beginning of narrations or with CONs beginning main narrations and with DMs initiating a closing. The level contour occurred more frequently towards the end of narrations.

In informal narrations the elements' mean normalized F0 is lower compared to that in formal narrations (formal: $\overline{x}$ = 0.7 ST; informal: $\overline{x}$ = 0.2 ST). While the F0 in FPs is comparable across formalities, the difference in CONs and DMs is above a JND. Including the overall speakers' mean F0 a tentative difference with higher F0 in informal compared to formal narrations can be observed (formal: $\overline{x}$ = 91 ST; informal: $\overline{x}$ = 92 ST). While FPs do not show differences in F0 they are shorter in formal compared to informal narrations (formal 0.82 and 1.02 related to the mean syllable duration). The predominant contour was falling in both formalities (formal: n = 13, 57%; informal n = 16, 57%). Rising contours are produced more frequently in informal compared to formal situations (formal: n = 4, 17%; informal n = 9, 32%). Table 3 shows a more detailed overview of prosodic aspects across the two formalities.

The 45 elements occurring at boundaries on their own are comparable in prosodic shape with a mean duration of 370 ms (i.e. 0.99 normalized by speech rate), a mean normalized F0 just about a JND ($\overline{x}$ = 0.6 ST), and predominantly falling F0 contour (n= 31, 69 %). The lexicalized elements CONs and DMs show longer durations than FPs and CONs show more variation regarding the F0 contour compared to DMs and FPs. Table 4 presents prosodic aspects of CONs, DMs and FPs as single elements at discourse boundaries.

The 15 elements occurring in pairs at discourse boundaries were analysed split into the first and the second part. The first element was produced with a mean duration of 332 ms (0.87 relative to the mean syllable duration) and a mean normalized

Table 4: *Prosodic aspects of single discourse particles: normalized mean duration related to the speech rate (% syll), the mean F0 difference to the speakers' mean ($F0_{norm}$) in ST, and the percentages of falling ($fall_{\%}$), rising ($rise_{\%}$) and level contours ($level_{\%}$).*

	n	**syll$_{\%}$**	**F0$_{norm}$**	**fall$_{\%}$**	**rise$_{\%}$**	**level$_{\%}$**
CON	14	0.99	-0.84	50	21	29
DM	12	1.20	1.98	83	8	8
FP	19	0.86	0.64	74	16	11

Table 5: *Prosodic aspects of discourse particles occurring in pairs: normalized mean duration related to the speech rate (% syll), the mean F0 difference to the speakers' mean ($F0_{norm}$) in ST, and the percentages of falling ($fall_{\%}$), rising ($rise_{\%}$) and level contours ($level_{\%}$).*

1st part	**n**	**syll$_{\%}$**	**F0$_{norm}$**	**fall$_{\%}$**	**rise$_{\%}$**	**level$_{\%}$**
CON	5	0.95	-3.83	80	20	-
DM	7	0.80	3.43	43	29	29
FP	3	0.91	-0.78	33	-	67
2nd part						
DM	12	0.73	-0.03	25	25	50
FP	3	1.03	1.81	33	67	-

F0 below a JND ($\overline{x}$ = -0.1 ST). Table 5 presents prosodic aspects of CONs, DMs and FPs as first and second part of combinations at discourse boundaries. CONs are generally longer and both CONs and FPs are realized below the speakers' mean compared to DMs. All three types are predominantly produced with a falling contour, yet, DMs and FPs show some variation. Next to overall falling contours FPs are also produced with level contour (n = 2) while DMs also with rising contour (n = 2). The respective second element was generally produced with a mean duration of 303 ms (0.79 related to the mean syllable duration) and a mean normalized F0 below a JND ($\overline{x}$ = 0.4 ST). They are produced with varying contours. The DMs were realized with shorter durations and also with varying contours. Rises were also produced in this context on FPs.

4. Discussion

The results of this exploratory study show that lexicalized DMs and CONs, as well as FPs, are used to mark boundaries in discourse structure in English.

There seems to be a tendency for different types of elements at different boundaries within the narration. The more frequent use of FPs towards the beginning is in line with their speech planning function [17]. CONs on the other hand are more frequent towards the middle and end of the narration connecting the individual paragraphs [14]. This difference of boundary type is also tentatively present in our prosodic analysis. The predominantly falling F0 contours with all particle types at all boundary types is in line with observations from discourse particles initiating a new paragraph or turn in other languages [33]. Further analyses of more speakers will provide further insight.

Across the different formalities there is less explicit marking of discourse boundaries towards the end of narrations in formal register. This could be explained with habitual ease; that is, speakers might be more familiar with formulaic language, including closing of narrative sequences, in informal narrations due to everyday use of voice messaging or telephone calls to friends or family. Voice messages directed to the police might not be as frequent as to a friend.

Additionally, the type of element is also influenced by register. FPs are more frequently used to mark discourse boundaries in formal settings, while the lexicalized elements are more frequently used in informal narrations. We provide two possible explanations here: The higher frequency of FPs has previously been associated with higher cognitive load [34, 35]. The cognitive load can be assumed to be higher in the formal narrations in our data due to the lower habitual ease when narrating a car accident to the police while the pressure on language form is also higher in more formal situations. Speech planning might therefore require more cognitive resources. The shared position at discourse boundaries, however, favours an alternative explanation focusing on the functional overlap of FPs, DMs, and CONs. The choice of element at discourse boundaries in different formalities might be related to speech register. DMs are associated with informal speech and are generally avoided in more formal, prepared speech [36]. Given that FPs are phonetically not salient and frequently missed by listeners [37], they might be better fitted than DMs for a formal register, since they could be less disruptive to the overall flow of speech. The prosodic analysis provides further evidence for their functional overlap. Overall, the prosodic shape of FPs, DMs, and CONs is comparable in the analysed data. The reduced F0 values in most of the cases can be interpreted as prosodic non-integration of the particles. In that way, prosody further helps to structure the discourse and marks the beginning of a new topic followed by a reset in F0 [28].

The variance found related to contour in combinations of elements could be related to the preceding and following higher F0 in cases of partial prosodic integration: a falling contour from the preceding context in the first element and the rising contour into the following context within the second element. The variance could also be linked to different pragmatic functions of the two elements e.g. *new topic* and *turn holding*. Future work will need to address these depending on their prosodic phrasing. There also seem to be more rises in the informal narrations along with an overall higher mean F0 in this setting. This is tentative evidence for a prosodic register difference, with possibly more prosodic variation in informal register, compared to prosody in formal contexts. Further analyses are necessary to confirm this.

This study included mono- and bilingually raised speakers with different language backgrounds and focused on the distribution, choice, prosodic form of different discourse particles in the speakers' majority language. Prior work has shown an influence on discourse particle use from the first to the second language [33]. Further analyses will address the possible differences across different speaker groups and different language backgrounds in both of these bilingual speakers' languages.

5. Conclusions and Outlook

This study investigated the distribution and prosodic form of different elements occurring at discourse boundaries and adds to the research on discourse pragmatics across different speech registers. The focus on similar contexts where a common discourse structuring function can be assumed allowed a comparison of different discourse particles and found prosodic similarities but distributional differences across register for the lexicalized and non-lexicalized particles. Further research needs to address the inter-speaker variability and the bilingual status of speakers in this data, addressing the question of cross-linguistic similarity and bilingual variation.

6. Acknowledgements

Thank you to Kalliopi Katsika for making the unpublished discourse annotations of the RUEG corpus available for our analysis and, together with Annika Labrenz and Kateryna Iefremenko, for the collaborations leading up to this analysis. Thank you also to the two anonymous reviewers for helpful feedback and suggestions for further research.

7. References

[1] L. Crible, L. Degand, and G. Gilquin, "The clustering of discourse markers and filled pauses: a corpus-based French-English study of (dis)fluency," *Languages in Contrast*, vol. 17, no. 1, pp. 69–95, feb 2017.

[2] M. S. Schmid and K. B. Fägersten, "Disfluency markers in L1 attrition," *Language learning*, vol. 60, no. 4, pp. 753–791, 2010.

[3] G. Gilquin, "Hesitation markers among EFL learners: Pragmatic deficiency or difference?" *Pragmatics and corpus linguistics: A mutualistic entente*, no. 2, pp. 119–149, 2008.

[4] B. Heine, G. Kaltenböck, T. Kuteva, and H. Long, "On the rise of discourse markers," *Studies at the Grammar-Discourse Interface: Discourse Markers and Discourse-Related Grammatical Phenomena. John Benjamins Publishing Company*, pp. 23–56, 2021.

[5] M. Swerts, "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, vol. 30, no. 4, pp. 485–496, 1998.

[6] J. E. Fox Tree, "Interpreting Pauses and *Um*s at Turn Exchanges," *Discourse Processes*, vol. 34, no. 1, pp. 37–55, Jul. 2002.

[7] K. Fischer, "Discourse particles, turn-taking, and the semantics-pragmatics interface," *Revue de s ´emantique et pragmatique*, vol. 8, pp. 111–132, 2000.

[8] B. Knudsen, A. Creemers, and A. S. Meyer, "Forgotten little words: How backchannels and particles may facilitate speech planning in conversation?" *Frontiers in Psychology*, vol. 11, p. 593671, 2020.

[9] E. A. Schegloff, "Some other "*uh*(*m*)" s," *Discourse processes*, vol. 47, no. 2, pp. 130–174, 2010.

[10] B. Fraser, "An account of discourse markers," *International review of Pragmatics*, vol. 1, no. 2, pp. 293–320, 2009.

[11] P. Auer and S. Günthner, *Die Entstehung von Diskursmarkern im Deutschen – ein Fall von Grammatikalisierung?* Berlin, New York: De Gruyter, 2005, pp. 335–362.

[12] L. Lee, D. Jouvet, K. Bartkova, Y. Keromnes, and M. Dargnat, "Correlation between prosody and pragmatics: case study of discourse markers in French and English," in *INTERSPEECH 2020*, 2020.

[13] C.-Y. Tseng, Z.-Y. Su, C.-H. Chang, and C.-H. Tai, "Prosodic fillers and discourse markers–discourse prosody and text prediction," in *Tonal Aspects of Languages*, 2006.

[14] H. Kuswoyo, E. T. S. Sujatna, L. M. Indrayani, and A. Rido, "Cohesive Conjunctions *and* and *so* as Discourse Strategies in English Native and Non-native Engineering Lecturers: A Corpus-Based Study," *International Journal of Advanced Science and Technology*, vol. 29, no. 7, pp. 2322–2335, 2020.

[15] H. Maclay and C. E. Osgood, "Hesitation Phenomena in Spontaneous English Speech," *Word*, vol. 15, no. 1, pp. 19–44, 1959.

[16] M. Belz, *Die Phonetik von äh und ähm*. Springer Berlin Heidelberg, 2021. [Online]. Available: http://link.springer.com/10.1007/978-3-662-62812-6

[17] H. H. Clark and J. E. FoxTree, "Using *uh* and *um* in spontaneous speaking," *Cognition*, vol. 84, no. 1, pp. 73–111, may 2002.

[18] G. Kjellmer, "Hesitation. In Defence of ER and ERM," *English Studies*, vol. 84, no. 2, pp. 170–198, apr 2003.

[19] E. Shriberg, "To 'errrr' is human: ecology and acoustics of speech disfluencies," *Journal of the International Phonetic Association*, vol. 31, no. 1, pp. 153–169, jun 2001.

[20] M. Belz and U. Reichel, "Pitch characteristics of filled pauses," in *The 7th Workshop on Disfluency in Spontaneous Speech*, aug 2015.

[21] R. Ogden, "Clicks and percussives in English conversation," *Journal of the International Phonetic Association*, vol. 43, no. 3, pp. 299–320, 2013.

[22] J. Trouvain, "On clicks in German," *Trends in Phonetics and Phonology: Studies from Germanspeaking Europe*, pp. 21–33, 2015.

[23] M. Zellers, "An overview of discourse clicks in Central Swedish," in *Annual Conference of the International Speech Communication Association (Interspeech), Incheon, Republic of Korea, September*, 2022, pp. 18–22.

[24] K. Aijmer and A.-M. Simon-Vandenbergen, "Pragmatic markers," *Discursive pragmatics*, vol. 8, pp. 223–247, 2011.

[25] E. A. Schegloff, *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press, 2007, vol. 1.

[26] E. A. Schegloff and H. Sacks, "Opening up Closings," *Semiotica*, vol. 8, no. 4, pp. 289–327, 1973.

[27] A. Labrenz, M. Böttcher, F. Groth, K. Iefremenko, and K. Katsika, "Discourse markers and filler particles at the boundary: A corpus study across languages, speaker groups, and communicative situations," in *Discourse Markers - Theories and Methods*, Paris, France, 2023.

[28] C.-Y. Tseng, "Higher Level Organization and Discourse Prosody," in *Tonal Aspects of Languages*, 2006.

[29] H. Wiese, A. Alexiadou, S. Allen, O. Bunk, N. Gagarina, K. Iefremenko, E. Jahns, M. Klotz, T. Krause, A. Labrenz, A. Lüdeling, M. Martynova, K. Neuhaus, T. Pashkova, V. Rizou, T. Rosemarie, C. Schroeder, L. Szucsich, W. Tsehaye, S. Zerbian, and Y. Zuban, "RUEG Corpus," 2021. [Online]. Available: https://doi.org/10.5281/zenodo.3236069

[30] P. Boersma and D. Weenink, "Praat: doing phonetics by computer. version 6.2.03," http://www. praat. org/, 2021.

[31] A. Jongman, Z. Qin, J. Zhang, and J. A. Sereno, "Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners," *The Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. EL163–EL169, 2017.

[32] S. Wehrle, "A multi-dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder," Ph.D. dissertation, University of Cologne, 2021.

[33] S. Sbranna, E. Möking, S. Wehrle, and M. Grice, "Backchannelling across languages: Rate, lexical choice and intonation in l1 italian, l1 german and l2 german," in *Proc. Speech Prosody 2022. Lisbon, Portugal*, vol. 2022, 2022, pp. 734–738.

[34] S. Betz, N. Bryhadyr, O. Türk, and P. Wagner, "Cognitive Load Increases Spoken and Gestural Hesitation Frequency," *Languages*, vol. 8, no. 1, p. 71, 2023.

[35] A. Berthold and A. Jameson, "Interpreting Symptoms of Cognitive Load in Speech Input," in *UM99 user modeling: Proceedings of the seventh international conference*, 1999, pp. 235–244.

[36] L. Crible, "Discourse Markers and (Dis) fluency," *Pragmatics & Beyond New Series*, vol. 286, 2018.

[37] O. Niebuhr and K. Fischer, "Do not hesitate! — Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance," in *Interspeech 2019*. ISCA, sep 2019, pp. 544–548.