PHONETIC DESCRIPTION OF FILLED PAUSES AS DISCOURSE MARKERS IN TOTOLI

Constantijn Kaland, Maria Bardají

Institute of Linguistics, University of Cologne, Germany
ckaland@uni-koeln.de, mbardaj1@uni-koeln.de

ABSTRACT

This study reports a phonetic analysis of filled pauses, similar to 'uh' and 'um' in English, in Totoli (Sulawesi, Indonesia). The focus of this study is on two types; 'a(h)' and 'e(h)'. The analysis was carried out on a corpus of spontaneous Pear Film retellings to investigate their frequency and position of occurrence, temporal characteristics and vowel quality. Results show that filled pauses function as markers of discourse structure, corroborating results for other languages. In particular, the Totoli filled pauses delimit two parts of the discourse that form different, semantically coherent units (i.e. 'paragraphs'). This function mainly applies to 'a'-like, rather than 'e'-like filled pauses. In addition, dialectal variation in the acoustic realisation of the filled pauses was found. The outcomes are interpreted as indications of the linguistic status of filled pauses in Totoli.

Keywords: filled pause, discourse marker, acoustic analysis, vowel quality, duration

1. INTRODUCTION

The endangered language Totoli is spoken in the Northern part of the Indonesian province Central Sulawesi by a decreasing number of speakers (approximately 7500 in [1]). The literature on this language is limited. Basic linguistic information and word lists are provided in [1], and a phonological and syntactic study of intonation units is reported in [2]. Verbal morpho-syntax has been studied to a larger extent ([3], [4], [5], [6], [7], [8], [9], [10], [11], [12]). Discourse structure has received limited attention, except for the recognition that multiple intonation units can form a paragraph. These often come with a specific sequence of intonation patterns, i.e. paragraph non-final intonation units end with a high boundary tone, whereas the final intonation unit in the paragraph ends with a low boundary tone [2].

Specific f0 movements in groups of subsequent intonation units (i.e. 'paratones') have been well-documented for English in read speech and spontaneous speech. They mark "a stretch of discourse presented by a speaker as forming a unit with a single topic" ([13], p.38), i.e. a paragraph. For Dutch, paragraphs showed an f0 downtrend over the course of the entire paragraph ('supradeclination', [14]). A follow-up experiment showed that listeners rate the paragraph as more natural when supradeclination is present than when it is absent. The perceptual relevance of paragraph intonation was furthermore shown for English listeners, who could hear whether a sentence was uttered in isolation or as part of a paragraph [15].

The literature has shown that there are other means to structure discourse at the level of the paragraph, such as pausing. For example, a study on Dutch showed that filled pauses (FPs) such as 'uh' and 'um' "carry information about larger-scale topical units" ([16], p.494). Filled pauses were more often present at strong breaks, which correspond to transitions between paragraphs, than at weak breaks, which correspond to within-paragraph boundaries. In addition, FPs were often preceded and followed by a silent pause at major breaks. If FPs occurred after a major break, they were most often found phrase-initially, i.e. still in the vicinity of the break. FPs were tentatively interpreted as linguistically relevant elements [16], which distinguishes them from 'placeholders' signaling whether the speaker has word-finding problems (e.g. [17]) and non-linguistic phenomena such as coughs and laughs.

Although traditional work studied FPs as disfluencies that are not part of the language (e.g. [18], [19]), more recent work has adopted linguistic interpretations. Thus, English 'uh' and 'um' were shown to be under control of the speaker just as any other word [20]. It was also shown that 'uh' occurs at minor breaks and 'um' at major ones and that their segmental makeup is very similar (central vowels) across languages spoken in Europe ([20], p.92) and also sound similar in Hebrew (e.g. 'eh', 'e-h', 'em', 'e-m', 'ah', 'a-m', e.g. [21]). Japanese FPs are segmentally different (e.g. 'eeto', 'etto', 'ano', 'anoo', 'uun', 'uunto', 'konoo', 'sonoo', 'jaa', e.g. [22]) and so do some Spanish ones ('este', e.g. [23]). It should be noted that

FPs in the latter two languages were derived from demonstratives [20]. Furthermore, studies reported dialectal variation in the production of FPs in English and Mandarin ([20], [24], [25]).

In the current study we explore the use of FPs in Totoli in order to advance our knowledge of their use in lesser documented languages. This is particularly valuable given the segmental similarity of FPs in well-studied European languages. There is no published documentation so far on Totoli FPs. The FPs in the available corpus [26] are mainly 'a(h)', 'e(h)' or 'mm(h)'. Given that the focus of the current study is on the potential discourse structuring function of Totoli FPs, 'mm(h)'-like FPs (backchannels) were excluded. We investigate 'a'- and 'e'-like FPs for their frequency of occurrence and their acoustic characteristics (temporal and vowel quality) in order to investigate the extent to which they contribute to discourse structure.

2. METHODOLOGY

This section describes the speech data, annotation and (acoustic) analyses. It is important to note that the notion of paragraph has been widely mentioned in the literature and proposed to be a grammatical unit at the level between discourse and phrase (e.g. [27]). However, to our knowledge, consensus on how to divide spontaneous spoken discourse into paragraphs is lacking. We will therefore outline the criteria used for Totoli in Section 2.2.

2.1. Participants and data collection procedure

Pear Film [28] monologue retellings [26] were elicited from 16 speakers (8F/8M; age range 31-82). All speakers were bilinguals (Totoli and Indonesian). Eight participants (four female, four male) were native speakers of the Totoli variety spoken in Tolitoli city and surrounding villages (henceforth 'southern dialect'), and the other eight participants (four female, four male) were native speakers of the variety of the Northern Tolitoli District (henceforth 'northern dialect').

The data was recorded with a Zoom Q8 camera and an AKG C520 external microphone. Participants were first shown the Pear Film, a 6-minute videoclip without speech widely used for linguistic elicitation [28], and were told to remember as many details as possible. Thereafter, participants were recorded whilst retelling the film to a Totoli interlocutor. The presence of the interlocutor ensured a more spontaneous nature of the recorded speech. Recordings in which there was a considerable amount of interaction between the interlocutors were excluded from analysis to ensure that all investigated speech was monologous.

2.2. Annotation

The selected recordings were transcribed in ELAN [29], translated into Indonesian by a Totoli native speaker and checked for consistency. Segmentation and annotation in Praat [30] consisted of two steps. First, segmentation into paragraphs and, second, segmentation and labeling of FPs with 'a' or 'e', depending on their perceived vowel quality. A paragraph is understood in this study as a part of discourse that is at least one intonation unit and is semantically coherent. Paragraph boundaries were determined based on prosodic, lexical, morpho-syntactic and semantic criteria. As for prosody, we followed the pattern of boundary tones described in Section 1 (i.e. [2], p.94). Final lengthening was also considered as a paragraph-final prosodic phenomenon. Lexical markers of discourse structure were words that (1) signal change of topic, e.g. *bali* 'so', *ingga (daan) noosa* 'not long after', *danna/daan* 'then', *tooka* 'finished', (2) introduce temporal clauses, e.g. *injan* 'after' or the Indonesian loan words *pas* 'exactly when' and *begitu* 'as soon as'. Morpho-syntactically, the repetition of the final part of the previous sentence as the initial part of the following sentence (tail-head linkage [31]) was taken as a possible paragraph transition. Semantically, paragraph boundaries were determined by dividing the discourse into coherent units, e.g. one event construal per paragraph. Not all prosodic, lexical or morpho-syntactic markers were consistently present, in which case we relied on semantic coherence to set the paragraph boundary.

2.3. Frequency of occurrence

The annotated FPs were categorized according to labeled vowel quality ('a' and 'e') and position relative to the paragraph (vowel quality measures in Section 2.5). Paragraph position was defined as occurring *between* two paragraphs, with the *start* of a paragraph, or *within* a paragraph. The items in each category were counted and a chi-square test was run to investigate whether their distribution was significantly different from chance level.

2.4. Temporal characteristics

The raw duration in ms was measured for all FPs. In addition, the distance to the start of the paragraph was measured for the *within* items. These distance measures were then analysed in a linear mixed model with distance measure (s) as response and vowel quality as factor. The model included a

random intercept for speaker.

2.5. Vowel quality

Acoustic measures of vowel quality were taken in order to confirm that the manual labels indeed correspond to different vowel targets acoustically. If the vowel qualities would overlap, it would be less plausible to assume a two-way distinction for the FPs. Vowel quality (F1 and F2) was measured following the procedure in [32]. This method takes formant measures from subintervals of the intervals corresponding to the FPs. The subinterval was set as the part of the FP for which the intensity dropped maximally 5% at either side of the intensity peak. Formant values are generally the most stable in this part. If the intensity level stayed within the 5% margin at the interval boundary, that boundary was taken as the boundary of the subinterval.

The formant measures were analysed in linear mixed models; one model for each formant. In both models the formant value in Bark was the response, with the interaction between labelled vowel quality (Vq: 'a','e') and dialect (Di: North, South) as factors, and with speaker and paragraph position (between, start, within) as random intercepts.

3. RESULTS

3.1. Frequency of occurrence

Table 1: Number of FPs (χ^2 residuals) split for vowel quality (Vq) and paragraph position.

Vq	Paragraph position		
	between	start	within
a	54 (0.59)	77 (0.90)	20 (-2.08)
e	12 (-1.04)	15 (-1.59)	22 (3.65)

The number of items occurring in each of the categories of FPs is given in Table 1. The results of the chi-square test show that their distribution was significantly different from chance level [$\chi^2(2, N = 200) = 22.42$, $p < 0.001$]. The chi-square Pearson residuals (observed - expected) furthermore indicate that 'a'-like FPs tend to occur in between paragraphs and at paragraph starts, whereas 'e'-like FPs tend to occur within paragraphs (Table 1).

3.2. Temporal characteristics

Table 2: Duration in ms (SD) of FPs split for vowel quality (Vq) and paragraph position.

Vq	Paragraph position		
	between	start	within
a	486.41 (138.67)	267.37 (101.96)	368.09 (199.41)
e	360.08 (148.87)	217.95 (80.18)	432.21 (212.12)

The raw duration measures (Table 2) show that the longest FPs (∼450 ms) are found between paragraphs for 'a' and within paragraphs for 'e'. Intermediate durations (∼360 ms) were found within paragraphs for 'a' and between paragraphs for 'e'. Both vowel categories showed the shortest FP durations (∼240 ms) at the start of paragraphs. The linear mixed model analysis showed that the distance to the start of the paragraph did not differ between 'a'-like items (μ = 4.69, sd = 3.89) and 'e'-like items (μ = 4.33, sd = 3.27): [β = -0.07, SE = 1.09, df = 39.91, t = -0.07, n.s.].

3.3. Vowel quality

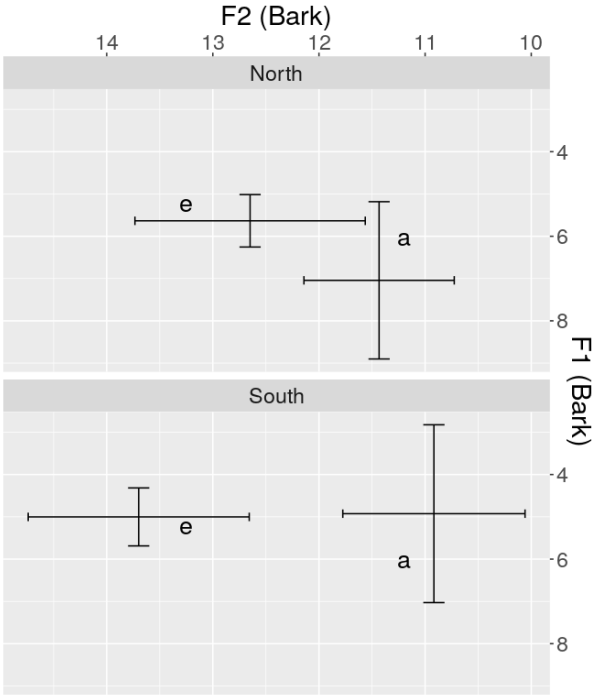

Figure 1: Formant measures for 'a'-like and 'e'-like FPs produced by speakers of the Northern (top) and Southern (bottom) dialect.

Table 3: Results of the linear mixed effects models for the formant measures (F1 and F2)

	Predictor	β	SE	df	t	p
F1	Intercept	7.26	0.47	13.05	15.57	< 0.001
	Vq	-1.22	0.32	186.68	-3.84	< 0.001
	Di	-2.38	0.61	13.29	-3.88	< 0.01
	Vq*Di	0.44	0.56	193.65	0.78	n.s.
F2	Intercept	11.51	0.21	12.43	56.06	< 0.001
	Vq	1.44	0.16	187.22	8.81	< 0.001
	Di	-0.44	0.31	13.17	-1.43	n.s.
	Vq*Di	0.70	0.29	194.09	2.40	< 0.05

The vowel quality results show that 'a' and 'e' were consistently labelled and differ acoustically in a significant way (Table 3). Dialect had a main

effect (F1) and an effect in interaction with vowel quality (F2). F1 values were lower for the Southern dialect than for the Northern dialect (Figure 1). The F2 interaction shows that the 'a' and 'e' values lie closer to each other in the Northen dialect than in the Southern dialect. In the Southern dialect, the F2 values are lower for 'a' and higher for 'e'.

4. DISCUSSION AND CONCLUSION

FPs in Totoli come in two vowel categories that show systematic differences according to paragraph position and speaker dialect. That is, the results revealed tendencies to use 'a'-like FPs near the boundaries of paragraphs (between or at their start), and 'e'-like FPs within paragraphs. Given that they were tendencies, there were still 'a'-like FPs used within pargraphs and 'e'-like FPs used near paragraph boundaries (Table 1). The tendencies were, however, further confirmed by the duration of the FPs; for both vowel categories the average FP was the longest at the position where it tended to occur more often (between and within paragraph). Note that 'a'-like FPs occurred often at the start of paragraphs but were the shortest at that position. This could be explained by the presence of following discourse material, that might have already been planned by the speaker.

The vowel quality measures furthermore confirmed that the vowel categories are acoustically different and that speakers from the northern Totoli dialect produce lower vowels which are more densely distributed within the acoustic space than the speakers of the southern dialect. The difference in vowel height (F1) and vowel place of articulation (F2) are interrelated. This can be explained according to the widely used visualisation of the acoustic space as a triangular shape (i.e. a vowel triangle). The lower the vowels (i.e. more toward the lower tip of the triangle), the closer their positions will be. In the current study, dialectel differences were the largest for 'a'; i.e. [a]-like targets in the southern dialect and [ɑ]-like in the northern dialect. Dialectal differences in FPs were also reported for other languages (e.g. [25]).

Taken together, the results of the current study point at a linguistic status of the FPs in Totoli. That is, they show clear distributional patterns with regard to their position in the paragraph. Their distribution suggests that they have a discourse marking function rather than being produced in an uncontrolled way. In particular the 'a'-like items seem to act as paragraph boundary markers. The possibility cannot be excluded, however, that among the 'e'-like items there might have been some hesitations, rather than discourse markers. This becomes particularly clear from the residuals of the FPs within paragraphs (Table 1). For those items the largest deviations from the expected values are observed, indicating the strongest tendency for 'a'-like items to *not* occur in that position as well as the strongest tendency for 'e'-like items to occur in that position. Thus, assuming that hesitations most often occur within paragraphs, where speakers might pause for planning reasons rather than for discourse structure reasons, the current data suggest that these would rather be 'e'-like. In addition, 'e'-like FPs have a wider spread in their F2 values (Figure 1), which indicates realisation toward the center of the acoustic space. Although central vowels have been reported to occur in linguistically relevant FPs crosslinguistically [20], they seem to be a minority in Totoli. Generally, FPs in Totoli occur as two vowels that are acoustically distinct. This could be taken as an additional indication of their linguistic nature. That is, their acoustic realisations map onto vowel targets and show dialectal variation. It could be speculated that 'a'-like FPs are derived from a demonstrative, i.e. *ia* (proximal), *ana* (medial) and (less likely) *itu* (distal), as listed in [1] (p.99). Also, it cannot be excluded that FPs originate from interjections such as *ya* or *ye* 'yes'. Totoli FPs thus show a clear correlation with linguistic units and are more articulate than central (schwa-like) vowels.

It remains to be seen whether there are more (discourse) structural differences in the use of 'a' or 'e' in Totoli FPs. One specific question pertains to the distribution of FPs at the start of paragraph-internal intonation units (the *within* items in this study). If 'e'-like items were indeed more often hesitations, it could be expected that they would not occur as often at the start of intonation units. That position could be occupied more often by 'a'-like items, just as they do at the level of paragraphs. This, as well as morpho-syntactic aspects of paragraphs, will be investigated in future studies.

5. ACKNOWLEDGEMENTS

Research for this paper was funded by the German Research Foundation (DFG) - Project-ID 281511265 - SFB 1252. The recordings were made by Christoph Bracks and the second author. The data (recordings, textgrids and dataframe) are made available at https://osf.io/yah89/. We gratefully acknowledge the Totoli speakers who participated in the recordings and contributed to the transcriptions.

6. REFERENCES

[1] N. P. Himmelmann, Ed., *Sourcebook on Tomini-Tolitoli languages: general information and word lists*, 1st ed., ser. Pacific linguistics. Canberra: Pac. Ling., Research School of Pacific and Asian Studies, Australian National U., 2001, no. 511.

[2] C. Bracks, "Intonation Units and grammatical units in Totoli," PhD Thesis, Universität zu Köln, 2020.

[3] S. Inghuong, A. Adnan, A. G. Hali, I. Halim, and N. Baso, *Morfologi dan sintaksis bahasa Totoli*. Palu: PPBSIDST, 1990.

[4] I. A. Sofyan, A. Adnan, Z. Mahmud, and M. Masyhuda, *Struktur bahasa Totoli*. Jakarta: Pusat Bahasa, 1991.

[5] N. P. Himmelmann and S. Riesberg, "Symmetrical Voice and Applicative Alternations: Evidence from Totoli," *Oceanic Linguistics*, vol. 52, no. 2, pp. 396–422, 2013.

[6] S. Riesberg, "A first take on information structure in Totoli - Reference management and its interrelation with voice selection," *Proceedings of the 2nd International Workshop on Information Structure of Austronesian Languages*, pp. 65–81, 2015.

[7] ——, "'Fern leaf, so you are challenging me?' Some observations on the Lelegesan, a form of verbal combat in Totoli," 2019.

[8] S. Riesberg, K. Malcher, and N. P. Himmelmann, "The many ways of transitivization in Totoli," in *Valency over Time*, S. Luraghi and E. Roma, Eds. De Gruyter, 2021, pp. 235–264.

[9] N. P. Himmelmann and S. Riesberg, "Expressions of directed caused accompanied motion events in Totoli, a western Austronesian language of Indonesia," in *Typological Studies in Language*, A. Margetts, S. Riesberg, and B. Hellwig, Eds. Amsterdam: John Benjamins Publishing Company, 2022, vol. 134, pp. 219–242.

[10] M. Bardají i Farré, S. Riesberg, and N. Himmelmann, "Limited-Control Predicates InWestern Austronesia: Stative, Dynamic, Or None Of The Above?" *Oceanic Linguistics*, 2021.

[11] S. Riesberg, M. Bardají i Farré, K. Malcher, and N. P. Himmelmann, "Predicting voice choice in symmetrical voice languages: All the things that do not work in Totoli," *Studies in Language*, vol. 46, no. 2, pp. 453–516, 2022.

[12] M. Bardají i Farré, "Nominalization in Totoli and other western Austronesian languages," PhD Thesis, Universität zu Köln, 2022.

[13] G. Yule, "Speakers' topics and major paratones," *Lingua*, vol. 52, no. 1-2, pp. 33–47, 1980.

[14] A. Sluijter and J. Terken, "Beyond Sentence Prosody: Paragraph Intonation in Dutch," *Phonetica*, vol. 50, no. 3, pp. 180–188, 1993.

[15] I. Lehiste, "The Phonetic Structure of Paragraphs," in *Structure and Process in Speech Perception*, K. S. Fu, W. D. Keidel, H. Wolter, A. Cohen, and S. G. Nooteboom, Eds. Berlin, Heidelberg: Springer, 1975, vol. 11, pp. 195–206.

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

[17] N. Amiridze, B. H. Davis, and M. Maclagan, Eds., *Fillers, pauses and placeholders*, ser. Typological studies in language. Amsterdam Philadelphia: Benjamins, 2010, no. 93.

[18] F. Goldman-Eisler, "A Comparative Study of two Hesitation Phenomena," *Language and Speech*, vol. 4, no. 1, pp. 18–26, 1961.

[19] M. Hayashi and K.-E. Yoon, "A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble," *Studies in Language*, vol. 30, no. 3, pp. 485–540, 2006.

[20] H. Clark and J. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, no. 1, pp. 73–111, 2002.

[21] Y. Maschler, "Discourse markers at frame shifts in Israeli Hebrew talk-in-interaction," *Pragmatics*, vol. 7, no. 2, pp. 183–211, 1997.

[22] N. Iwasaki, "18 Japanese fillers as discourse markers: Meanings of "meaningless" elements," in *Handbook of Japanese Semantics and Pragmatics*, W. M. Jacobsen and Y. Takubo, Eds. De Gruyter, 2020, pp. 799–838.

[23] J. Brody, "Particles Borrowed from Spanish as Discourse Markers in Mayan Languages," *Anthropological Linguistics*, vol. 29, no. 4, pp. 507–521, 1987.

[24] Y. Zhao and D. Jurafsky, "A preliminary study of Mandarin filled pauses," in *Disfluency in Spontaneous Speech*, 2005, pp. 179–182.

[25] Y. Tian, T. Maruyama, and J. Ginzburg, "Self Addressed Questions and Filled Pauses: A Cross-linguistic Investigation," *Journal of Psycholinguistic Research*, vol. 46, no. 4, pp. 905–922, 2017.

[26] M. Bardají i Farré and C. Kaland, "Phonetic description of filled pauses as discourse markers in Totoli," 2022, https://osf.io/yah89/.

[27] R. Longacre, "The Paragraph as a Grammatical Unit," in *Discourse and Syntax*, T. Givon, Ed. BRILL, 1979, pp. 113–134.

[28] W. L. Chafe, *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, N.J: Praeger, 1980.

[29] Max Planck Institute for Psycholinguistics, "ELAN," Nijmegen, 2022, https://archive.mpi.nl/tla/elan.

[30] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer," 2022, http://www.praat.org/.

[31] L. De Vries, "Towards a typology of tail–head linkage in Papuan languages," *Studies in Language*, vol. 29, no. 2, pp. 363–384, 2005.

[32] C. Kaland, "Acoustic correlates of word stress in Papuan Malay," *Journal of Phonetics*, vol. 74, pp. 55–74, 2019.