

The intra-word pause and disfluency in Dalabon

Janet Fletcher*, Nicholas Evans*, & Belinda Ross**

* University of Melbourne, Australia

** University of Aarhus, Denmark

Abstract

Earlier impressionistic analyses of Dalabon indicate that the grammatical word is often realized as either an accentual or an intonational phrase, followed by a pause. Unusually, it can also be interrupted by a silent pause, with each section being potentially (although not necessarily) realized as separate intonational phrases. Our analyses of pause duration and pause placement within grammatical words support these earlier impressions, although this use of the silent pause appears to be restricted to certain affix boundaries, and other phonological constraints relating to the following surrounding linguistic material. These interruptions also share certain characteristics of “normal” disfluencies however.

1. Introduction

There are relatively few experimental studies of disfluencies and repairs in typologically unusual languages. This paper examines silent pause placement duration and disfluency in Dalabon, a polysynthetic Australian language spoken in Arnhem Land in the Northern Territory of Australia (see Fig. 1). One interesting feature of Dalabon is that grammatical words can be interrupted by placing a silent pause after a pronominal prefix. Evans et al. [1] found that this is not a widespread feature of pausing in the language. It is also not at all clear what additional “functions” these pauses might have, if we can assume that speakers use pausing strategies as one means of structuring their spontaneous discourse. It is also not clear how these silent pauses differ from those that might more generally be associated with disfluency on the one hand, or the marking of larger discourse segments on the other. Figure 2 shows a typical example of this phenomenon. The prefix “kenh” is isolated from the rest of the following grammatical word by a long pause. The second (much shorter) pause also interrupts the fully inflected grammatical word.

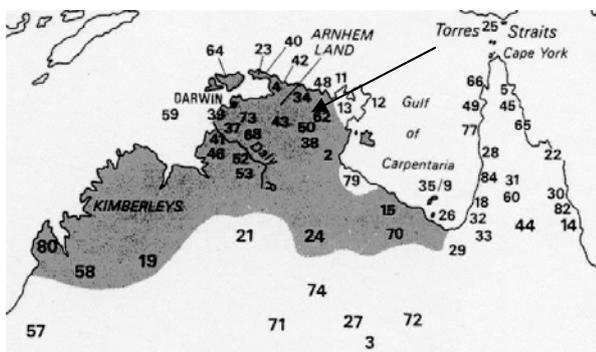


Figure 1: A map of Northern Australia showing where Dalabon is still spoken. The map also shows the location of other indigenous languages that are still spoken in Northern Australia

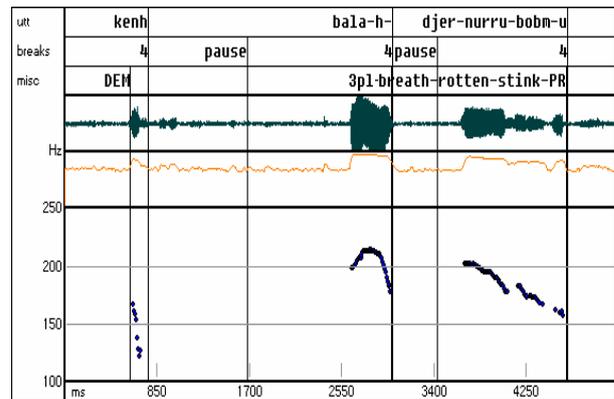


Figure 2: Speech waveform, fundamental frequency contour and rms amplitude trace for a stretch of speech illustrating an isolated prefix “kenh”, a demonstrative.

Recent analyses of disfluencies (e.g. Shriberg, [2]) suggest that a distinction needs to be drawn between unfilled pauses that are part of the “editing” phase in the speech production process, and those that might be considered planning pauses. The temporary interruption may signal a following repair or correction. This does not appear to be the case in the above example. The first interruption in the grammatical word illustrated in Figure 2 is likely to be planned, in so far as the initial isolated unit reflects the properties of a full intonational phrase in Dalabon, with an intonational pitch accent and a final falling boundary tone. However, it is not clear whether the second pause (evident by the second silent stretch in the speech waveform) is the same kind of pause.

It appears that disfluency in Dalabon is rarely signaled by filled pauses like “um” or “uh” that you find in General Australian English or “eh” in Scottish English, for example. While it remains to be empirically tested, silent pauses seem to be the main cue to signal interruption to the speech production process in this language.

If we take into consideration, the three regions in a disfluency, (after Levelt, [3] & Shriberg, [2]), the Reparandum, the Editing phrase, and the Repair, it may well be that we need to examine the characteristics of the first and last regions with respect to disfluencies in Dalabon. Disfluency pauses appear to coincide with repetition, deletion or substitution of units, such as pronominal prefixes, which should otherwise be attached to a following verbal unit. When there is a true Repair, that is when the morpheme or morpheme sequence is either repeated or changed to a different unit, there is still a perceptibly clear silent pause. Initial observations also suggest that the intonation contour is slightly different at the edge of the Reparandum region, with a less clearly defined falling intonation pattern. Rather, the tune tends to be either slightly falling, or sustained mid-level.

Figure 3 below shows an example of an interrupted verbal word complex that shows the regions in disfluency. From the speech waveform, it is evident that there are two major silent pauses. From left to right, we see a Reparandum, which is then repaired, but this in turn is effectively both a Repair and a Reparandum, which by the third and final stretch of speech shown below, is a true Repair. It remains to be seen however, whether this is a typical pattern in spoken Dalabon narratives.

The aims of this study were to investigate intra-word pauses, and silent pauses in disfluency regions to see whether there was any significant difference between the duration of these pause-types, and the pauses that coincide with larger discourse segments. An additional aim was to see whether the right boundary of any Reparandum was also signaled by a different F0 level to those that usually signal the edge of complete grammatical units that coincide with full intonational phrase edges, or units consisting of detached prefixes that may or may not coincide with intonational phrase boundaries.

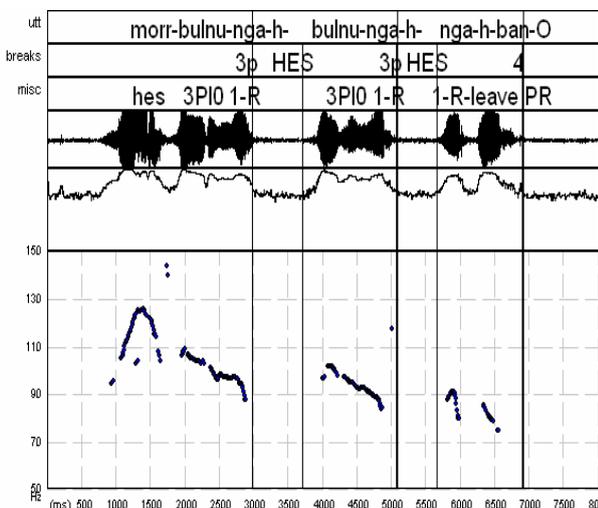


Figure 3. Speech waveform, F0 & Rms contour for a Dalabon utterance showing 2 unfilled pauses in Dalabon preceding 2 Repairs for the utterance: ‘them, I, them, I, I leave them’

2. Method

The corpus in this study consisted of two narratives produced by a male and female speaker of Dalabon. Approximately 30 minutes of connected speech were analysed for the female speaker (MT), and 15 minutes for the male speaker (JC). The corpus was recorded by the second author in Northern Arnhem Land during a linguistic fieldtrip. The field tapes were transcribed in practical orthography (see Tables 1 & 2), and also glossed and translated by the second author.. The recordings were subsequently digitized at 22 KHz using eps/Waves+, running on a SUN work station in the Phonetics Laboratory, University of Melbourne. The signal files were then annotated using EMU (Cassidy & Harrington, [4]). Silent pauses were identified and annotated using the acoustic waveform and spectrogram as a guide. Any silent gap in the waveform of 200 ms or more was labeled as an unfilled pause. A conservative duration threshold was chosen because long stops in this language are often between 150-180 ms or more in duration (Evans et al [1]).

Break indices were also annotated. Earlier prosodic studies of Dalabon (e.g. Fletcher & Evans, [5]), as well as the current one, are located within the autosegmental-metrical (A-M) intonational framework (e.g. Pierrehumbert, [6] Ladd, [7]) among others. A version of ToBI (Tones and Break Indices)

has also been devised for the language, whereby levels of prosodic constituency are also annotated, along with intonational targets (after Beckman & Ayers-Elam, [8]). Earlier work suggests that minimally three levels of constituency need to be acknowledged in Dalabon: a break index value of 1 indicates there is minimal juncture between adjacent words; a break index value of 3 indicates tonal juncture of some kind, i.e. a falling or rising intonation contour the end of a word, followed by a pitch restart on a following word. Provisionally, we will describe this as an accentual phrase. Finally, a break index value of 4 indicates a full intonational phrase. A break index (BI) 4 marks the highest degree of perceived juncture. The phonetic cue indicating a break 4 constituent is a phrase-final intonational movement, lengthening of a final syllable, and generally a silent pause. It was expected that detached prefixes that displayed the above characteristics were realized as well-formed prosodic constituents. In the case of disfluency regions however, a BI value of 4p was used because it was not at all clear whether the Reparandum region of disfluency was actually a full intonational phrase break. F0 values were extracted at these label points to get an indication of pitch height at the edge of Reparanda, versus full intonational phrase boundaries marked by a BI 4. The distribution of silent pauses, median pause duration was calculated. Instances of a grammatical word being interrupted by a pause were also noted, as were disfluencies.

Table 1. Consonant contrasts in Dalabon. The practical orthography used for each sound is included in parentheses.

		Place of Articulation					
		Peripheral		Apico-		Lam ino- pala tal	Glott al
		Bilab ial	Vel ar	alveol ar	retro flex		
Manner of Articulation	Short stop	b) . . . □	(k) . □	(d) . □ . .	□	□	(dj) . . .
					(rd)		
	Long stop	□□	□□	□□	□□	□□	
		(bb)	(kk)	(dd)	(rdd)	(djdj)	
	Nasal	□ .	□ .	□ .	□	æ□	
		(m)	(ng)	(n)	(rn)	(nj)	
Lateral			□	□			
			(l)	(rl)			
Rhotic			□	□			
			(rr)	(r)			
Semi-vowel	□ .				□ .		
	(w)				(y)		

Table 2. Vowel contrasts in Dalabon

	Front	Central	Back
Close	i		u
Mid	e		o
Open		a	

minimal juncture between adjacent words; a break index value of 3 indicates tonal juncture of some kind, i.e. a falling or rising intonation contour the end of a word, followed by a pitch restart on a following word. Provisionally, we will describe this as an accentual phrase. Finally, a break index value of 4 indicates a full intonational phrase. A break index

(BI) 4 marks the highest degree of perceived juncture. The phonetic cue indicating a break 4 constituent is a phrase-final intonational movement, lengthening of a final syllable, and generally a silent pause. It was expected that detached prefixes that displayed the above characteristics were realized as well-formed prosodic constituents. In the case of disfluency regions however, a BI value of 1p was used because it was not at all clear whether the Reparandum region of disfluency was actually a full intonational phrase break. F0 values were extracted at these label points to get an indication of pitch height at the edge of Reparanda, versus full intonational phrase boundaries marked by a BI 4. The distribution of silent pauses, and mean pause duration were calculated. Instances of a grammatical word being interrupted by a pause were also noted, as were disfluency regions, where the editing phase was a silent pause. That is to say, we separated those kinds of grammatical word interruptions that involved the detachment of a prefix, from other kinds of interruptions that were more like a “conventional departure from fluency” (Shriberg, [2:160]).

3. Results

3.1. Pause duration and pause distribution

Figures 4-6 plot the durational distribution and length (ms) of all unfilled silent pauses, pauses that are part of the editing phase of a disfluency region, and pauses within grammatical words (after detached prefixes) for speakers MT and JC. Altogether, 648 silent pauses were measured across the corpus. As would be expected, the distribution of pauses is skewed to the left. The mean duration of silent pauses that are preceded by full intonational phrase boundaries, that coincide with fully inflect grammatical words is 941ms for speaker MT and 612 ms for speaker JC. For ease of exposition, we refer to these pauses as “standard” pauses. The pauses range in duration from 207 ms to 6160 ms for MT, and 200ms and 1874 ms for JC. The latter speaker pauses less often than speaker MT

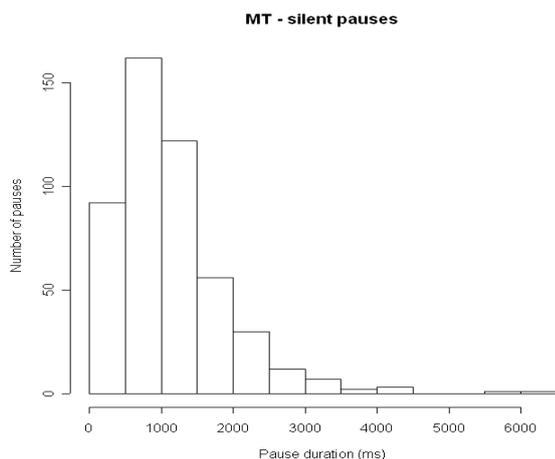


Figure 4a. Distribution and duration of intonational phrase final silent pauses produced by speaker MT.

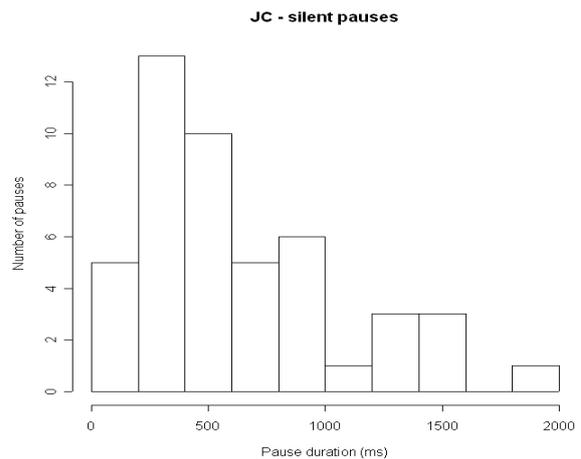


Figure 4b. Distribution and duration of intonational phrase final silent pauses produced by speaker JC.

The results of a one-way ANOVA are highly significant for speaker MT ($F=17.11$, $p<0.0001$) suggesting that there are clear differences in duration for the different pause types for this speaker. The differences are less significant for speaker JC ($F=3.42$, $p<0.05$), although a similar trend is apparent. Figures 4a and 4b, show that the proportion of hesitation silent pauses (i.e. pauses that constitute the “editing” phase of a Reparandum) are fewer in number than “standard” silent pauses, and tend to be somewhat shorter. There are 66 instances of silent pauses following a Reparandum and preceding a Repair for speaker MT, and only 18 for speaker JC. Comparing the pause durations plotted in Figures 5a & b, there are no significant differences between the mean duration of disfluency pauses (481ms) and within-word pauses, however ($t=1.03$, $p>0.05$) for speaker MT.

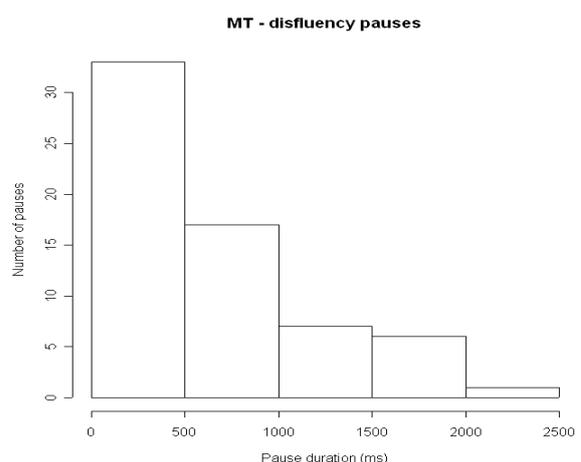


Figure 5a. Distribution and duration of “disfluency” silent pauses produced by speaker MT.

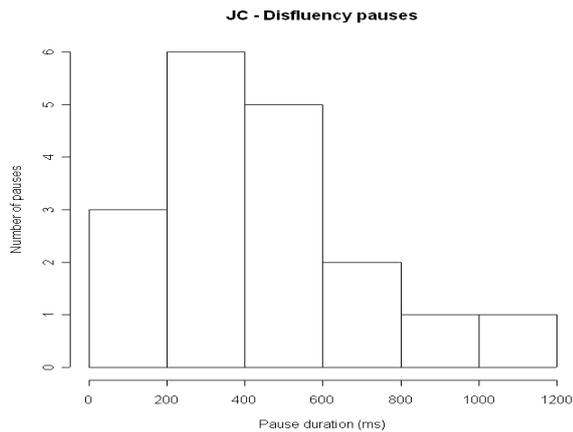


Figure 5b. Distribution and duration of “disfluency” silent pauses produced by speaker JC.

Figure 6 shows there are far fewer pauses within grammatical words, i.e. after detached prefixes, than standard pauses that occur after “complete” grammatical words (27 versus 488 instances for speaker MT). Speaker JC produced only 4 instances of detached prefixes followed by a pause. For MT, the difference in mean duration between the two types of pause is highly significant ($t=5.05$; $p<0.0001$), with intra-word pauses being more than half the length of standard pauses (440 ms versus 981ms). The four intra-word pauses produced by JC are also very short, being around 200-250 ms in length.

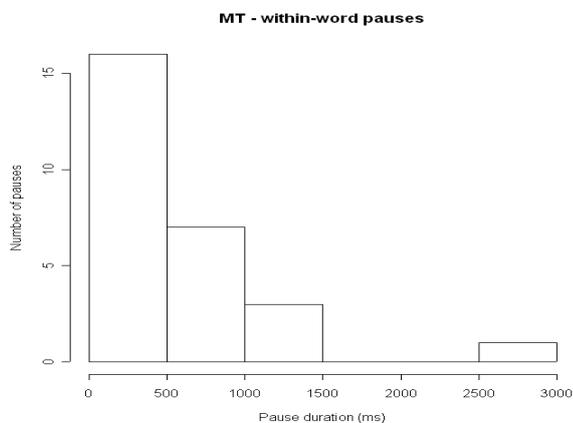


Figure 6. Distribution and duration of “within-word” silent pauses (not hesitation disfluencies) produced by speaker MT.

3.2 F0 analysis

Figures 7a & b plot the final measurable F0 value in the final voiced segment of the constituent preceding a) a “legal” within word pause i.e. at the right edge of a detached prefix b) a Reparandum, and c) a fully formed grammatical constituent and intonational phrase. Both a) and b) are labeled as 4p on the Break Index tier. The analysis of F0 values associated with the edge of BI 4 and 4p constituents shows that for speaker MT, (Figure 4a) there is a small difference between all three boundaries. Most BI4 prosodic constituents have a lower F0 value (116 Hz) at their right edge, than BI 4p constituents (138 Hz). This difference is weakly significant ($w=5050$; $p<0.05$). However there are no significant differences between F0

values at the edge of BI4p constituents that are detached prefixes (the left most plot in Figure 6a), and F0 values at the edge of Reparanda (the middle plot).

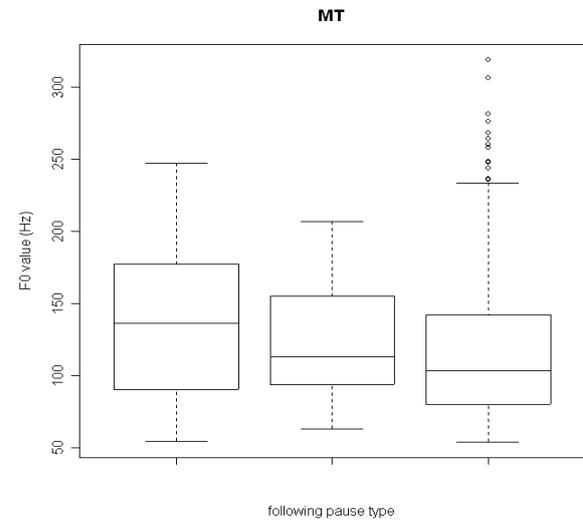


Figure 7a. F0 values extracted at the edge of verbal material preceding the three different pause types: within-word pauses (i.e. after detached prefixes), editing phase interruptions, and intonational phrase-final pauses for speaker MT.

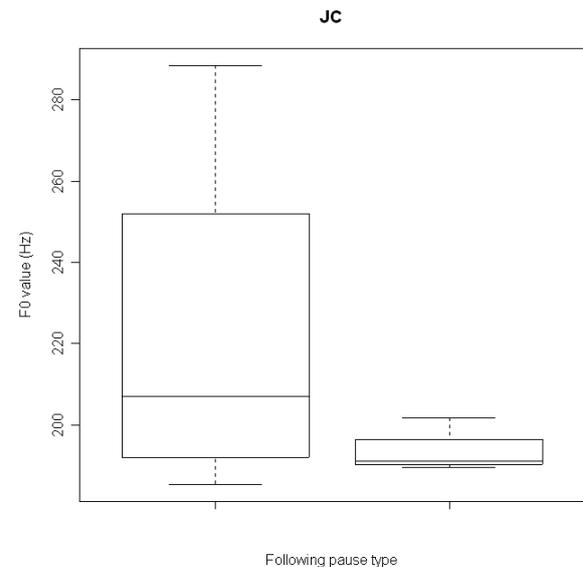


Figure 7b. F0 values extracted at the edge of verbal material preceding two different pause types: editing phase interruptions and intonational phrase-final pauses for speaker JC.

There are differences in the mean F0 values at BI4p and BI4 boundaries for Speaker JC, but these are not statistically significant (Figure 7b). However, speaker JC produced a large number of H% boundaries at BI4 edges which will have significantly influenced the F0 analysis shown in Figure 7b. At BI4p edges, a range of tune levels were observed, which is also clearly reflected here. Speaker MT, on the other hand produced relatively more falling tunes at the edge of BI4 constituents (i.e. L% boundary tones). Relatively low tune values were also apparent at BI4p edges but these were not as low as at BI4 edges. Presumably local pitch range effects like final lowering at the right edge of major discourse segments

will have also influenced the BI4 F0 values observed for both speakers.

4. Discussion

Our results suggest that there may not be significant phonetic differences between the phonetic cues at the right edges of Reparandum regions of disfluencies and detached prefixes. While there are clear pause duration differences at BI4 edges (i.e. “standard” pauses) and at interruption points in the discourse for both speakers, within this category, intra-word pauses (i.e. after detached prefixes) are of similar duration to editing phase pauses. Results of the F0 analysis are less clear cut. However, these values also reflect the different boundary tones used by speakers at BI4 constituent edges. Nevertheless, detached prefixes can exhibit similar boundary configurations to those observed at full intonational phrase edges, but these are also observed at the edges of Reparanda. A closer analysis of these boundaries will be undertaken to determine how many of the detached prefixes really do exhibit full prosodic characteristics of an intonational phrase.

It is perhaps no surprise that we have observed similar patterns between the two kinds of interruptions, particularly when we look at the conditions surrounding “detachment” of prefixes. The detached prefixes are the clearest example of a non-isomorphism between a grammatical word and a phonological word. The two relevant phonological requirements that must be met before pause can occur are that

(a) This must not split a foot, though since feet are constructed over morphemes in virtually all cases, this could also be phrased as a morphological requirement. Because feet have a bimoraic minimum, this effectively means that the prefix must be closed, if monosyllabic, and otherwise satisfy a disyllabic minimum

(b) The remainder of the word must have at least two syllables.

The detached prefixes in this case may have a fully formed intonational boundary. Hesitation pauses function similarly to traditional notions of the Editing phase of a Disfluency region. They occur in connection with repetition, deletion or substitution of units. Interestingly, many of the disfluency regions in this Dalabon corpus involve verbal words that include pronominal prefixes. In other words, the detached prefixes are often the elements that are repeated, or deleted, and are usually re-attached to the preceding word in the discourse. Editing phase interruptions differ from the intra-word pauses in this corpus, in that the unit preceding a pause is either repeated or changed to a different unit, whereas the detached prefix that precedes an intra-word pause is more likely (although not always) to constitute its own prosodic phrase (or BI4 constituent) with a fully formed intonational contour. Once again, a closer analysis of the tonal characteristics at the edge of these units needs to be undertaken.

Diachronically, there are reasons to regard the possibility of breaking a single verbal word into a number of “pause units” as a Dalabon innovation compared to related languages of the region. Comparison with other Gunwinyguan languages, spoken in the same region of Northern Arnhem Land, shows the situation in Bininj Gun-wok, rather than Dalabon, to be the norm – in other words, prefixes do not detach in these languages. Moreover, the emergence of sub-word pause units appears to be linked to a morphological innovation in Dalabon which has had important phonological consequences: the

extension of a codal glottal-stop following pronominal prefixes to become the unmarked TAM value, signaling assertativity, rather than the marked type that it is in BGW, where it is confined to the much rarer ‘immediate’ aspect. We suggest speculatively that this is a historical evolution – what may have started out as a “normal” interruption or a disfluency many well have resulted in a process whereby the relatively long grammatical words of polysynthetic languages may well contain smaller phonological words with relevant prosodic characteristics associated with this level of prosodic constituency. The intra-word pause may well be an indicator of this change in progress in Dalabon.

Whilst preliminary in nature, these findings suggest that an examination of spoken discourse in languages other than the mainstream European or Asian languages may shed light on a number of phonetic and phonological issues, including the role of pause in the structuring of spoken communication

5. Acknowledgements

This research was supported by the Australian Research Council. Thanks to Debbie Loakes who assisted in the analysis of Speaker JC’s narrative.

6. References

- [8] Beckman, M. & Ayers-Elam, G. (1994). Guidelines for ToBI labelling http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/singer_tobi.html
- [4] Cassidy, S. and Harrington, J. (2001). Multi-level annotation in the Emu speech database management system. *Speech Communication*, 33, 61-77.
- [1] Evans, N., Fletcher, J. & Ross, B. (in review) Big words, small phrases: mismatches between pause units and the polysynthetic word in Dalabon. *Linguistics*
- [5] Fletcher, J. & Evans, N. (2002). An acoustic intonational study of intonational prominence in two Australian languages. *Journal of the International Phonetic Association*, 32, 123-140.
- [7] Ladd, D.R (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- [3] Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition* 14, 41-104.
- [6] Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Unpublished MIT PhD thesis.
- [2] Shriberg, E. (2001). To ‘err’ is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31, 153-169.