A cross-linguistic study of prosodic lengthening in child-directed speech

Elinor Payne[1], *Brechtje Post*[2], *Lluïsa Astruc*[3], *Pilar Prieto*[4,5], *Maria del Mar Vanrell*[4]

[1] Phonetics Laboratory, University of Oxford, and St Hilda's College, Oxford, UK
[2] RCEAL, Cambridge University, and Jesus College, Cambridge, UK
[3]Department of Languages, The Open University, Milton Keynes, UK
[4] Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Spain
[5] ICREA, Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

elinor.payne@phon.ox.ac.uk, bmbp2@cam.ac.uk,
mla28@cam.ac.uk,mariadelmar.vanrell@upf.edu, Pilar.prieto@upf.edu,

Abstract

We compared the manifestation of prosodic lengthening in child-directed speech with the adult-directed speech of 6 Catalan and English women. Results showed a less variegated system of prosodic lengthening in CDS, with the selective enhancement of certain structures, notably phrase-final lengthening and (in English) nuclear accented syllables.

Index Terms: child-directed speech, prosodic lengthening, Catalan, English.

1. Introduction

It has been widely documented that child-directed speech (CDS) may differ notably from adult-directed speech in terms of prosody (e.g. [1] [2] [3] [4]). Common features include higher and greater range of pitch, especially in stressed syllables; longer duration of individual words, more prominent final lengthening, slower speech rate, higher amplitude, shorter utterances and longer pauses, more even rhythm, and more reliable positioning of pauses at phrase boundaries ([2] [5] [6] [7]).

From a functional point of view, one thesis is that the richness of prosodic cues in CDS may function to attract and maintain the attention of the child (e.g. [2] [5]) and communicate affective information [8]. Furthermore, it has been argued that CDS may actively facilitate language acquisition by, for example, highlighting linguistic structure and aiding word identification [9] [10] [11] [12]. However, evidence also suggests that certain kinds of prosodic modification (e.g. rhythmic) may be the result of an accommodation towards the child ([6] [7]). More generally, research has shown that young children and infants prefer listening to CDS to listening to ADS ([9] [13] [8]).

Cross-linguistic studies have shown certain features to be particularly prevalent: for example higher pitch and wider pitch range [14] [2] [4] [15] [16]. However, the extent of exaggeration varies and exceptions exist: [4] report less exaggerated pitch characteristics in CDS in Mandarin Chinese (a tone language) than in American English. Most studies, be they monolingual or crosslinguistic, investigate the degree to which certain phonetic parameters of prosody are amplified in CDS. A little studied question is the extent to which certain aspects of prosody may be prioritized or even suppressed. In other words, does CDS merely exaggerate the marking of prosodic structure in ADS, or is it more selective?

As an initial foray into this issue, this paper examines the phenomenon of prosodic lengthening in English and Catalan. It asks the following research questions: i) is there evidence of greater prosodic lengthening in CDS?; ii) if so, are certain prosodic structures highlighted more than others; iii) and if so, are there cross-linguistic similarities in this, or is the selection more language-specific?

2. Methodology

2.1 Participants

This study forms part of a wider project investigating the acquisition of rhythmic and intonational properties in Catalan, English and Spanish. For the present study, we investigated the CDS and ADS of 3 Catalan-speaking and 3 English-speaking female adults, interacting with their 2-year old children and an adult interviewer.

2.2 Materials and elicitation

The data consisted of short question-and-answer dialogues, elicited through a structured game, based on short, animated clips, shown on Powerpoint slides on a laptop screen. The clips showed simple, everyday scenes, which could easily be described in words that were highly familiar to the children. For example, one scene showed a little girl blowing soap bubbles, another showed a little boy playing with building blocks. For the CDS, the mother was instructed to ask her child to describe what was happening in each clip, then to praise the child for getting it right, and repeat what the child had said. For the ADS, the mothers were recorded doing the same task, but interacting with an adult (the interviewer). A typical dialogue went thus:

CDS context
Mother: "*What's happening? What's the girl doing?"*
Child: *"(She's) blowing bubbles!"*
Mother: "*That's right!* ***She's blowing bubbles****!*"

ADS context
Mother: "*What's happening? What's the girl doing?"*
Interviewer: *"She's blowing bubbles!"*
Mother: "*That's right!* ***She's blowing bubbles!"***

The utterances used for the present study were the target utterances spoken by the mother (in the example, the instances of "*She's blowing bubbles*" in bold and underlined). Recordings were made in the participants' homes in Cambridge and Barcelona, using a Marantz PMD660 recorder and Shure PG81 microphones for the Catalan recordings, and a Tascam HD-P2 recorder with AKG C3000B microphones for the English recordings.

2.3 Labelling and analysis

Syllables were segmented from the waveform and spectrogram and start-points and end-points labelled on a syllabic tier by a research assistant using *Praat*. Vocalic and consonantal segmentation was performed with reference to standard criteria. Boundary placement between vocalic and consonantal intervals was guided primarily by the presence of a sudden, significant drop in amplitude and a break in the formant structure, particularly F2. Marking the onset of

consonants was facilitated by various cues, according to the manner of consonant. For instance, the onset of a fricative was marked at the start of high frequency energy (visible frication). The onset of a nasal consonant was marked by the presence of nasal formant structure and low amplitude in the waveform. Any aspiration following stop release was included as part of the consonant interval.

In addition to syllable boundaries, syllables were also labelled according to segmental content, e.g. CV, CVC, CCV, CVCC etc, and according to level of prominence. Although twelve different syllable structure types were extracted from the speech files, the incidence of most of these was very low, and in many cases did not contain the complete range of other factors being tested. It was thus decided that only the two most common syllable structure types, CV and CVC, would be used. These also had the advantage of being minimally contrastive, and contrasting in a critical way, representing the simplest kind of open and closed syllables respectively. For prosodic prominence, syllables were labelled as belonging to one of four levels: i) lexically unstressed; ii) lexically stressed but not accented; iii) lexically stressed and accented; iv) nuclear accented. Nuclear pitch accents always fell on the final word in the phrase.

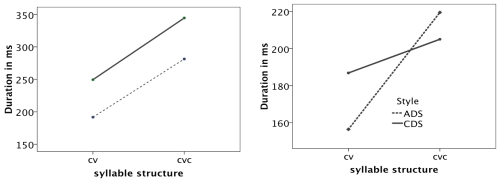
Intonational phrase boundaries and word boundaries were marked on a further two tiers. Both intermediate and intonational phrases were identified. However, since there were very few examples of intermediate phrases, only intonational phrases were included in the present analysis. By referring to these two tiers, syllables could be categorised according to phrase position (initial, medial or final) and word position (initial, medial or final). In the case of monosyllabic words, syllables aligned with both the beginning and end of a word boundary, and these were labelled as a separate category. Syllable durations, together with all other prosodic and word boundary information, were extracted using a Praat script. This yielded a total of 1170 syllables, across language and speech style.

3. Results

3.1 Comparison of speech styles

A repeated measures ANOVA was run on the combined dataset of ADS and CDS results for each language, to verify any significant difference between speech styles. The results suggest language-specific modifications in the parameters of prosodic lengthening for CDS.

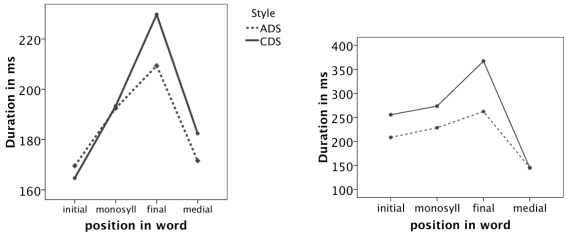
Figures 1a (left) and 1b (right): mean syllable duration (in ms) for different syllable structures in CDS and ADS for English (1a) and Catalan (1b)



Speech style had a main effect in English, with mean syllable duration longer in CDS than in ADS. There were no significant interactions between speech style and any of the other variables. In Catalan, on the other hand, though speech style did not prove to have a main effect on syllable duration, there was a significant interaction between style and syllable structure ($F(1,696)=10.16$, $p<0.05$). As Figures 1a and 1b show, while in English, syllables are longer in CDS regardless of syllable structure, in Catalan, open syllables are longer in CDS, but closed syllables are actually *shorter*. This results in greater uniformity of duration across syllable structure types in CDS than in ADS, for Catalan.

In Catalan there was also a significant interaction between style and word position ($F1(3,696)=5.02$, $p<0.05$). As Figures 2a and 2b show, while in English syllables are longer in CDS roughly to the same degree across the word (except word-medially), in Catalan, they are longer in CDS specifically when word-final in a polysyllabic word (and to a lesser extent word-medially).

Figures 2a (left) and 2b (right): mean syllable duration (in ms) for different word position in CDS and ADS for Catalan (1a) and English (1b)

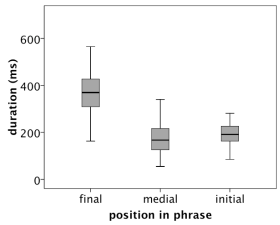


Following this, for each speech style and within each language, a series of univariate Anovas was run with syllable structure, level of prosodic prominence, position in the word and position in the phrase as factors. Through this analysis, lengthening patterns were first established for ADS and then compared with patterns found in CDS.

3.2 English ADS

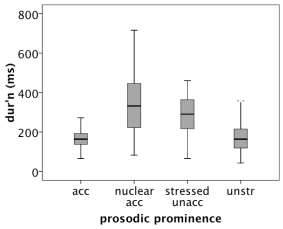
The following variables were found to have a main effect on syllable duration in ADS: syllable structure (closed syllables longer) ($F(1,251)=77.01$, $p<0.001$), prosodic prominence ($F(3,251)=6.18$, $p<0.001$) and phrase position ($F(2, 251)=39.58$, $p<0.001$). Word position did not have a main effect on syllable duration and there were no interactions.

Figure 3: mean syllable duration (in ms) for different phrase positions for English ADS



Bonferroni post hoc comparisons showed syllables in phrase-final position to be significantly longer than in phrase-medial ($p<0.001$) and phrase-initial ($p<0.001$) position (see Figure 3). In addition, phrase-initial syllables were marginally longer than phrase-medial syllables, and this proved significant ($p<0.05$).

Figure 4: mean syllable duration (in ms) for different prominence levels for English ADS



Bonferroni post hoc comparisons revealed nuclear accented syllables to be significantly longer than other accented syllables ($p<0.001$), stressed unaccented syllables ($p<0.001$) and unstressed syllables ($p<0.001$) (see Figure 4). Stressed unaccented syllables were also significantly longer than unstressed syllables ($p<0.001$) and, somewhat anomalously, (non-nuclear) accented syllables ($p<0.001$), which were themselves no longer than unstressed syllables. In other words, only if a pitch accent was nuclear did it increase

duration. Furthermore, while a lexical stress also increased duration, this effect actually appeared to be suppressed when coincident with a (non-nuclear) pitch accent. There is, it seems, a possible trading effect between pitch movement and durational cues. We might postulate that the suppression of lengthening effects in a non-nuclear position functions to increase the salience of the nuclear accent, where the cues of pitch movement and lengthening cumulatively gather.

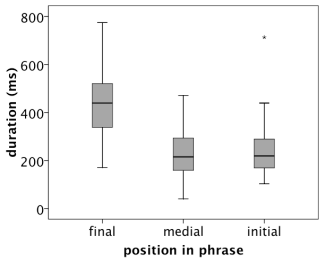
In summary, English ADS was characterized by phrase-boundary lengthening, with the greatest degree of lengthening occurring phrase-finally. In addition, lengthening is induced by the presence of a nuclear accent or the presence of a lexical stress when not coincident with an accent. These results reinforce findings for English ADS obtained through controlled readings (see [17]).

3.3 English CDS

As with ADS, the following variables were found to have a main effect on syllable duration: syllable structure (closed syllables longer) ($p<0.05$; $F(1,247)=11.43$); prosodic prominence ($F(3,247)=4.52$, $p<0.05$) and phrase position ($F(2, 247)=6.63$, $p<0.05$). As with ADS, word position did not have a main effect and there were no interactions.

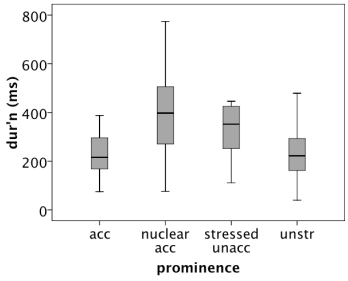
Bonferroni post hoc comparisons revealed phrase-final syllables to be longer than both phrase-medial syllables ($p<0.001$) and phrase-initial syllables (0.001). Unlike ADS, there was no difference between phrase-initial and phrase-medial syllable duration, i.e. no evidence of boundary-initial lengthening (see Figure 5).

Figure 5: mean syllable duration (in ms) for different phrase positions for English CDS



Bonferroni post hoc comparisons also revealed nuclear accented syllables to be longer than both unstressed syllables ($p<0.001$) and other accented syllables ($p<0.05$) (see Figure 5). No other comparisons were significant: of particular note, there was no difference between stressed and unstressed syllables (unless the stressed syllable was nuclear accented).

Figure 6: mean syllable duration (in ms) for different prominence levels for English CDS

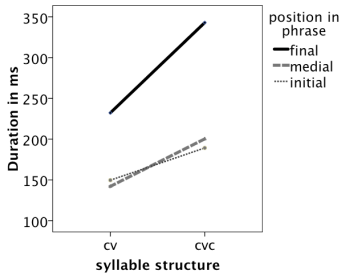


In summary, English CDS did not show the same degree of boundary lengthening as evident in ADS. Specifically, although there were clear phrase-final effects (and indeed effects were stronger), phrase-initial lengthening was not evident. In terms of lexical stress and accent marking, there was clear evidence of lengthening to differentiate nuclear accented syllables from other accented syllables and unstressed syllables. However, unlike ADS, there is no clear distinction between nuclear accented syllables and stressed unaccented syllables. Furthermore, away from the nuclear focus, there was no clear distinction between stressed and unstressed syllables.

3.4 Catalan ADS

The following variables had a main effect on syllable duration: syllable structure (closed syllables longer) ($F(1,306)=49.26$, $p<0.001$), prosodic prominence ($F(3,306)=12.92$, $p<0.001$) and phrase position ($F(2,306)=66.51$, $p<0.001$). Word position did not have a main effect, however, there was a significant interaction between prominence and word position ($F(1, 306)=2.52$, $p<0.05$). While nuclear accented syllables were longer than other types of syllables in all word positions, they were exceptionally longer in word-final position, for both monosyllables and polysyllables. There was also an interaction between syllable structure and phrase position ($F(1, 306)=9.73$, $p<0.001$). Although closed syllables were longer than open syllables for all positions, the difference was much larger for phrase-final position (see Figure 7)

Figure 7: mean syllable duration (in ms) for different syllable structures in different phrase positions for Catalan ADS



Bonferroni post hoc analyses showed phrase-final syllables to be longer than phrase-medial ($p<0.001$) and phrase-initial ($p<0.001$) (see Figure 8). Unlike English, however, there was no evidence of phrase-initial lengthening).

Figure 8: mean syllable duration (in ms) for different phrase positions for Catalan ADS

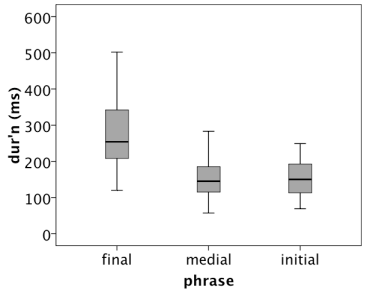
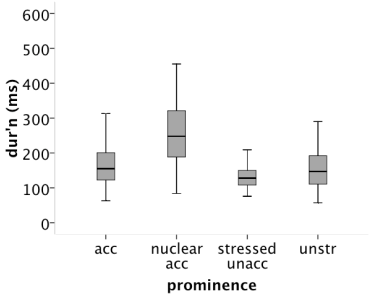


Figure 9: mean syllable duration (in ms) for different prominence levels for Catalan ADS



Bonferroni post hoc comparisons revealed nuclear accented syllables to be longer than other accented syllables ($p<0.001$), stressed unaccented syllables ($p<0.001$) and unstressed syllables ($p<0.001$). Other accented syllables were longer

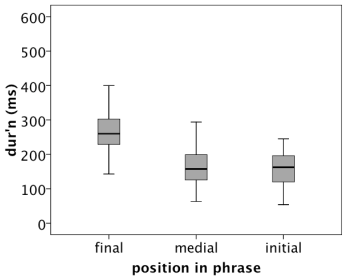
than stressed unaccented syllables ($p<0.001$), but not significantly different from unstressed syllables (see Figure 9). It appears, then, that a nuclear accent lengthens duration, but a lexical stress does not.

In summary, Catalan ADS was characterized by phrase-final lengthening in addition to lengthening induced by a nuclear accent, but not lexical stress alone. Again, these results reinforce findings for Catalan ADS obtained through controlled readings (see [17]).

3.5 Catalan CDS

As in ADS, position in the phrase proved to have a main effect on syllable duration ($F(2, 364)=10.8$, $p<0.001$) and Bonferroni post hoc comparisons revealed that phrase-final syllables were longer than both phrase-medial ($p<0.001$) and phrase-initial syllables (0.001) (see Figure 10). However, unlike ADS, no other variables proved to have a main effect. There were no interactions between variables.

Figure 10: mean syllable duration (in ms) for different phrase positions for Catalan CDS



To sum up, while Catalan CDS shows evidence of phrase-final lengthening, unlike ADS it shows no evidence of marking nuclear accents through lengthening.

4. Conclusions

This study reinforces the picture of prosodic lengthening patterns in English and Catalan ADS ([17]) and gives evidence for modification of these patterns in CDS. In both languages, CDS presents a less variegated system of prosodic marking than ADS. One result of this is that there is greater uniformity of syllable duration, which provides a probable explanation for the more even rhythm reported for English and Catalan CDS (see [6] [7]). The results also show the selective (and language-specific) enhancement of certain structures, at least via durational cues. In both languages, phrase-final lengthening is prioritized. In English, durational marking of the nuclear accented syllable is additionally prioritized. Other prosodic markers evident in ADS (initial boundary lengthening and unstressed syllable reduction in English; nuclear accented syllable marking in Catalan) are not clearly evident in CDS, at least in terms of duration. It remains to be investigated whether other phonetic cues pattern in this way. One possibility is that nuclear accented syllables in Catalan CDS rely more on exaggerated pitch movement and/or amplitude.

5. Acknowledgements

We would like to thank N. Argemí, A. Barberà, M. Jean Bell, A. Estrella, and F. Torres-Tamarit for recording the data in the three languages, and to N. Hilton for performing the segmentation and coding of the data. This research has been funded by two Batista i Roca research projects (Refs: 2007 PBR 29 and 2009 PBR 00018 respectively) awarded by the Generalitat de Catalunya, and by the project SG-51777 awarded by the British Academy.

6. References

[1] Albin, D. D., & Echols, C. H. (1996). Characteristics of stressed and word-final syllables in infant-directed speech: Implications for word-level segmentation. Infant Behavior and Development, 19, 401-418.

[2] Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20(1), 104-113.

[3] Garnica, O., (1977). Some prosodic and paralinguistic features of speech to young children. In Snow, C.E. and Ferguson, C.A., (Eds.), Talking to children: Language input and acquisition, (63-88), Cambridge, MA: CUP.

[4] Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental Psychology, 24(1), 14-20.

[5] Stern, D., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, *18*, 727–735.

[6] Payne, E.; Post, B.; Astruc, L.; Prieto, P.; Vanrell, M. (2009). "Rhythmic modification in Child Directed Speech" in O. Parker Jones & E. Payne (eds) *Oxford Working Papers in Linguistics, Philology and Phonetics,* 12: 123-144

[7] Payne, E., B. Post, L. Astruc, P. Prieto, & M. Vanrell (submitted). Rhythmic modification in CDS. in M. Russo (ed.), *Atti del convegno di Prosodia. Gli universali prosodoci: confronto e ricerche sulla modellizzazione ritmica e sulle tipologie ritmiche*. Roma: Aracne Biblioteca di Linguistica.

[8] Werker, J., & McLeod, P. (1989) Infant preference for both male and female infant-directed talk: a developmental study of attentional affective responsiveness *Canadian Journal of Psychology* 230–246.

[9] Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

[10] Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: initial representations for language learning. In F. Newmeyer (ed.) *Linguistics: The Cambridge survey Vol. 3*. Cambridge: CUP.

[11] Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*, 269–286.

[12] Kemler Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., and Cassidy, K.W. (1989). How the prosodic cues in motherese might assist language learning. Journal of Child Language, 16, 55-68.

[13] Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: evidence from 7-week-old infants. *Infant Behavior and Development*, *15*, 325–345.

[14] Shute, B., & Wheldall, K. (1989). Pitch alternations in British motherese: Some preliminary acoustic data. *Journal of Child Language*, *16*, 503–512.

[15] Papousek, M., & Hwang, S. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: comparison with registers of adult conversation and foreign language instruction. *Applied Linguistics*, *12*, 481–504.

[16] Kitamura, C., C. Thanavishuth, D. Burnham, S. Luksaneeyanawin (2002) Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language *Infant Behavior & Development* 24 (2002) 372–392.

[17] Prieto, P., M. Vanrell, L. Astruc, E. Payne & B. Post (*this conference*) Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English and Spanish.