PRE-SCHOOLERS USE HEAD GESTURES RATHER THAN DURATION OR PITCH RANGE TO SIGNAL NARROW FOCUS IN FRENCH

Núria Esteve-Gibert[1,2], Hélène Lœvenbruck[3], Marion Dohen[4], Mariapaola D'Imperio[5,2]

[1]Universitat Oberta de Catalunya, [2]Aix Marseille University, CNRS, LPL UMR 7309, Aix en Provence, France, [3]Univ. Grenoble Alpes, CNRS, LPNC, 38000 Grenoble, France [4]Univ. Grenoble Alpes, Grenoble INP, CNRS, GIPSA-lab, 38000 Grenoble, France [5]Rutgers University, USA
nesteve@uoc.edu, Helene.Loevenbruck@univ-grenoble-alpes.fr, marion.dohen@grenoble-inp.fr, mariapaola.dimperio@rutgers.edu

ABSTRACT

Previous research on the development of children's marking of new referents in speech has traditionally neglected one source of relevant information: visual cues to prosodic structure. In light of previous findings showing that prosody allies with non-referential body movements in the expression of information structure, we explore whether pre-schoolers mark focused information in the discourse gesturally and/or prosodically. A group of French-speaking pre-schoolers were audio-visually recorded while producing semi-spontaneous utterances in 3 focus conditions (broad focus; contrastive narrow focus; corrective narrow focus). The acoustic (duration and pitch range at the word level) and visual (head gestures) analyses showed a higher rate of head gesturing in the narrow focus conditions (corrective>contrastive>broad), but no effect of focus condition on word duration nor on pitch range values. These results indicate that French pre-schoolers use visual prosody to highlight new/contrastive referents in the discourse before developing the ability to use acoustic cues of prosody.

Keywords: narrow focus, children, head gestures, prosody, French intonation

1. INTRODUCTION

There is now considerable amount of evidence showing that prosodic features of speech are tightly coordinated with body movements, at the temporal level and also at the (semantic and pragmatic) meaning level (eg. [1], [2]). At the temporal level, prosodic landmarks are found to serve as anchoring points for body movements to align with specific speech locations. Bi-phasic body movements (like pointing gestures, manual beats, or head nods) all have a prominent phase (called 'stroke' if it is an interval or 'apex' if it is a specific point in time) that is found to be coupled with pitch-accented syllables [3]–[6]. Furthermore, prosodic edges (phrase boundaries) are also found to determine the temporal positioning of co-speech gestures [7]–[9]. In development, children acquire these temporal alignment patterns as soon as they are able to combine speech with gestures as a single meaning unit [10].

At the meaning level, adult speakers use both acoustic (prosodic) and visual (gestural) modalities to structure information in speech, to indicate sentence type, or to express emotional and epistemic meaning (see [2] for a review). This multimodal integration has also been observed in development. Before entering the lexical stage, young infants comprehend basic pragmatic meanings like request or assertion through prosody and hand gestures, while pre-schoolers process facial expressions and intonation cues as a marker of epistemic meanings before they are able to use lexical means to do so (eg. [11], [12]).

Despite previous research suggesting that children integrate prosody and body gestures as markers of pragmatic meaning, a pragmatic component has been understudied: information structure. Information structure refers to the marking of the informational status of discourse referents [13], [14], and prosody is one of the main strategies that can be used to signal if a referent is new or given in the discourse. In French, for instance, the initial and last syllable of focused words is expected to be lengthened because speakers insert a prosodic break before and after the focused element (and lengthening is a marker of phrase boundary marking [15]-[17]). However, recent findings suggest that speakers may also use body movements to mark new referents in the discourse, with head nods being one of the most frequent gestural means for this purpose ([18]–[21]).

In development, previous research on children's ability to distinguish between new and given discourse referents has exclusively focused on the prosodic modality. Current findings suggest an early use of acoustic marking (by means of pausing, for instance, [22]) and only a later mastering of adult-like prosodic-phonological patterns (i.e. adult-like pitch accent type and placement; see [23] for a review).

Studies of young children's ability to mark information structure with gestures are scarce. [24] reported that Australian 6-year-olds can use hand beat

gestures aligned with lexical content words to emphasize discourse referents, and that these gestures are not always accompanied by pitch accents. It is still unclear, however, whether children can use gestural strategies to focus lexical items at earlier stages in development (i.e. pre-schoolers), and whether they are able to combine gestural and prosodic marking.

The present study aims at answering these questions. Given that prosody and gesture go hand in hand in the development of other pragmatic components, this could also be the case for the expression of information status. We expected children to gesturally and prosodically mark focused words, and that this marking would be more frequent for corrective than contrastive focus. We specifically examined head gestures (nods and tilts) which have been shown to be frequently used in adults.

2. METHODS

2.1. Participants

A total of 24 French-speaking pre-school aged children participated in our study (mean age: 60 months; age range: 50-67 months; 8 boys). Two additional children were tested but excluded from the final sample (one due to colour-blindness issues that could affect the results of the task, and the other one due to fussiness).

2.2. Materials

Children produced a total of 60 sentences containing Noun Phrases that had the following shape: Article + disyllabic Noun + disyllabic Adjective (eg. *Prends la valise violette* 'Take the purple suitcase'). Two variables were manipulated and fully crossed: the type of information status (broad focus, contrastive narrow focus, or corrective narrow focus), and the position of the new referent within the sentence (either at object noun or at the phrase-final adjective position). This resulted in 5 experimental conditions (N=12 sentences per condition), summarized in Table 1. All nouns and adjectives were elicited in each experimental condition to rule out potential effects of segmental and syllabic structure.

The visual display consisted of a picture of a girl at the bottom left corner of the screen, who was the character with whom children were asked to interact. At the centre of the screen, there was a big bag containing different objects depending on the experimental condition. In the broad focus condition, only one object was shown in the bag; in the contrastive and corrective focus conditions, two or more items were shown, either differing in colour (if the new referent to be emphasized was the adjective) or in nature (if the new referent to be emphasized was the noun). At the top right corner of the screen, an event was depicted (eg. closed eyes to be opened), together with one of the objects from inside the bag (see section 2.3 for further details of the game).

Table 1: Example of sentences in each experimental condition. Capital letters indicate contrastive focus, bold letters indicating corrective focus.

Type	*New referent*	*Example for each condition*
Broad	None	*Prends la valise violette* Take the suitcase purple
Contrastive	Noun (non-phrase-final)	*Prends la VALISE violette* Take the SUITCASE purple
	Adjective (phrase-final)	*Prends la valise VIOLETTE* Take the suitcase PURPLE
Corrective	Noun (non-phrase-final)	*Prends la* ***VALISE*** *violette* Take the SUITCASE purple
	Adjective (phrase-final)	*Prends la valise* ***VIOLETTE*** Take the suitcase PURPLE

2.3. Procedure

The game went as follows: children were told that in order for Claire (the girl's name) to launch the events (eg. opening the eyes), she had to pick the right object from inside the bag (the target object was shown next to the image depicting the event). Since Claire could not see the object herself, children had to give her instructions about which target object to pick. When there was a single object inside the bag, children were expected to produce a sentence in broad focus condition; when there were several objects inside the bag, the contrastive focus condition was expected (emphasizing either the noun if the objects were equal in colour but differing in shape, or emphasizing the adjective if they were equal in shape but differing in colour). If Claire took the wrong object and children were induced to repeat the instruction, corrective (i.e. stronger) focus was expected on the target word (emphasizing the noun or the adjective accordingly).

Children were audio-visually recorded: a camera was placed in front of them, and a microphone was placed next to the child.

2.4. Analysis

Children's productions were coded both acoustically and gesturally. The ELAN annotation tool was used for the head gesture coding [25], with several tiers: word by word orthographic transcription

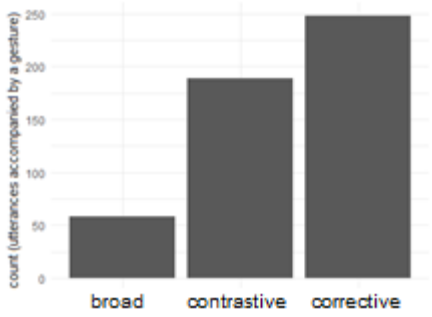
of the children's speech (Tier 1), experimental condition (Tier 2), presence or absence of a head gesture (yes/no; Tier 3), target word accompanying the prominent phase of the head movement (none/noun/adjective; Tier 4), and type of head movement (none / nod / tilt / chin pointing/ eyebrow raising; Tier 5). We expected children to produce a higher amount of head gestures in the corrective than in the contrastive than in the broad focus conditions, and to produce these gestures on the new referent within the utterance (either noun of adjective).

PRAAT was used for the acoustic coding [26], annotating the orthographic transcription of the utterance (Tier 2), its word by word (Tier 2) and syllable by syllable (Tier 3) segmentation, the target experimental condition (Tier 4), and the F0min and F0max values within the target noun and target adjective (Tier 5). Word and syllable segmentations were automatically performed using SPPAS [27] and later checked manually. Only full target utterances (including both Noun and Adjective) were prosodically analysed, therefore excluding instances of only Noun or only Adjective, N=186 out of a total of 1,235). Following adult studies on French prosodic marking of focus, we expected children to produce longer word duration and wider F0 range values for both new referents in the discourse, with even greater values under corrective focus (corrective > contrastive > broad). Finally, we expected children to produce longer syllables at the end of focused referents, as a sign of Accentual Phrase break insertion, a common prosodic strategy to mark focus in French.

3. RESULTS

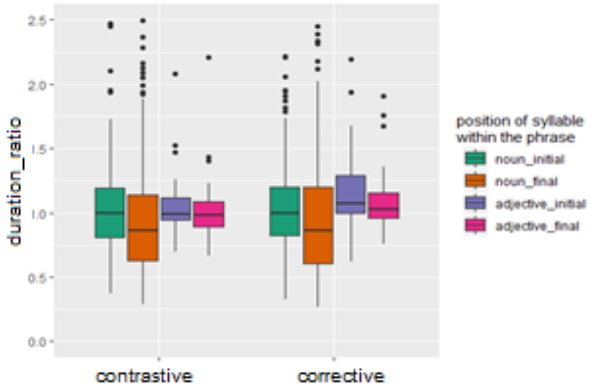
Figure 1 shows that, as expected, children produced a higher proportion of head gestures in the corrective compared to the contrastive condition and in the contrastive compared to the broad focus condition (all gesture types collapsed). We performed Linear Mixed Models analyses [28] in order to statistically evaluate the results. A first glmer model with presence/absence of head gesture as dependent variable, experimental condition as fixed factor, and participant and item as random factors, confirmed these significant differences (broad vs. contrastive: β=.479, z=1.91, p=.05; broad vs. corrective: β=1.072, z=4.27, p<.001; contrastive vs. corrective: β=.593, z=3.02, p<.01). The majority (73%; N=363) of head gestures were correctly aligned (i.e. the head gesture accompanied the focused item). A subsequent *glmer* model with correct/incorrect alignment as dependent variable, and focus type (contrastive vs. corrective) and focus position (on noun vs. on adjective), and participant and item as random effects, further revealed a main effect of focus position (p<.001, with more incorrect alignment when the focalised element was the non-phrase-final noun), but no main effect of focus type or any interaction.

Figure 1. Amount of utterances accompanied by a head gesture in each experimental condition (independently of the position of the gesture within the utterance).



As for the prosodic analysis, a first 'lmer' model examined whether word duration (calculated as a ratio between the narrow focus conditions and the broad focus condition used as a baseline) was influenced by focus condition and by position of the target word within the phrase (participant and item as random factors). Results revealed no main effect of focus condition (p=.13), no main effect of position of the target word within the phrase (p=.94), nor an interaction between these two (p=.37).

Figure 2. Syllable duration values across conditions and their position within the phrase. A ratio of 1 means equal length compared to the broad focus condition.

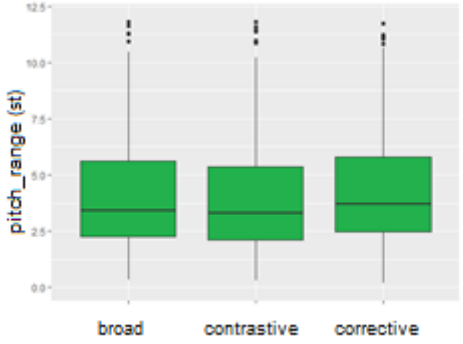


A second model explored whether syllable duration (calculated as a ratio between the narrow focus conditions and the broad focus condition as a baseline) was influenced by focus condition and by position of the syllable within the phrase (first syllable of the noun/non-phrase-final element, last syllable of the noun/non-phrase-final element, first syllable of the adjective or phrase-final element, last syllable of the adjective/phrase-final element), with participants and items as random factors. Results

showed no main effect of focus condition (p=.06), nor of position of syllable within the phrase (p=.67), and no interaction (p=.87) (see Figure 2).

A third model investigated whether children used F0 excursion values (measured as the difference between F0max and F0min in each word, transformed in semitones) to distinguish between focus conditions (participant and item treated as random factors again). Results showed no main effect of focus condition on the F0 range values of the target elements (p=.66) (see Figure 3). An additional analysis was carried out to explore whether children compressed the F0 range of non-focused elements instead of expanding the F0 range of focused elements. The results revealed that the status of the element (focused/non-focused) did not interact with focus condition (p=.07).

Figure 3. Pitch range values (in semitones) across the three experimental conditions.



4. DISCUSSION

Previous studies on children's ability to use prosodic cues to emphasize new elements in the discourse had suggested a late development of adult-like prosodic patterns (pitch accent type and pitch accent placement). As a consequence, it has been proposed that children's mastery of prosodic focus is only manifest at around 7-8 years of age (see [23] for an overview). However, one communicative modality has been omitted in this previous research: the visual marking of informational focus. Previous findings on the gesture domain reveal that speakers highlight new discourse referents not only through acoustic strategies but also by producing head gestures (i.e. head nods) and facial expressions (i.e. eyebrow raising) (see [2] for a review). The present study aimed at exploring whether this multimodal marking of focus is already exploited by pre-school aged children, still developing language.

The results of our study suggest that French pre-schoolers do not seem to use acoustic-prosodic marking to distinguish between old and new information in the discourse, or between new information and corrected information in the discourse. Neither pitch range nor word and syllable duration values varied across conditions. In the particular case of syllable duration values, children did not lengthen syllables at the right-edge nor at the left-edge of focused elements. This contrasts with what French-speaking adults do: the focused element would be phrased into an Accentual Phrase by lengthening the AP-initial or AP-final syllables and increasing their pitch excursions ([15]-[17]). It would be expected that the AP-final lengthening would be even clearer in utterance-final position (in our utterances, the adjective-final syllable given the typical word order in French), but syllable position was a non-significant factor in our data.

Instead, pre-school children do use head gestures for that purpose. Children produced a higher rate of head gestures accompanying utterances in a corrective narrow focus condition than in contrastive narrow focus or broad focus conditions. It appears that French pre-schoolers are able to highlight new and corrected referents in the discourse by means of visual communicative strategies before they are able to use typical acoustic prosodic strategies.

A close inspection of the position of the head gesture within the utterance has revealed that it is easier for children to mark focalised elements with a head gesture when these are in a phrase-final position (as we found more cases incorrect gesture-speech alignment when the focused element was the non-phrase-final noun). These are interesting findings given that adults systematically align acoustically prominent syllables with the prominent phrases of the gesture movement [3]–[9]. We assume that the pre-schoolers' prosody-gesture misalignment could be due to their failure to acoustically highlight the target syllables, and/or also due to their inability to rephrase the focused elements into APs. Thus the utterance-final element became their default anchoring position for gestural apex alignment.

Our results are in line with previous studies on the development of pragmatic meanings that take into account multimodal cues. These studies have in fact observed that children first comprehend pragmatic (epistemic and ironic) meanings through visual cues than through prosodic and lexical cues (see [1] for a review). Body gestures might therefore scaffold the acquisition of linguistic marking of informational structure just like they scaffold the acquisition of other complex pragmatic meanings.

Further cross-linguistic studies that take into account all communicative modalities are needed to investigate whether visual cues precede acoustic, cues independently of the linguistic structure of the target language, or whether children foster one modality over the other as a way to overcome incompletely mastered grammatical complexity.

5. CONCLUSION

An audio-visual production study on a group of French pre-schoolers has shown that while co-speech, head gestures, are employed to mark contrastive and corrective referents, typical acoustic/prosodic means are not yet acquired. Specifically, a differential increase in word or syllable duration and/or F0 range on focused items was not found in our data. Head gestures, on the other hand, tended to be correctly aligned with the focused item (this being more frequent when the focused item was the phrase-final adjective). These results therefore suggest that gestural marking of information structure might be acquired earlier than acoustic-prosodic means.

6. REFERENCES

[1] N. Esteve-Gibert and B. Guellaï, "Prosody in the auditory and visual domains: A developmental perspective," *Front. Psychol.*, vol. 9, pp. 1–10, 2018.

[2] P. Wagner, Z. Malisz, and S. Kopp, "Gesture and speech in interaction: An overview," *Speech Commun.*, vol. 57, pp. 209–232, 2014.

[3] A. Kendon, "Gesticulation and speech: two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, M. R. Key, Ed. The Hague: Mouton, 1980, pp. 207–227.

[4] T. Leonard and F. Cummins, "The temporal relation between beat gestures and speech," *Lang. Cogn. Process.*, vol. 26, no. 10, pp. 1457–1471, 2011.

[5] E. Krahmer and M. Swerts, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *J. Mem. Lang.*, vol. 57, no. 3, pp. 396–414, 2007.

[6] H. L. Rusiewicz, S. Shaiman, J. M. Iverson, and N. Szuminsky, "Effects of Prosody and Position on the Timing of Deictic Gestures," *J Speech Lang Hear Res.* vol. 56, no. 2, pp. 458-470, 2013.

[7] N. Esteve-Gibert, J. Borràs-Comes, E. Asor, M. Swerts, and P. Prieto, "The timing of head movements: The role of prosodic heads and edges," *J. Acoust. Soc. Am.*, vol. 141, no.6,pp. 4727–39, 2017.

[8] J. Krivokapic, M. K. Tiede, and M. E. Tyrone, "A Kinematic Analysis of Prosodic Structure in Speech and Manual Gestures," in *Proceedings of the 18th International Congress of Phonetic Sciences*, 2015.

[9] S. Shattuck-hufnagel, P. L. Ren, E. Tauscher, and C. Ma, "Are torso movements during speech timed with intonational phrases?," in *Proceedings of the Speech Prosody*, 2010.

[10] N. Esteve-Gibert and P. Prieto, "Infants temporally coordinate gesture-speech combinations before they produce their first words," *Speech Commun.*, vol. 57, pp. 301–316, 2014.

[11] I. Hübscher, N. Esteve-Gibert, A. Igualada, and P. Prieto, "Intonation and gesture as bootstrapping devices in speaker uncertainty," *First Lang.*, vol. 37, no. 1, pp. 24–41, 2017.

[12] I. Hübscher, L. Wagner, and P. Prieto, "Young children's sensitivity to polite stance expressed through audiovisual prosody in requests," *Proc. Int. Conf. Speech Prosody*, pp. 897–901, 2016.

[13] B. Daniel, "Semantics , Intonation and Information Structure Focus – Background Preliminaries on Focus Realization," pp. 1–36, 2005.

[14] E. Vallduví, *The informational component*. New York, New York, USA: Garland, 1991.

[15] A. Michelas and M. D'Imperio, "Prosodic boundary strength guides syntactic parsing of French utterances.," *Lab. Phonol.*, vol. 6, no. 1, pp. 119–146, 2015.

[16] C. Féry, "Focus and phrasing in French," *Audiatur Vox Sapientiae- A Festschrift Arnim von Stechow*, no. October 1998, pp. 153–181, 2001.

[17] M. Dohen and H. Lœvenbruck, "Pre-focal Rephrasing, Focal Enhancement and Post-focal Deaccentuation in French," *International Conference on Spoken Language Processing*, vol. 1, pp. 785-788, 2004.

[18] G. Ambranzaitis and D. House, "Multimodal prominences: Exploring the patterning and usage of focal pitch accents , head beats and eyebrow beats in Swedish television news readings," *Speech Commun.*

[19] P. Barkhuysen, E. Krahmer, and M. Swerts, "The interplay between the auditory and visual modality for end-of-utterance detection.," *J. Acoust. Soc. Am.*, vol. 123, pp. 354–365, 2008.

[20] U. Hadar, T. J. Steiner, E. C. Grant, and F. C. Rose, "Head movement correlates of juncture and stress at sentence level," *Lang. Speech*, vol. 26, no. 2, 1983.

[21] C. T. Ishi, H. Ishiguro, and N. Hagita, "Analysis of relationship between head motion events and speech in dialogue conversations," *Speech Commun.*, vol. 57, pp. 233–243, 2014.

[22] A. S. H. Romoren and A. Chen, "Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch," *Lang. Speech*, vol. 58, pp. 8–23, 2015.

[23] A. Chen, "Get the focus right: acquisition of prosodic focus-marking across languages," in *The Develompent of Prosody in First Language Acquistion*, P. Prieto and N. Esteve-Gibert, Eds. John Benjamins, pp. 295–313, 2018.

[24] M. Mathew, I. Yuen, and K. Demuth, "Talking to the beat: Six-year-olds' use of stroke-defined non-referential gestures," *First Lang.*, p. 0142723717734949, 2017.

[25] H. Lausberg and H. Sloetjes, "Coding gestural behavior with the NEUROGES-ELAN system.," *Behav. Res. Methods, Instruments, Comput.*, vol. 41, no. 3, pp. 841–849, 2009.

[26] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2012.

[27] B. Bigi, "SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech," *Phonetician - Int. Soc. Phonetic Sci.*, vol. 111–112, pp. 54–69, 2015.

[28] T. F. Jaeger, "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," *J. Mem. Lang.*, vol.59, pp.434-46, 2008.