Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

The prosodic and gestural marking of the information status of referents in children's narrative speech: A longitudinal study

Sara Muñoz-Coego[1], *Júlia Florit-Pons*[1], *Patrick Louis Rohrer*[1,2], *Ingrid Vilà-Giménez* [1,3], *Pilar Prieto*[4,1]

[1]Dept. of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona
[2] Nantes Université, Laboratoire de Linguistique de Nantes - UMR 6310, Nantes
[3]Dept. of Subject-Specific Education, Universitat de Girona, Girona
[4]Institució Catalana de Recerca i Estudis Avançats, Barcelona

sara.munoz@upf.edu, julia.florit@upf.edu, patrick.rohrer@upf.edu, ingrid.vila@udg.edu, pilar.prieto@upf.edu

Abstract

Developmental studies have claimed that 5- to 7-year-old children use prosodic prominence to mark the status of referential expressions in discourse, where new and accessible referents tend to be prosodically more prominent than given referents [1], [2]. Studies on the gestural marking of information structure have shown that gestures tend to co-occur more frequently with new and focused referents in adult and child speech [3], [4]. Although prosody and gesture are tightly integrated in speech [5], to our knowledge no previous approach has jointly looked at the development of these multimodal cues as markers of the information status of referents. The present study investigates how Catalan-speaking children use prosody and manual gestures to signal the information status of referents in narrative speech. A longitudinal database was used containing elicited narratives by 83 children at two points in development (5-6 and 7-9 years old). Results showed that new referents (in contrast with accessible and given referents) were signaled through the use of non-referential gestures, specially at the ages of 7-9, and through nuclear accentuation already at the ages of 5-6. Results also revealed that all types of referents were pitch accented at both points in development, regardless of their information status.

Index Terms: information status of referents, multimodality, non-referential gestures, prosody

1. Introduction

Human language constitutes a multimodal system; that is, speakers integrate both speech and gesture to communicate meaning [5]. Prosody and gesture encode a wide range of pragmatic functions, among them the structuring of information in discourse. Prosodic and gestural prominence, which tend to occur simultaneously [5], have been reported to encode novel and contrastive information in discourse both in adult speech [6], [7], [8], [9] and in child speech [1], [3], [10].

1.1. The gestural and prosodic marking of information status in adult speech

The term Information Structure (IS) is used to refer to the ways in which information is packaged in discourse with the aim of updating the common knowledge (i.e., common ground) of interlocutors [11]. Out of the three IS levels (i.e., Focus/Background, Topic/Comment, and Information Status of Referents) distinguished in [11] and [12], in the present study we will focus on the level of Information Status of Referents (henceforth, information status). Referents are discourse entities introduced in speech through a noun phrase or a prepositional phrase. By entity here we mean individuals, objects, places, times, events, situations, or propositions, whether real or imagined, which can be retrieved by anaphoric expressions [10]. According to their retrievability, referents can be of three types, from the most to the least cognitively active for the interlocutor: *given*, *accessible*, or *new*. Given referents are cognitively active as they have been explicitly introduced previously in discourse. Accessible referents are newly introduced entities in discourse, which are part of the speakers' common ground due to being inferable from their (situational or linguistic) context, or from cultural and general knowledge. New referents are also newly introduced entities in discourse, but are less cognitively active than accessible referents, as they cannot be inferred from world or contextual knowledge.

Evidence from Germanic and Slavic languages has shown that given referents tend to be associated with less prominent pitch accents, and to be deaccented more often than new referents [6], [13], [14], [15]. Studies on Romance languages [16], [17], [18] claim that nuclear accent placement may have an important role in the marking of IS categories too. Regarding gesture production, gestures have been found to mark information status as well. Authors generally coincide on reporting that co-speech gestures occur more frequently with new and accessible referents in discourse [9], [19]. Research has therefore shown that adult speakers tend to make use of prosodic prominence and gesture to signal information status, but by analyzing the two components separately. As far as we know, only one study [20] has looked simultaneously at prosody, gesture, and information status. However, the authors did not consider the role of prosody in marking information status, looking exclusively at how prosody associates with gesture, and how gesture associates with information status. The main goal of our study is precisely to analyze the role of prosody and co-speech manual gesture in information status marking, but from the point of view of its acquisition by children and its development.

1.2. The development of information status marking

Developmental research has shown that children start using prosodic and gestural strategies to mark information status at around the age of 5 [1], [3], [21] (see [22] for a review). On the one hand, in relation with prosodic marking, German-speaking children have been reported to deaccent given referents and accent new and accessible in an adult-like manner between the

10.21437/SpeechProsody.2022-82

ages of 5-7 [1]. Other studies have found patterns of deaccentuation of given referential expressions in German children's spontaneous interactions already at the age of 3;5 [2]. On the other hand, regarding gestural marking, previous studies have reported that head nods associate more strongly with focused information in 5-year-old children's speech [e.g., 3]. To our knowledge, only one study has analyzed specifically the relationship between manual gestures and information status marking in children's speech [21]. Using the same narrative corpora that was used for the present study, the authors found that non-referential gestures, i.e., gestures that do not visually represent the semantic content of speech (see section 2.3.1), associate significantly more with new referents in discourse than given and accessible at the ages of 5-6, a tendency that becomes more pronounced at 7-9. The present study is an extension of [21]. While [21] investigated specifically the gestural perspective (i.e., when gesture is produced, which information status categories are being marked), the present study takes a more holistic point of view, in that it looks at which categories of information status are being marked in gesture as well as in prosody.

1.3. Motivation of the study

Previous literature has shown that narrative abilities are a good measure of children's language skills [23]. Because of this, we consider that assessing the first stages of narrative speech development allows us to analyze how children start acquiring the multimodal marking of information status in discourse. Developmental studies on the acquisition of prosodic means for information status marking have mainly looked at Germanic languages, therefore, it is still unclear how the prosody of Romance languages is acquired for this purpose. Moreover, provided that nuclear accentuation is reported to be essential in IS marking in Romance languages [16], [17], [18], we think it is important that our analysis accounts not only for accentuation and deaccentuation patterns, but also for the nuclear/prenuclear contrast. Moreover, to better assess the acquisition of multimodal cues for information status marking, the current study will assess how children use prosodic and gestural marking in a truly comparable way. Therefore, in comparison with [21], the present analysis will analyze the gestural and prosodic marking of information status by taking as a point of departure the behavior of all referents appearing in the narratives (and not just the ones that are associated with gestures, as in [21]). Finally, as in [21], we believe it is necessary to consider the contribution of gesture referentiality when it comes to the marking of IS, given that non-referential gestures have been said to be special markers of IS categories such as focus [5], [24]. Moreover, we believe that the role of non-referential gestures as markers of IS may be a factor in explaining why they benefit narrative recall and comprehension [25], [26] and narrative performance [27], [28]. All in all, we believe that non-referential gestures could work jointly with prosody in the framing of narrative discourse from early on in terms of IS marking (and more specifically, the marking of information status). Assessing the prosodic and gestural marking of information status will thus allow us to have a more complete view on the acquisition of children's multimodal language.

1.4. Research questions and hypotheses

The current study analyzes the development of the children's multimodal marking of information status in narrative speech through the use of manual co-speech gestures and prosodic prominence (in terms of presence/absence of pitch accentuation and nuclear/prenuclear accentuation). For this reason, a longitudinal perspective based on a corpus of children's narrative speech is adopted. Two main research questions are addressed: 1) Do children use both (referential and non-referential) gestures and prosodic prominence (presence/absence of accentuation, and prenuclear/nuclear accentuation) to mark the status of referents in discourse at both times in development (5-6 and 7-9 years of age)? and 2) How does the gestural and prosodic marking of information status evolve over the developmental period from 5-6 to 7-9 years old?

2. Methods

To address our research questions, we followed a longitudinal approach. We analyzed the narrative retellings contained in the *Audiovisual corpus of Catalan children's narrative speech development* [29].

2.1. Participants

The *Audiovisual corpus of Catalan children's narrative speech development* contains narrative retellings from 83 children (43 girls and 40 boys), who were video recorded when they were between 5 and 6 years of age (*M* age = 5.9; *SD* = 0.55; Time 1) and, two years later, when they were between 7 and 9 (*M* age = 7.98; *SD* = 0.66; Time 2).

Participants, who were Catalan dominant (*M* dominance = 90.4%; *SD* = 9.5) Spanish-Catalan bilinguals with no prior history of language disorders. For more details, see [29].

2.2. Materials and procedure

The 331 narratives were collected through a narrative elicitation task. Participants were asked to watch two wordless cartoons (*Die Sendung mit der Maus*, approximate length 41-50*s*). They first watched a cartoon and then were asked to stand up and retell the story to the experimenter, who pretended not to know what the story was about. After that, the same process was repeated with a second story. At Time 2 the same procedure was repeated.

2.3. Data coding

Children's narratives were annotated for information status, gesture referentiality, and speech prosody. Annotations were carried out independently of each other to avoid any bias from the different domains. Annotators first orthographically transcribed the narratives in ELAN [30]. Then, gestural annotation was carried out in ELAN (see subsection 2.3.1.). The annotation of referent status and prosody was then carried out in Praat [31] (see subsections 2.3.2. and 2.3.3. respectively). Once all the annotations were completed, the data from Praat was imported into ELAN and the full dataset was then exported for statistical analyses.

2.3.1. Gesture referentiality

For the annotation of co-speech gestures, we followed an adapted version of McNeill's [5], [32] standard classification of gestures, and [33] for the definition of non-referential gestures. For the current study, we distinguished between *referential* and *non-referential* gestures. Referential gestures encode a lexico-semantic meaning as they visually display actions or properties or spatial locations of entities in speech via iconicity, metaphoricity, or deixis. Non-referential gestures, in turn, do

not represent any semantic content of speech, and have been traditionally described as encoding rhythmic information as well as discursive and pragmatic meanings.

2.3.2. *Information status*

For the annotation of information status, we followed the annotation scheme proposed by [12]. Only noun phrases and prepositional phrases which were referential expressions (i.e., referred to some entity) were considered as referents. Adverbial phrases, even if they contained a noun, were not considered referents. Noun phrases which were part of idiomatic expressions along with expressions that have a fixed use (e.g., *once upon a time*) were left out from the coding. After identifying discourse referents, they were annotated as *given*, *accessible*, or *new* depending on their degree of retrievability for the listener. The coding was carried out on the basis of the text transcription.

2.3.3. *Prosody*

Before coding prosody, to ensure that their prosodic quality was sufficient for further annotation, narratives were evaluated using an adapted version of the Comprehensive Oral Reading Fluency Scale (CORFS) [34]. This rubric focused on two categories, *Expression/Volume* and *Smoothness*, considered on a 1-4 Likert scale. Narratives that received average scores of 1.5 or less were excluded from further prosodic annotations (see section 2.3.4). For the annotation of speech prosody, we followed the Cat_ToBI labeling system [35]. First, accented syllables were identified, and further pitch accent types associated with them. Both were annotated according to the coder's perception of prominence and of variations in pitch. Second, for annotating nuclear accent placement, two categories were coded, e.g., *intonational phrase* (i.e., a major prosodic group, annotated with a "%"), and *intermediate phrase* (i.e., a minor prosodic group less strongly marked than the intonational phrase, annotated with a "-"). Disfluencies (filled pauses, self-corrections, repetitions, and interruptions, as per [36]), non-speech events (e.g., background noises) and words which were unintelligible or inaudible were annotated in a miscellaneous tier, and subsequently removed from analysis.

2.3.4. *Statistical analyses*

Before the statistical analysis, a total of 9 narratives were discarded: 4 of them were eliminated due to insufficient prosodic quality, and another 5 were discarded because the child did not remember the story at all, and thus was unable to produce a narrative. As a consequence, the final sample consisted of 157 narratives at Time 1 and 166 at Time 2, totaling 323 narratives. Within each of these narratives, referents which were clitic pronouns were excluded from the statistical models, as clitics are considered unaccented words per definition and depend on another word's prosodic features.

In order to simplify the interpretation and explanation of the results, we run three separate Linear Mixed Effects Models (LME) using the *lme4* package in R [37]. The first one assessed the interaction between gesture referentiality and information status; The second one assessed the relationship between pitch accentuation and information status, and a third one assessed the relationship of nuclear accentuation with information status. The dependent variable in all three models was the number of referents produced per participant. In the first model, the fixed factors were Time (2 levels: Time 1/Time 2), Information Status (3 levels: Given/ Accessible/New), and Gesture Referentiality (2 levels: Referential/Non-referential), as well as their three-way interaction. In the second model, the fixed factors were Time, Information Status, and Accentuation (2 levels: Accented/Deaccented), as well as their three-way interaction. In the third model, the fixed factors included Time, Information Status, and Nuclear Accentuation (2 levels: Prenuclear/Nuclear), as well as their three-way interaction. To determine the random effects structure, we tested multiple possible random effects structures and chose the one that best described the data and that did not show any convergence issues. For this, we used the "compare performance" function from the *performance* package [38]. While in the first model the random effects structure included by-Participant varying intercepts and by-Participant varying slopes for Time and Gesture Referentiality, in the second and the third models, the random effects structure included random slopes and intercepts for Participant.

3. Results

The first LME analysis showed a significant main effect of Time ($\chi^2(1) = 34.135$, $p < .001$), a significant main effect of Information Status ($\chi^2(2) = 269.961$, $p < .001$), a significant main effect of Gesture Referentiality ($\chi^2(2) = 432.637$, $p < .001$), and a significant 3-way interaction ($\chi^2(4) = 30.193$, $p < .001$). Post-hoc pairwise contrasts showed that all referent types occurred significantly more often with no gesture than with referential and non-referential gestures at both periods in development. However, while there was a significant increase in the number of given referents marked with no gesture from Time 1 to Time 2 ($d = 1.34$, $p < .001$), the number of new and accessible referents marked with no gesture did not increase significantly. When looking at referents which were gesturally

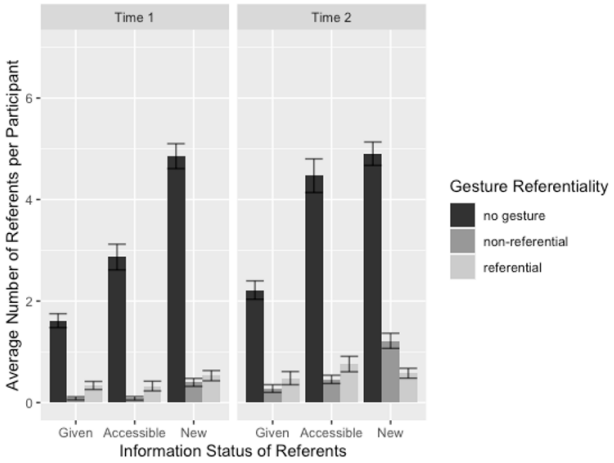

Figure 1. Graph representing the average number of gestures marking referents per participant at Time 1 and Time 2.

marked, the number of new referents marked with non-referential gestures significantly increased from Time 1 to Time 2 ($d = 0.685$, $p < .001$). Results also showed that, at Time 2, new referents were marked significantly more by non-referential gestures than by referential gestures ($d = 0.534$, $p = .037$). Furthermore, the number of new referents marked with a non-referential gesture was significantly higher than the number of given and accessible referents marked with non-referential gestures ($d = 0.101$, $p = .002$; $d = 0.786$, $p < .001$, respectively).

The second LME analysis showed a significant main effect of Time ($\chi^2(1) = 58.916$, $p < .001$), a significant main effect of Information Status ($\chi^2(2) = 268.411$, $p < .001$), a significant main effect of Accentuation ($\chi^2(1) = 2069.147$, $p < .001$), and a significant 3-way interaction (Time x Accentuation x Information Status: $\chi^2(2) = 8.511$, $p = .014$). Post-hoc pairwise contrasts showed that all referents at Time 1 and at Time 2 were significantly more accented than deaccented (*new*: $d = 2.709$, $p < .001$; *accessible*: $d = 0.998$, $p < .001$; *given*: $d = 1.358$, $p < .001$; see Fig. 1). Contrasting from the perspective of Time, while there were significantly more referents in general, there were no significant differences between accentuation patterns.

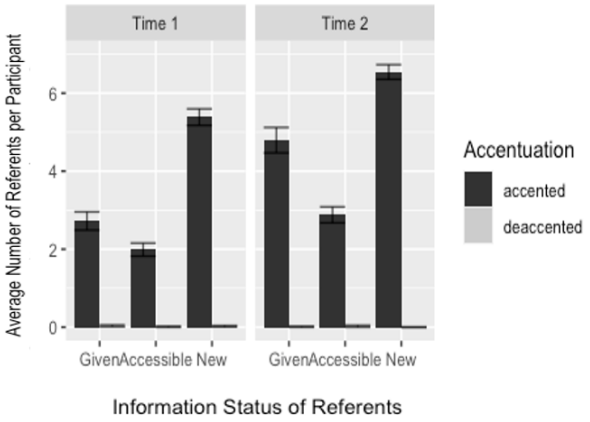

Figure 2. Graph representing the average number of accented and deaccented referents per participant at Time 1 and Time 2.

The last LME analysis showed a significant main effect of Time ($\chi^2(1) = 63.153$, $p < .001$), a significant main effect of Information Status ($\chi^2(2) = 294.359$, $p < .001$), a significant main effect of Nuclear Accentuation ($\chi^2(1) = 700.048$, $p < .001$), and a significant 3-way interaction (Time x Information status x Nuclear Accentuation: $\chi^2(2) = 6.858$, $p = .032$).

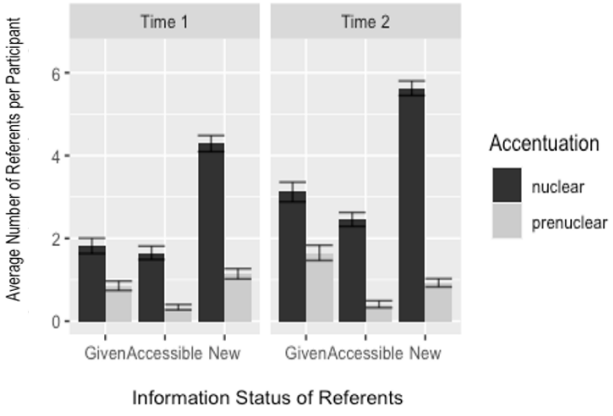

Figure 3. Graph representing the average number of prenuclear and nuclear referents per participant at Time 1 and Time 2.

Post-hoc pairwise contrasts revealed that, both at Time 1 and at Time 2, given, accessible, and new referents were produced significantly more with nuclear accentuation than with prenuclear accentuation (at T1, new: $d = 1.64$, $p < .001$; accessible: $d = 0.685$, $p < .001$; given: $d = 0.503$, $p < .001$; at T2, new: $d = 1.64$, $p < .001$; accessible: $d = 1.07$, $p < .001$; given: $d = 0.767$, $p < .001$, see Figure 2). At both times in development, the number of nuclear new referents was significantly higher than the number of nuclear given and accessible referents (at T1, given: $d = 1.29$, $p < .001$; accessible: $d = 1.38$, $p < .001$; at T2, given: $d = 1,31$, $p < .001$; accessible: $d = 1.67$, $p < .001$). Interestingly, given referents were marked with a prenuclear accent significantly more at Time 2 than at Time 1 ($d = 0.415$, $p = .004$), while no such significance was found for new and accessible referents.

4. Discussion and Conclusions

The current study presents the results of a longitudinal investigation that assesses the development of the prosodic (through patterns of accentuation and nuclear accent placement) and gestural marking of referent status at two time points in development. Our two main objectives were (1) to investigate if children at these two points in development use both prosodic prominence (presence/absence of accentuation, and prenuclear/nuclear accentuation) and gesture referentiality (referential/non-referential) to mark the information status of referents in narrative discourse, and (2) to evaluate how the multimodal marking of information status evolves over the developmental period from 5-6 to 7-9 years old.

First, regarding the role of co-speech gesture, our results showed that children at both points in development introduced given, accessible, and new referents in discourse significantly more often without gesture than with gesture. While not significant, a visual inspection of the data shows a clear tendency for gestures to mark new information at Time 1. At Time 2, the number of new referents marked with non-referential gestures is significantly higher than the number of given and accessible referents marked with non-referential gestures. This finding provides empirical support for claims highlighting the role of non-referential gestures in the marking of new information [5]. Our results further showed that new referents were marked significantly more with non-referential than referential gestures at the ages of 7-9, and that the number of non-referential gestures marking new referents significantly increased from the ages of 4-5 to 7-9. This supports previous findings that point towards the existence of a strong connection between non-referential gestures and the structuring of information in discourse [25], [27]. Second, results revealed that children at both points in development tend to accent all three types of referents. These results are not in tune with those in previous research which described patterns of deaccentuation of given referents for German children [1], [2]. In our corpus, deaccentuation of referential expressions is scarce. Behind these differing results may lie the different prosodic profiles of Germanic and Romance languages, something which future research needs to assess. Finally, regarding nuclear accentuation, we found that, although all three types of referents were significantly more accented with a nuclear pitch accent, the number of new referential expressions in nuclear position was significantly higher compared to given and accessible referents already at the ages of 4-5. Moreover, the tendency to have given referents in prenuclear positions increased significantly from the ages of 4-5 to 7-9. Research is still needed to examine the role of other types of prosodic prominence (e.g., pitch accent types) in information status marking and its relationship with gestural marking.

5. References

1. De Ruiter, L. (2013). How German children use intonation to signal information status in narrative discourse. *Journal of Child Language* 41, 1015–1061. doi: 10.1017/S0305000913000251
2. Grünloh, T., Lieven, E., & Tomasello, M. (2015). Young Children's Intonational Marking of New, Given and Contrastive Referents. *Language Learning and Development*, *11*(2), 95– 127. doi: 10.1080/15475441.2014.889530
3. Esteve-Gibert, N., Loevenbruck, H., Dohen M., & d'Imperio, M. (2019). Pre-schoolers use head gestures rather than duration or pitch range to signal narrow focus in French. ICPhS 2019 - 19th International Congress of Phonetic Sciences, Aug 2019, Melbourne, Australia. hal-02367780
4. Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties, in *Information structure, linguistic structure and the dynamics of language acquisition*, eds. C. Dimroth and M. Starren (Amsterdam, The Netherlands: John Benjamins), 311–328.
5. McNeill, D. (1992). *Hand and Mind*. Chicago, IL: University of Chicago Press.
6. Baumann, S., & Schumacher, P. B. (2020). The incremental processing of focus, givenness and prosodic prominence. *Glossa: a journal of general linguistics 5*(1). 1–30. doi: 10.5334/gjgl.914
7. Marslen-Wilson, W. D., Levy, E., and Komisarjevsky Tyler, L. (1982). Producing interpretable discourse: the establishment and maintenance of reference. *Language, place, and action: Studies in deixis and related topics*, eds. R. J. Jarvella and W. Klein (Chichester, UK: Wiley), 339–378
8. Skopeteas, S. (2016). Information Structure in Modern Greek. In C. Féry & S. Ishihara (Eds.), *Oxford Handbook of Information Structure* (pp. 686–708). Oxford: Oxford University Press.
9. Debreslioska, S., & Gullberg, M. (2020). What's New? Gestures Accompany Inferable Rather Than Brand-New Referents in Discourse. *Frontiers in Psychology*, *11,* 1935. doi: 10.3389/fpsyg.2020.01935
10. Chen, A. (2011). Tuning information packaging: intonational realization of topic and focus in child Dutch. *Child Language, 38*, 1055–1083. doi: 10.1017/S0305000910000541
11. Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3-4), 243–276. doi: 10.1556/ALing.55.2008.3-4.2
12. Götze, M., Weskott, T., Endriss C., Fiedler, I., Hinterwimmer, S., Petrova, S., Schwarz, A., Skopeteas, S., & Stoel, R. (2007). Interdisciplinary Studies on Information Structure. *Working Papers of the SFB 632*, Potsdam, (147–187).
13. Baumann, S., & Riester, A. (2012). Referential and lexical givenness: Semantic, prosodic, and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (pp. 119–162). Berlin: De Gruyter Mouton.
14. Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*(4), 680–703. doi: 10.1016/j.wocn.2008.05.001
15. Jasinskaja, E. (2016). Information Structure in Slavik. In C. Féry & S. Ishihara (Eds.), *Oxford Handbook of Information Structure* (pp. 709–732). Oxford: Oxford University Press.
16. Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. *Natural Language & Linguistic Theory*, *23*(3), 687–755. doi: 10.1007/s11049- 004-2874-7
17. Dufter, A., & Gabriel, C. (2016). Information structure, prosody, and word order. In S. Fischer & C. Gabriel (Eds.) *Manual of Grammatical Interfaces in Romance* (pp. 419–455). Berlin: De Gruyter.
18. Gutiérrez-Bravo, R. (2002). Focus, Word Order Variation and Intonation in Spanish and English: An OT account. *Amsterdam Studies in the Theory and History of Linguistic Science. Series IV, Current Issues in Linguistic Theory*, *217*, 39–54. doi: 10.1075/ cilt.217.06gut
19. Im, S., and Baumann, S. (2020). "Probabilistic relation between co-speech gestures, pitch accents and information status," in *Proc. Linguist. Soc. Am.* 5: 685. doi: 10.3765/plsa.v5i1.4755
20. Yoshioka, K. (2008). Gesture and information structure in first and second language. *Gesture* 8, 236- 255. doi: 10.1075/gest.8.2.07yos
21. Rohrer, P.L., Florit-Pons, J., Vilà-Giménez, I. & Prieto, P. (2021). Children use non-referential gestures in narrative speech to mark discourse elements which update common ground. *Frontiers in Psychology, 12*. doi: 10.3389/fpsyg.2021.661339
22. Chen, A. (2018). Get the focus right across languages: Acquisition of prosodic focus marking in production. In P. Prieto & N. Esteve-Gibert (Eds.) *The Development of Prosody in First Language Acquisition* (pp. 295–317). Amsterdam: John Benjamins.
23. Demir, Ö. E., Levine, S. C., & Goldin-Meadow, S. (2012). Getting on track early for school success: The STEP assessment system to support effective instruction (PreK-3rd). Creating an assessment of literacy skills for three- and four-year-olds. *Foundation for Child Development.* Available from: https://www.fcd-us.org/getting-on-track-early-for-school-success-the-step-assessment-system-to-support-effective-instruction-prek-3rd/
24. Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol.* 3, 71–89. doi: 10.1515/lp-2012-0006
25. Igualada, A., Esteve-Gibert, N., and Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. J. Exp. Child Psychol. 156, 99–112. doi: 10. 1016/j.jecp.2016.11.017
26. Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., & Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. Journal of Experimental Child Psychology, 172(8), 168–188. doi: 10.1016/j.jecp.2018.02.004
27. Vilà-Giménez, I., Igualada, A., and Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Dev. Psychol.* 55, 250–262. doi: 10.1037/dev0000604
28. Vilà-Giménez, I., and Prieto, P. (2020). Encouraging kids to beat: Children's beat gesture production boosts their narrative performance. *Dev. Sci.* 23, 1–14. doi: 10.1111/desc.12967
29. Vilà-Giménez, I., Florit-Pons, J., Rohrer, P. L., Muñoz-Coego, S., Gurrado, G., & Prieto, P. (2021, June 8). Catalan children's multimodal development in narrative discourse. Retrieved from osf.io/npz3w
30. ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan
31. Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.08. Last accessed on the 29th of February of 2020 from http://www .praat.org/
32. McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226514642.001.0001.
33. Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: a labelling proposal. In *Proc. 9th International Conference on Speech Prosody 2018*, 201–205.
34. Rasinski, T. (2004). Creating fluent readers. *Educational Leadership*, *61*(6), 46.
35. Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., Sichel-Bazin, R., & Vanrell, M. M. (2015). Intonational phonology of Catalan and its dialectal varieties. In S. Frota, & P. Prieto (Eds.), *Intonation in Romance* (pp. 9–62). Oxford: Oxford University Press.
36. Cravotta, A., Busà M. G., & Prieto, P. (2019). Effects of Encouraging the Use of Gestures on Speech. *Journal of Speech, Language, and Hearing Research,* 1-16. doi:10.1044/2019_JSLHR-S-18-0493
37. Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.
38. Lüdecke, D., Makowski, D., Ben-Shachar, M., Patil, I., Waggoner, P., Wiernik, B., and Arel-Bundock, V. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software 6,* 3139. doi: 10.21105/joss.03139.