A longitudinal study of children’s intonation in narrative speech

Jeffrey Kallay and *Melissa A. Redford*

Department of Linguistics, University of Oregon, Eugene, Oregon, USA

jkallay@uoregon.edu, redford@uoregon.edu

Abstract

Adults’ narratives are hierarchically structured. This structure is evident in the linguistic and prosodic domains. Children’s narratives have a flatter structure. This structure is evident in the linguistic domain, but less is known about the prosodic domain. Here, we report results from a longitudinal study of children’s narratives that enhance our understanding of the development of discourse prosody. Spontaneous narratives were obtained from 60 children (aged 5 to 7) over a 3-year period. F0 was tracked to obtain absolute measures of slope steepness and linearity for every utterance of each narrative. These measures are known correlates of syntactic and semantic complexity. Slope direction and inter-utterance continuity in F0 were also calculated. These measures are known correlates of event boundaries in adult discourse. The results indicated systematic developmental changes related to age and year for all measures except slope steepness, consistent with developmental increases in linguistic complexity and the production of more adult-like narratives. The evidence also indicates that developmental change is most pronounced between the ages of 5 and 7 years, and levels out afterwards.

Index Terms: prosodic acquisition, discourse prosody, intonation.

1. Introduction

Children’s narratives are known to be qualitatively different than adults’. Whereas adult discourse displays a hierarchical structure such that a stratified organization of ideas emerges through the embedding of locally related events within larger segments [1], children’s narratives display a relatively flatter structure. The flatter structure of children’s narratives is evident at multiple levels of linguistic analysis, including at the level of syntactic complexity [2], discourse marking [3], information structure [4], and thematic coherence [5]. For example, 5-year-old children are much more likely than adults to sequentially chain events through the use of connectives such as *and* or *then*, lending a “narrowly local level of organization” to their narratives [3, p. 67]. By age 9 years, children produce narratives comprised of sentences that are syntactically complex, but they still do not embed events within a larger network of ideas in an adult-like manner.

The hierarchical structure of adult narratives is also evident in the prosodic domain. For example, the overall steepness of the intonation (F0) contour and extent of pitch accenting correlates with the syntactic and semantic complexity of the language produced. Specifically, the steepness of falling contours in adult discourse is correlated with sentence length such that longer utterances are produced with shallower F0 slopes than shorter utterances [6, 7]. Also, variability of F0 across the utterance is an indication of utterance-level information structure conveyed by semantic focus and boundary tones [8, 9, 10]. Finally, global information structure is marked by pausing patterns, low terminal tones, and pitch resets across strong boundaries [1, 11, 12, 13].

The developmental trajectory of prosodic cues to narrative structure is not well-established. The few studies that have been completed suggest the gradual emergence of adult-like discourse prosody. For example, Vion and Colas [11] investigated French-speaking children’s prosodic marking of utterances in narratives using perceptual judgments and found that “the number of prosodic units indicating discourse continuation increases as the linguistic and communicative skills of the narrator improve” [p. 496].

Vion and Colas’ [11] finding that prosodic cues to continuation increase in children’s narratives over developmental time is also a bit surprising if children’s flatter discourse structure is cued in large part by the sequential chaining of events using connectors. More specifically, English-speaking children’s heavy use of *and* and *then* might be expected to correlate early on with more prosodic continuity across utterances rather than less. A preliminary study of children’s discourse prosody provides some support for this expectation. Redford et al. [14] found that “adults produced a higher proportion of phrases [acoustically] marked for completion than children did” [p. 3]. But we clearly need more data on the development of discourse prosody to understand its emergence. The goal of the current study was to provide some such data.

Spontaneous narratives were obtained from 60 children over a 3-year period. Children were between 5 and 7 years old at the outset of the study. The recorded narratives were acoustically segmented into pause-delimited utterances. F0 was then tracked to obtain objective descriptions of intonational structure relevant to discourse prosody. In particular, we obtained absolute measures of slope steepness and linearity for every utterance. The direction of the slopes were then categorized as falling or rising. We also obtained a measure of the relative degree of continuity between the F0 contours on either side of uttterance boundaries (i.e., pauses) to identify discontinuities that would signal pitch resets. If discourse prosody develops in parallel with increasing utterance length and narrative complexity, we expected to find that children’s intonational patterns become more adult-like over developmental time. Specifically, we expected slope steepness would decline as children’s mean length of utterance increased; the linearity of F0 contours within the utterance would decrease due to increasing prominence marking of salient information and the presence of boundary tones within an utterance; the proportion of falling contours would increase as children move away from an “event-chaining” narrative strategy; and discontinuity of F0 between utterances would increase for the same reason.

2. Methods

2.1. Participants

The speech samples used in the current study were obtained from 60 typically developing children. Our longitudinal corpus contains recordings of 71 children in total, but 11 of these were excluded from the present analyses due to noisy recordings. In total, 29 children were male and 31 were female. Each child was recorded once per year over 3 years. Children ranged in age from 5;2 to 7;10 at the time of the first recording. Of the 60 children, 16 (12 male) were 5 years old (M = 5;6), 20 (7 male) were 6 years old (M = 6;6), and 24 (10 male) were 7 years olds (M = 7;5) in the first year of study.

2.2. Speech Samples

Spontaneous narratives were elicited using the "Frog Story" picture books by Mercer Mayer. The task was completed by both a caregiver and the child to create a more natural storytelling situation, namely, one with interlocutors. Each participant chose a different Frog Story from which to base their narrative. While the caregiver familiarized him/herself with their picture book story, the experimenter helped the child look through his/her choice, drawing the child's attention to particular events at predetermined locations in the book. The goal was to help the child conceptualize a coherent narrative. Following familiarization, either the child or caregiver told their story to the other. When the first participant completed his/her story, the second participant told his/her story. After the first round of tellings was complete, a second round was initiated so that each child (and caregiver) told the same story twice. The goal was to minimize language planning and word-finding effects on prosody. The second telling was used for the current study, as it was more likely to represent fluent speech and a higher quality narrative (see [15] for task details).

The narratives were digitally audio recorded using a wireless microphone attached to a hat or headband that participants' wore during storytelling. Each recording was pause-delimited into individual utterances and transcribed orthographically by a trained researcher. Disfluencies were specially coded. These were defined as word interruptions, repetition-restarts, and prolongations [16]. For the purpose of marking utterance boundaries, a pause was defined as any period of non-speech that lasted for at least 150 milliseconds. Mean length of utterance was calculated from the transcriptions.

2.3. Prosodic Measures

Each pause-delimited utterance was first extracted from within each of the audio recorded narratives using Praat software. The utterances were tagged for position within the narrative so that sequencing and inter-utterance characteristics could be assessed. The sonorant portions of the utterances were then identified, extracted, and concatenated to create new sonorant-only utterances over which F0 could be continuously tracked without the spurious errors that occur with abrupt changes in voicing. F0 was calculated at 10 millisecond intervals across each sonorant-only utterance. The measures were then normalized across narratives (and thus across speakers and speakers' age) by subtracting the mean F0 for a narrative from each pitch point of every sonorant-only utterance to center the F0 contour around zero.

Once centered, a predicted loess (local regression) line was fit to each F0 contour using R software. The slopes of the contours were obtained by fitting a linear regression model to the loess line and extracting the regression coefficient. The slope steepness was the absolute value of the coefficient, and the slope direction its sign. The relative linearity of the contour was measured according to the fit of the linear model in terms of the R^2 value. A higher R^2 value indicated a more linear contour, while a lower R^2 a contour with a higher degree of local F0 fluctuations.

To measure inter-utterance continuity, at each utterance boundary, the final pitch of the pre-boundary utterance was calculated as the average F0 over the last 50 milliseconds, and the initial pitch of the post-boundary utterance as the average F0 over the first 50 milliseconds. By averaging the F0 over 5 pitch points for each of these, it was hoped that the undue influence of any spurious F0 measurements would be minimized, yielding a more accurate reflection of the actual boundary values. Continuity of the F0 contour across utterances was then measured according to the absolute differences between the pre- and post-boundary F0 values. For each speaker, the standard deviation of the differences was calculated, and any interval with a difference greater than one standard deviation was categorized as discontinuous, while the rest were considered to be continuous. All cases in which there was a disfluency immediately adjacent to either side of the boundary were excluded from this analysis, as these were deemed likely to influence the degree of continuity.

2.4. Analyses

The continuous variables of slope steepness and linearity were analyzed using linear mixed effect models. These were constructed so that age at study outset (a between-subjects factor), study year (a within-subjects factor), and their interaction were included as fixed effects. Speaker was included as a random effect. In order to test the effect of each of the predictor variables on overall significance, reduced models were constructed by removing each effect in turn, and comparing the fit to the full model using ANOVA analyses. The resulting χ^2 statistic was used to assess effect significance. The categorical variables of slope direction and inter-utterance continuity of F0 contours were analyzed using logistic regression models, with the same effect structure as above.

3. Results and Discussion

3.1. Narrative Characteristics

Across the 3 years of study, the average length of narratives was 169.2 seconds, with unsystematic differences between Year 1 (M = 160.2 sec.), Year 2 (M = 182.4 sec.) and Year 3 (M = 164.7 sec.). There were also no systematic differences in the average number of utterances per narrative per year. In Year 1 the mean number of utterances was 67.75, in Year 2 it was 73.02, and in Year 3 it was 70.98. The average number of fluently produced words per narrative was 353, with Year 1 narratives containing significantly fewer fluently produced words (M = 317.6) than Year 2 (M = 370.4) and Year 3 (M = 372.9) narratives. There was a corresponding progressive increase in the mean length of utterance (MLU) as measured by the number of fluently produced words per utterance: Year 1, M = 4.90; Year 2, M = 5.29; and, Year 3, M = 5.54. A similar age-related increase was found between the age groups: 5-year-olds in Year 1 had MLU of 4.79 words; the 6-year-olds' MLU was 5.26 words; and the 7-year-olds' MLU was 5.59 words. Overall, the changes in number of fluent words produced per narrative and the time and age-related increases in MLU provide evidence that children's narrative language became more complex with developmental time.

3.2. Within-utterance Prosodic Measures

3.2.1. Slope Steepness

Comparing the full mixed effects model to the reduced ones, it was found that both age [$\chi^2(2) = 6.60$, $p = .03$] and study year [$\chi^2(2) = 19.63$, $p < .001$] had a significant effect on the steepness of the slopes. There was also found to be a significant interaction between age and study year in this regard [$\chi^2(1) = 4.71$, $p = .03$]. These results can be seen in Figure 1.

While the mixed effects analysis revealed significant main effects of age and study year on the steepness of the F0 contour slopes, there is a strong indication from the pattern in Figure 1 that neither of these effects were systematic. While slope steepness decreased from Years 1 to 3 for the speakers in the 5-year-old group, there was an increase each year for the 6-year-old group, and an initial increase followed by a decrease for the 7-year-old group. Overall, this result is not what we would expect if slope steepness correlates with utterance length, since we know there were systematic increases in MLU across the study years.

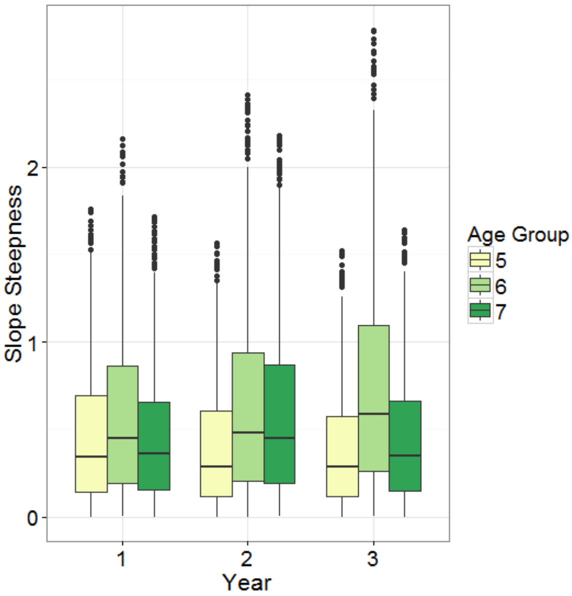

Figure 1: *Slope steepness by age group and year*.

3.2.2. Slope Linearity

A similar comparison of the full and reduced models also revealed that there were significant effects of both age [$\chi^2(2) = 12.72$, $p < .01$] and study year [$\chi^2(2) = 17.13$, $p < .001$] on the relative linearity of the pitch contour slopes. Once again, there was also a significant interaction between age and study year [$\chi^2(2) = 12.37$, $p < .001$]. These effects can be seen in Figure 2.

Figure 2 shows a decrease in linearity for the 5-year-old group from Year 1 ($M = .36$) to Year 2 ($M = .30$) and Year 3 ($M = .29$). This result is more in line with expectations stemming from the systematic increase in MLU over the same period of time. The figure also shows that changes in slope linearity were less pronounced in Year 2 and Year 3 of the study, suggesting that by age 7 years children have settled into a fairly stable pattern of focus and boundary marking within utterances.

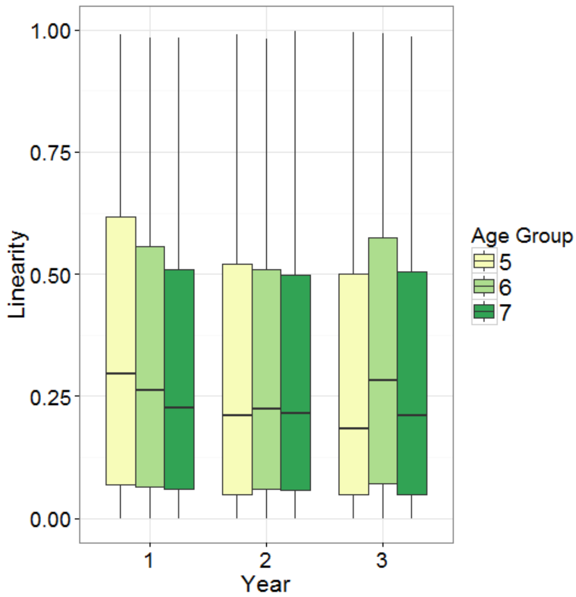

Figure 2: *Slope linearity by age group and year*.

3.2.3. Slope Direction

For the direction of the pitch contour slopes, a logistic regression analysis revealed that there were once again significant effects of age [$z = -2.04$, $p = .04$] and study year [$z = -3.77$, $p < .001$]. However, the interaction between these two factors was not significant [$z = 1.70$, $p = .09$]. These results can be seen in Figure 3 where the proportions of falling versus rising contours is plotted by the actual age of the speaker at the time of recording.

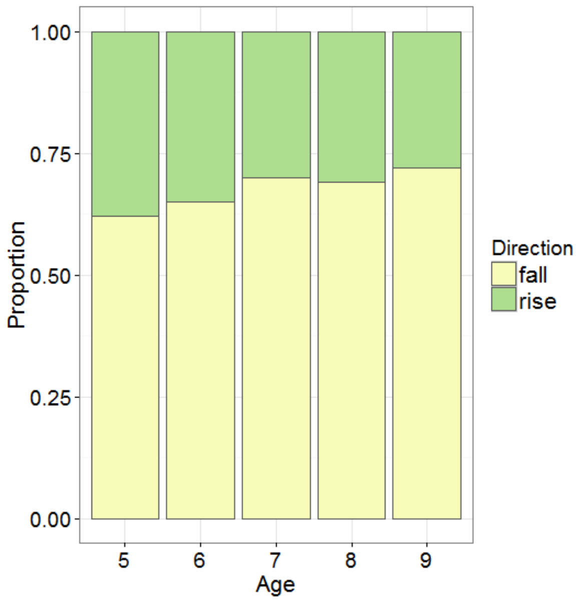

Figure 3: *Proportion of falling vs. rising contours by age*.

Figure 3 shows the systematic change in the relative proportions of contour types over developmental time. Ten percent more of the total utterances were realized with a falling contour at age 9 than at age 5. This finding is in line with expectations based on adult narratives. It is also worth noting that, similar to the slope linearity, the change related to slope direction appears to level out by age 7 years. The proportions of falling contours for each group are given in Table 1.

5	6	7	8	9
.62	.65	.70	.69	.72

Table 1. *Proportion of falling contours by age.*

3.3. Inter-utterance Prosodic Measure

The measure of inter-utterance continuity in F0 was also analyzed using a logistic regression analysis. The results indicated a significant effect of study year [z = -2.21, p = .03], but not of age [z = -1.12, p = .26]. The interaction between age and study year was not significant either [z = 1.14, p = .26]. When considered in terms of children's actual ages at time of recording, a systematic pattern emerges. Figure 4 shows that younger children generally produced fewer continuous contours from utterance-to-utterance than older children. This finding parallels Vion and Colas' [11] finding that discourse continuation marking increases in parallel with improving linguistic and communicative skills, but it is at odds with the expectation that young children's use of an event-chaining narrative strategy should result in more continuous intonation contours from utterance to utterance. Either way, the finding suggests more frequent pitch resetting at utterance initial position in older children's narratives than in younger children's narratives. Again, the data suggest that developmental changes in discourse prosody plateau around 7 years old, similar to the pattern that was found in the linearity of the slopes over time. The plateauing with age, seen in Figure 4, likely accounts for the lack of a significant effect of age on the measure of inter-utterance F0 continuity/ discontinuity.

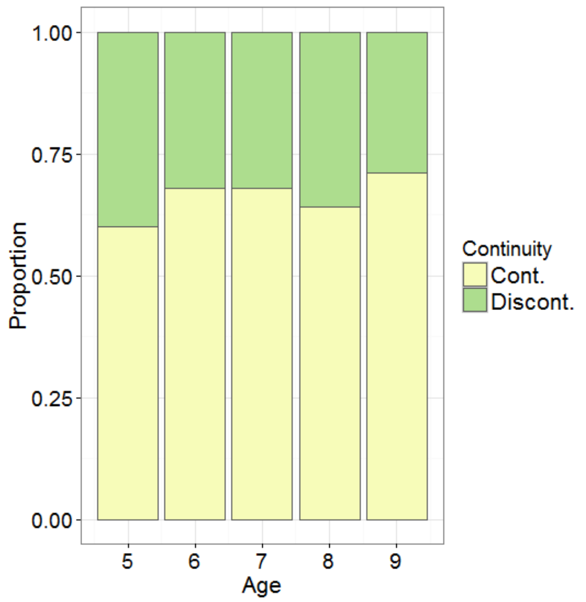

Figure 4: *Slope continuity by actual age.*

4. Conclusion

In general, the results of this study suggest that school-aged children's narratives gradually become more adult-like in the prosodic domain, just as they become more adult-like in the syntactic, semantic, and discourse domains. Of the prosodic features investigated here, all but slope steepness showed developmental change over time and all were in the expected direction. The results also strongly suggest that patterns of slope linearity, direction, and continuity change significantly between age 5 and 7 years, then slow or plateau from age 7 to 9 years.

The one feature of discourse prosody that did not change within the period of development studied was the relative steepness of F0 contours for individual utterances. Barring further investigation, it is not entirely clear why this feature may have differed from the others. One possibility is that it is covariate with some other characteristic of typical adult discourse which was not included in our analyses. For example, slope steepness is strongly correlated with utterance length [6]. But it is also the case that declination patterns are less consistent in spontaneous speech than in read-aloud speech [17]. Thus, it could be that slope steepness varies automatically with utterance length in prepared (pre-planned) speech, rather than with discourse structure per se.

Acknowledgements

This work was supported in part by a grant from the National Institute of Child Health and Human Development (NICHD), #R01HD061458. The content is solely the responsibility of the authors and does not necessarily reflect the views of NICHD.

5. References

[1] M. Swerts, "Prosodic features at discourse boundaries of different strength," *Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 514-521, 1997.

[2] H. Diessel and M. Tomasello, "A New Look at the Acquisition of Relative Clauses," *Language*, vol. 81, no. 4, pp. 882-906, 2005.

[3] R.A. Berman and D.I. Slobin (eds.), *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1994.

[4] A. Gopnik, S. Choi and T. Baumberger, "Cross-linguistic differences in early semantic and cognitive development," *Cognition*, 11, pp. 197-227, 1996.

[5] C. Peterson and A. McCabe (eds.), *Developing Narrative Structure*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1991.

[6] M. Swerts, E. Strangert, and M. Heldner. "F0 Declination in Read-Aloud and Spontaneous Speech," *Proceedings., Fourth International Conference on Spoken Language*, vol. 3, pp. 1501-1504, 1996.

[7] J. Yuan and M. Liberman. "F0 Declination in English and Mandarin Broadcast News Speech," *Speech Communication*, vol. 65, pp. 67-74, 2014.

[8] J. Hirschberg and J. Pierrehumbert. "The intonational structuring of discourse," *Proceedings, 24th Annual Meeting of the Association for Computational Linguistics*, pp. 136-144, 1986.

[9] J. Pierrehumbert and J. Hirschberg. "The Meaning of Intonational Contours in the Interpretation of Discourse," in P. Cohen, J. Morgan and M. Pollack, editors, *Intentions in Communication*, Cambridge, MA: MIT Press, 1990.

[10] M. Steedman. "Information Structure and the Syntax-Phonology Interface," *Linguistic Inquiry*, vol. 31, no. 4, pp. 649-689, 2000.

[11] M. Vion and A. Colas, "Intentional control and operational constaints in prosodic phrasing: a study of picture elicited narrations by French children," *Language and Speech*, vol. 52, no. 4, pp. 481-513, 2009.

[12] M. Swerts and R. Geluykens, "Prosody as a marker of information flow in spoken discourse," *Language and Speech*, vol. 37, no. 1, pp. 21-43, 1994.

[13] C.H. Nakatani, J. Hirschberg and B.J. Grosz. "Discourse Structure in Spoken Language: Studies on Speech Corpora," Paper presented at *AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Palo Alto, CA, USA, March 27-29, 1995.

[14] M. A. Redford, L.C Dilley, J.L. Gamache and E.A. Wieland. "Prosodic Marking of Continuation versus Completion in Children's Narratives," Paper presented at *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association*, Portland, OR, USA, Sept. 9-3, 2012.

[15] M.A. Redford. "A comparative analysis of pausing in child and adult storytelling," *Applied Psycholinguistics*, vol. 34, pp. 569-589, 2013.

[16] W.J.M. Levelt. "Monitoring and self-repair in speech," *Cognition*, vol. 14, pp. 41-104, 1983

[17] P. Lieberman, W. Katz, A. Jongman, R. Zimmerman, and M. Miller. "Measures of the sentence intonation of read and spontaneous speech in American English," *Journal of the Acoustical Society of America*, vol. 77, no. 2, pp. 649-657, 1985.