



# The role of Intentional control and operational constraints in Prosodic phrasing

Monique Vion, Annie Colas

## ► To cite this version:

Monique Vion, Annie Colas. The role of Intentional control and operational constraints in Prosodic phrasing. 2005, pp.1-14. hal-00136764

**HAL Id: hal-00136764**

**<https://hal.science/hal-00136764>**

Submitted on 15 Mar 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The role of Intentional control and operational constraints in Prosodic phrasing

*Monique Vion & Annie Colas*

*E-mail: [mvion@up.univ-mrs.fr](mailto:mvion@up.univ-mrs.fr) ; [acolas@up.univ-mrs.fr](mailto:acolas@up.univ-mrs.fr)*

**Summary:** Tape-recorded narrations told by seven- to eleven-year-old children were evaluated perceptually. The children were shown wordless comic strip stories. The narrative picture sequences were either arbitrary or ordered. In the arbitrary sequence, the events depicted in each picture, although presented as a sequence, could have occurred in any order. In the ordered sequence, the events could only occur in one order. Moreover in the later sequence, the ordinary succession of events was modified by the occurrence of a surprise event. Stories were produced according to simultaneous or consecutive pictures' display mode. In simultaneous mode, all pictures were on a single page. In consecutive mode, the comic strip was presented in booklet format with one picture per page and the storyteller was not allowed to look ahead or backtrack. Naïve French listeners were asked to segment the recordings using intonation variations as a criterion, and decide whether each prosodic segment was "conclusive" or "continuative".

The results showed: (1) that the listeners identified more continuative segments in the narrations of the eleven-year-old children than in the seven-year-old children's one; (2) that the listeners identified more conclusive segments in narratives produced in consecutive mode than in simultaneous mode; (3) the prosodic marking of the surprise event of the ordered sequence in consecutive mode.

## 1. Introduction

The idea that the prosody of spontaneous spoken discourse is made up of units that display prosodic grouping is largely acknowledged (Hirst and Di Cristo 1998). These units, which are characterized by a coherent intonation structure, vary in size. The size of the intonation units in a discourse can range from a single syllable to a syntactic phrase, a clause, a sentence, or larger utterances (Botinis, Granström, and Möbius 2001). A general issue concerning the purpose of intonation units is finding out what determines their construction during speech. The present study tries to experimentally identify some of the psycholinguistic constraints likely to influence the way in which a speaker divides the speech flow into discourse segments by means of prosody.

While speaking, speakers express the temporarily active contents of their mind. Under the assumption that speech production is incremental in nature, the information to be expressed is displayed in an ordered and cumulative sequence of content fragments. The fragments are verbalized one after the other. They are processed from top to bottom by various components of the processing system working in parallel (from the message component to the articulatory component via the grammatical and phonological components; Levelt 1989; Bock 1995).

Various reasons have been proposed to explain the prosodic segmentation of speech. Levelt (1989), for example, lists five reasons pertaining to the speaker's intentional control, speaking rate, syntax, semantics and operational constraints. Two of them are studied here. The first one concerns the intentional control exerted by the speaker on his/her speech. The speaker can use prosodic units as pragmatic devices to be highly intelligible to the listener. Di Cristo (2000) defends the idea that the overall function of prosody is to assist in the coding and decoding of speech, which essentially serves

to interpret utterances semantically and pragmatically. According to this view, breaks that are formally introduced into discourse while prosodic units are being generated are considered to be prosodic boundaries. That is to say, they are functional markers that indicate the completion (temporary or final) of both the acoustic event and the intended content. The second reason pertains to operational constraints. According to Levelt (1989), the speaker can take a break during incremental processing in order to wait for a new block of materials for phonological encoding. The lack of new ammunition can come from a difficulty at grammatical encoding level or be caused by a planning problem at the message level.

The present study was based on a double assumption: first, cues to the cognitive activities that occur during planning are provided by the way words are grouped prosodically; second, the resulting prosodic boundaries are utilized by the listener as he/she is interpreting the produced utterance. The first aim of the study was to show that prosodic structure varies as a function of both the speaker's intentional control (which is dependent on his/her linguistic abilities and communicative skills) and memory-related operational constraints and inference making. Our second aim was to demonstrate that prosodic structure can help listeners interpret an utterance. The field of the study is the development of narration. The speech materials consisted of children's narrations elicited by picture sequences.

Narrative skills are known to develop significantly during the elementary school years (Botvin and Sutton-Smith 1977). The present study focuses on the productions of elementary school children. Children were tape-recorded while telling stories depicted in comic strips with no text. The general hypothesis was that the linguistic expressions that structure discourse are manifestations of the conceptual constraints imposed by the information management process (Bronckart 1985; Chafe 1986). Four production situations were defined for the pictorial-narrative verbalization task by varying the cognitive constraints of the task. First, the establishment of temporal and/or causal relations between the events depicted in each frame was guided to different extents by the drawings, the inference making constraints were manipulated by using an ordered vs. arbitrary sequence of events. Second, the memory-searching constraints on planning were manipulated by varying the way in which the frames were displayed (all at once vs. gradually). By taking into account all combinations of these two variables, each with two categories, we conducted a series of studies to compare the textual organization of the children's productions (Vion and Colas 1998, 1999 a and b, 2001, 2004, 2005).

The aim of the present study was to show that the segmentation of a speech production into prosodic units is dependent upon (1) the child's narrative skills, (2) how explicitly the pictures depict the links between the events to be related, and (3) the way the picture story is displayed. The experimental task was a perceptual evaluation task. It required metalinguistic activity to identify the prosodic structure. The assumption was that the hearer can use his/her spontaneous capacity in a conscious and thoughtful way to make use of prosodic cues in the task. The empirical studies on this point are scarce and rarely take into account inter-speaker variations in suprasegmental cue configuration, which may be very large (see reviews by Cutler, Dahan, and van Donselaar 1997, and by Fon 2002). These studies have shown that intonative information (F0 and pitch range variations, phrase accents, and boundary tones) is used to process the relationships between the current utterance and those that precede and follow it (Hirschberg and Pierrehumbert 1986; Pierrehumbert and Hirschberg 1990; Grosz and Hirschberg 1992; Grosz, Hirschberg and Nakatani 1994; Swerts and Geluykens 1994; Auran, Portes, Rami, and Rigaud 2001; Portes 2002).

Judges who were not linguists performed the auditory analysis of the prosodic structure of the narrations. They were asked to: (1) segment the productions using melodic variations as a criterion, (2) decide whether a segment was "conclusive" or "continuative", and (3) assess the strength of the

boundary thus identified. The procedure was based on Rotondo's argument (1984), quoted by Swerts (1997), that group segmentation data can be used to derive the hierarchical structure of discourse. The proportion of listeners agreeing on a given break was seen as a measure of discourse boundary strength. The procedure was also based on the results of the study by Auran et al. (2002), which showed that listeners are able to make a categorical difference between "continuation" and "finality" in spontaneous spoken discourse.

Considering the variety of text organizations used in this study, not only differences in the prosodic structure of the productions but also regularities should be expected. The regularities should result partly from the operational constraints manipulated and partly from the development of children's communicative skills (intentional control). The predictions for each age and production situation will be presented below and take into account the possible interaction of these two dimensions.

## 2. Method

### 2.1. Speech Materials

The audio-recorded productions that were to be used in the material for perceptual evaluation were taken from the corpus collected earlier using a pictorial-narrative verbalization task (Vion and Colas 2000). Tape-recording took place in a quiet but not soundproof room in the children's school. Testing was individual. There were three persons in the room: the speaker, the experimenter, and a listener (who was a same-age, same-grade peer of the speaker and acted as the listener for only one speaker during the experiment). The two child participants were seated facing each other at a table so that the listener could not see the pictures and the speaker could not assume shared background knowledge of the narrated content. Each speaker was tested on one picture sequence display mode. In the *simultaneous* mode (S), the speaker saw the frames all at once (on one page), and he/she was asked to look at the entire comic strip and prepare to tell the story immediately afterwards. In the *consecutive* mode (C), the comic strip was presented in booklet format, with one picture per page. The speaker was asked to turn the pages one by one, and to say what was happening as each new picture was discovered. The instructions given to the listeners were the same in both conditions. Listeners were told to be attentive and to listen carefully to the stories in order to understand them, but they were not supposed to talk.

The materials for the present experiment were based on two picture sequences (comic strips) in which a character was carrying out various activities. Each comic strip comprised eight frames (8 x 8 cm). The comic strips differed as to how obvious the links were between the successive frames were (Figure 1). In the *arbitrary* sequence (A) (Figure 1a), the events, although presented as a sequence, could have occurred in any order. The activities were relatively independent of each other: the picture of the woman changing clothes could have been placed after the one where she putting on makeup (or anywhere else in the sequence for that matter). It was the speaker's task to infer links between the pictures in order to devise an overall representation of the story. In the *ordered* sequence (O) (Figure 1b), the events could only occur in a single order. Here, before potentially catching a fish, the boy had to put on his fishing gear, go to the water's edge, and cast the line. Although the events were proposed in a chronological order that was more constrained than in the arbitrary sequence, the ordered story still did not have a script structure because the ordinary sequence of events was modified by a surprise event (the boy catches a shoe). Each speaker was tested on one type of event sequence.

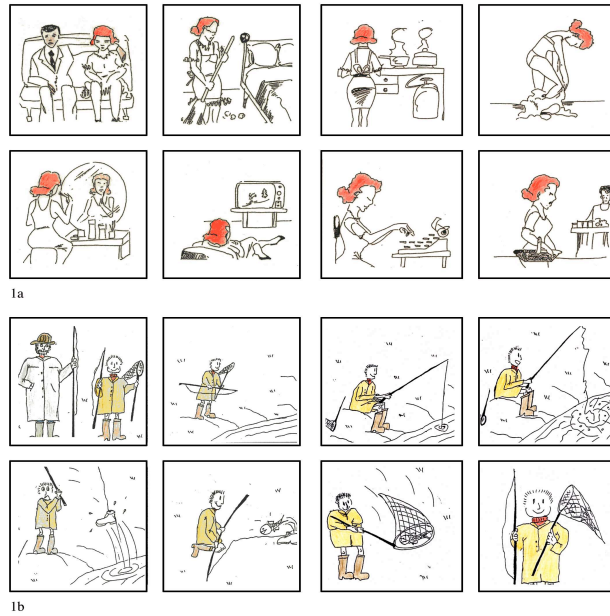


Figure 1: Materials (1a: arbitrary sequence, 1b: ordered sequence).

To sum up, the tape-recordings selected as the materials of the present study were produced in one of four conditions (AC, AS, OC, and OS), obtained by taking all combinations of the two variables “frame display mode” and “type of event sequence”. There were eighty-four speakers (native French-speaking boys and girls ages 7 to 11: 28 seven-year-olds, 28 nine-year-olds, and 28 eleven-year-olds). The productions were tape-recorded first, then converted into files using the Sound Edit 16 system (one sound file per narration). The total duration of the recorded materials was 21 minutes.

## 2.2. Labeling-and-Scaling Task

Segmentation into prosodic segments is based on a perceptual analysis similar to that used by Portes (2000, 2002) and Auran et al. (2001). The labeling-and-scaling task designed for this experiment was more directly derived from the study by Portes (2000, 2002). Judges had to segment the productions on-line. They were instructed to use a binary response system: one slash (/) to indicate continuity and a double slash (//) to indicate finality. The slash(es) were to be inserted between a word that ended a segment and the word that started the next segment. On a second pass through to material they also had to rate the boundary strength using a four-point scale, with “1” meaning “very weak boundary” and “4” meaning “very strong boundary”. The scale was taken from the ToBI standard for prosodic transcription (“Tone and Break Indices”: Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, and Hirschberg 1992), a standard for the hand labeling of the prosodic features of computerized English corpora using a five-point scale (from 0 to 4, with 0 meaning “no-boundary”).

## 2.3. Data Collection Design

Because of the total duration of the materials (21 minutes) and the concentration necessitated by the task, one participant could not judge all the materials. So the materials were divided into 14 blocks composed of six productions each. In each block, the productions were chosen so that a participant could listen to two productions from each of the three age groups (seven-, nine- and eleven-year-olds), three productions from each of the two types of event sequence (A and O), and three productions from each of the two frame-display modes (C and S). This made two groups of blocks with a complementary composition, seven blocks with the composition [9OC - 7OS - 11OS - 9 AS - 11AC - 7AC] and seven

blocks with the composition [9OS - 7OC - 11OC - 9AC - 11AS - 7 AS]. A participant listened to the six productions in a block, in the order indicated above or in the opposite order (each production was thus proposed on two lists). The segmentation of all the materials required setting up 28 lists made up of a sequence of six sound files. Four different participants labeled and scaled the six productions on a list.

## 2.4. Participants

The participants (hereafter called judges) were native French speakers who were not linguists. They were young adult students at the University of Provence in Aix-en-Provence, France. The perceptual evaluation of the materials required one hundred and twelve judges in all (4 judges x 28 lists).

## 2.5. Procedure

The experiment was computerized and took place in a soundproof room. The four judges assigned to a given list were provided with headphones and did the task individually, but simultaneously. The programs, written using the software PsyScope (Cohen, MacWhinney, Flatt, and Provost 1993), presented the sound files.

The experimenter began by telling the participants the purpose of the study. Then in order to familiarize them with the task, the experimenter told the participants they would perform a four-stage training task. During the first stage, they simply had to listen to a sound file. At this stage, no visual support was given: the participants did not see any transcriptions or comic strips. Then they received an orthographically transcribed version of the sound file they had just heard without punctuation nor capitals. On stage 2, they listened to the sound file a second time, and had to write down the breaks they perceived. During the third stage, they were asked to indicate (on a four-point scale) the strength of the boundaries they had written down during stage 2. On the fourth stage, a final hearing of the sound file enabled the listeners to check and perhaps modify their responses. The same stages were used to process the six productions each participant had to segment during the experiment proper.

The testing lasted approximately 30 minutes. During the experiment, the experimenter allowed the participants to rest when requested. At the end of the experiment, the participants were asked to write down the criteria they used to decide whether an utterance would end or continue after the break they had inserted.

## 2.6. Predictions

Many authors agree that prosodic phrasing is related to information units. Given that each comic strip was composed of eight frames, and the content of each frame could be uttered in one proposition, the children were expected to produce at least eight prosodic units per comic strip, but the number of prosodic units as well as their prosodic pattern (continuative vs. conclusive) should vary according to the interaction between the manipulated variables.

The frame display mode and the type of event sequence variables provided a way to manipulate the cognitive constraints (memory searching and inference making) that affect the establishment and management of links between events. Recall that the analysis of the entire corpus from which we drew the present materials showed that the *simultaneous* display mode and the *ordered* sequences are more conducive to establishing conceptual relationships and engaging into a narrative process than the consecutive display mode and arbitrary sequences. It was assumed that this would have repercussions on prosodic phrasing.

When the children saw the pictures all at once and/or when they saw the ordered sequence, they could anticipate, or try to prepare, an overall prosodic plan for their discourse. When carrying out a narrative process, speakers could add information about a character's states or motivations ("he is

happy”), which should also have an effect on prosodic phrasing. In this case, the number of breaks as well as their prosodic pattern (continuative vs. conclusive) should vary more than when the speakers discovered the pictures as they talked or told the *arbitrary* sequence (which induced a strong tendency to describe each picture independently).

The age variable was used to manipulate intentional control. As children grow older, they improve the way they structure and tell stories. In the picture storytelling task proposed here, the development of narrative skills should have less effect on the number of breaks than on their prosodic pattern. Younger children should use the continuative pattern less than older children, and for the same reasons, they should use the conclusive pattern more.

Thus, under the assumption that the break options used by speakers while producing discourse provide listeners with perceptual cues to grasp the discourse structure, interactions between the manipulated variables should be observed in the way the judges segment and label the children’s productions.

### **3. Results**

Recall that the segmentation procedure used in the present experiment was based on the assumption that one can derive the hierarchical structure of discourse from the auditory evaluation made by (several) listeners. For each production, a single segmentation was obtained by comparing the auditory evaluations of all eight listeners. Thus the segmented transcription combined all information about the breaks they noted (location, prosodic pattern, and strength).

#### **3.1. Preparing the Database**

To determine the prosodic boundaries, the proportion of listeners that agreed on each break was computed by performing the following operations. In order to get an odd number of answers, all the answers for a production given by one of the eight listeners (chosen at random) were eliminated. This allowed us to retain the breaks identified by at least four judges out of seven (majority). Then for each production, the break marks (slashes) were inserted in the transcription set for statistical analysis. The number of judges was also noted (which ranged between 4 and 7 for a given break). The prosodic pattern (one vs. two slashes) indicated by the greatest number of judges who noted the break was also written in the transcription, along with the number of judges who chose that pattern (between 2 and 4). Boundary strength was computed by averaging the ratings made by these judges.

#### **3.2. Coding the Data**

Each transcription was then converted into a data file (one per speaker) and coded for computer analysis. Each break was labeled with a five-part code. The first part indicated the number of judges who heard the break; the second, the prosodic pattern (coded c for “/” and t for “//”); the third, the number of judges who gave that prosodic pattern; the fourth, the strength of the break; and the fifth, the prosodic pattern (C vs. T) given for the break by the authors before collecting the data for the present experiment. The fifth piece of information allowed us to decide between the judges’ answers when their judgments were equally allocated to each prosodic pattern. In this case, the strength of the boundary was computed by averaging the weights given by the judges that shared the authors’ opinion. Note that the overall agreement rate between the judges and the authors on the labeling of the prosodic segments as conclusive or continuative was 85.79 %.

The 84 coded transcriptions were then analyzed using the Child Data Exchange System (CHILDES; MacWhinney, 1991). The data were entered in the format defined by the CHAT system

(*Codes for the Human Analysis of Transcripts*), which made it possible to search in the transcriptions using the CLAN module (*Computerized Language Analysis*) for text strings.

### 3.3. Analyses

To study the effects of intentional control and inferential and mnesic constraints, various analyses of variance were carried out with age, display mode, and type of event sequence as between subjects factors. The analysis design was: (3) Age (7, 9, 11 yrs) x (2) Display mode (*simultaneous, consecutive*) x (2) Type of event sequence (*arbitrary, ordered*). Various dependent variables were derived from the coded information concerning the breaks and their prosodic pattern. The strength information is not analyzed here. In the presentation below, only those analyses yielding effects that were significant at 0.05 or less are mentioned, and their results are discussed.

#### 3.3.1. Production Duration

The duration of the productions tended to decrease as the children grew older (approximately 55 seconds at age 7, 49 seconds at age 9 and 46 seconds at age 11). But the analysis of variance carried out on the data showed that the duration did not differ significantly across the ages. There was a significant effect of display mode: it took more time to tell the comic strip when the children discovered the pictures one at a time than when they could see all the pictures before narrating [ $C = 58.6$  vs.  $S = 41.9$ ;  $F(1,72) = 14.32$ ,  $p < 0003$ ].

#### 3.3.2. Prosodic Segmentation

The mean number of prosodic segments was 13.11 per production. Sixty-one percent of the segments achieved a consensus (unanimity). The analysis of variance with the ***number of segments per production*** as the dependent variable yielded a significant effect of display mode: the judges identified more segments in the productions of children who discovered the pictures one by one than those of children who could see all the pictures before narrating [labeled by the majority of judges:  $C = 14.4$  vs.  $S = 11.9$ ;  $F(1,72) = 7.22$ ,  $p < 0089$ ].

To go further into the analysis, each production was divided into two parts: the first concerned setting and presentation of the characters (Frame 1), and the second concerned the various activities carried out by the main character (Frames 2 to 8). The first part of each production was discarded from the remaining analyses, based on the assumption that the data from the first part of the productions might contain substantial variability and mask the effect of the other two manipulated variables on prosodic phrasing.

The mean number of prosodic segments was 10.56 in the second part of the productions (Frames 2 to 8). Note that 65% of the segments were agreed up by all judges (unanimity). The analysis of variance on these data indicated a significant effect of the display mode [majority:  $C = 11.8$  vs.  $S = 9.3$ ;  $F(1,72) = 9.10$ ,  $p < 0035$ ], as above for the whole production. Listeners identified more segments in the productions of children who discovered the pictures one by one while speaking than in those of children who could see all the pictures before narrating. The analysis carried out with the consensual segments as the dependent variable confirmed this result [unanimity:  $S = 5.9$ ,  $C = 7.7$ ;  $F(1,72) = 13.51$ ,  $p < 0005$ ].

To determine the effect of the other two manipulated variables, the data for Frames 2 to 8 were then processed as follows. The prosodic segments were separated into two categories. The first category concerned segments with a phatic function (um, so um, and um), or word choice ("he takes / he takes the / he takes the net"; "then / it is / the husband / of / finally..."), and sometimes syntactic phrasing ("i-i-it takes a net then after goes/it it/it tries to fish"). The classification was based on the assumption that



these segments reflected activities (maintenance of communication, lexico-syntactic choices) that did not alter the speaker's general communicative plan. The second category included segments related to the general structure of discourse. Each author did the classification independently. Disagreements were scarce and were resolved by discussion.

The subsequent analyses concerned the second category (i.e. the segments pertaining to the speaker's general communicative plan). The analysis of variance with ***the number of segments per production*** as the dependent variable yielded a significant effect of the display mode similar to that obtained above [majority:  $C = 12.5$  vs.  $S = 10.1$ ;  $F(1,72) = 8.09$ ,  $p < 0058$ ]. The analysis on the unanimous segments also yielded a significant effect of the display mode [unanimity:  $C = 7.9$  vs.  $S = 6.1$ ;  $F(1,72) = 11.30$ ,  $p < 0012$ ]. The last two analyses, which are in line with the preceding ones, indicate the impact of the amount of available information on children's prosodic planning. As expected, gradually discovering the contents to be reported led the speakers to plan with little lookahead and to generate a greater number of prosodic units.

### 3.3.3. Prosodic Patterns

The analysis of the segments involved in the general structure of the discourse was conducted, by examining their prosodic pattern (continuative vs. conclusive).

#### *Continuative pattern*

The analysis with ***the number of continuative patterns*** per production (majority) as the dependent variable yielded no significant effects. The analysis carried on the number of unanimous continuative patterns yielded an age effect: as the children grew older, the number of continuative patterns perceived by the listeners in the productions increased [unanimity: age 7 = 3.0, age 9 = 3.3 and age 11 = 4.6;  $F(2,72) = 3.28$ ,  $p < 0434$ ]. A pair-wise comparison using the Newman-Keuls test (at a p-level of 0.05) showed that the observed difference between the evaluations of the 7- and 11-year-old children's production was significant. This result (grounded in a consensus) provides some developmental information: as expected, it argues in favor of the increased use of prosodic units indicating discourse continuation by the oldest speakers in the study.

#### *Conclusive pattern*

The analysis with ***the number of conclusive patterns*** as the dependent variable yielded a significant effect of the display mode: the judges identified more conclusive patterns in the productions of the children who discovered the pictures one by one than in those of the children who could see all the pictures at once [majority:  $C = 3.9$  vs.  $S = 2.1$ ;  $F(1, 72) = 26.59$ ,  $p < 00001$ ]. The analysis with the number of unanimous conclusive patterns also yielded a significant display mode effect, as described above [unanimity:  $C = 3.4$  vs.  $S = 1.8$ ;  $F(1,72) = 29.94$ ,  $p < 00001$ ]. Moreover, there was an interaction between the display mode and the type of event sequence [unanimity:  $F(1,72) = 4.59$ ,  $p < 0356$ ] (Figure 2).

Number of conclusive patterns

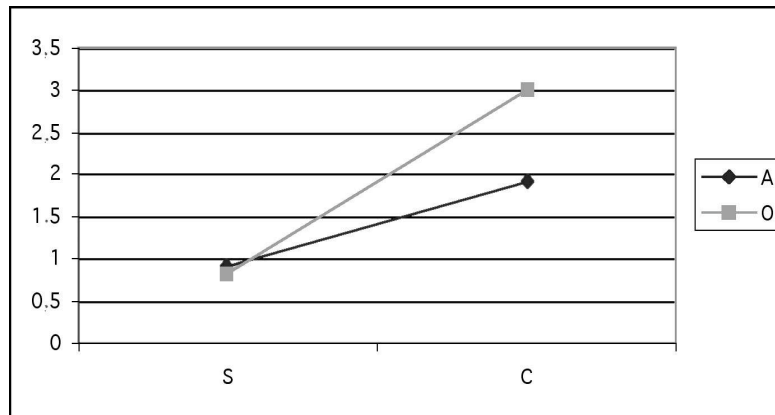


Figure 2: Frames 2 to 8, conclusive pattern (unanimity): display-by-type of event sequence interaction.  
Legend.

A: arbitrary sequence, O: ordered sequence, S: simultaneous display, C: consecutive display.

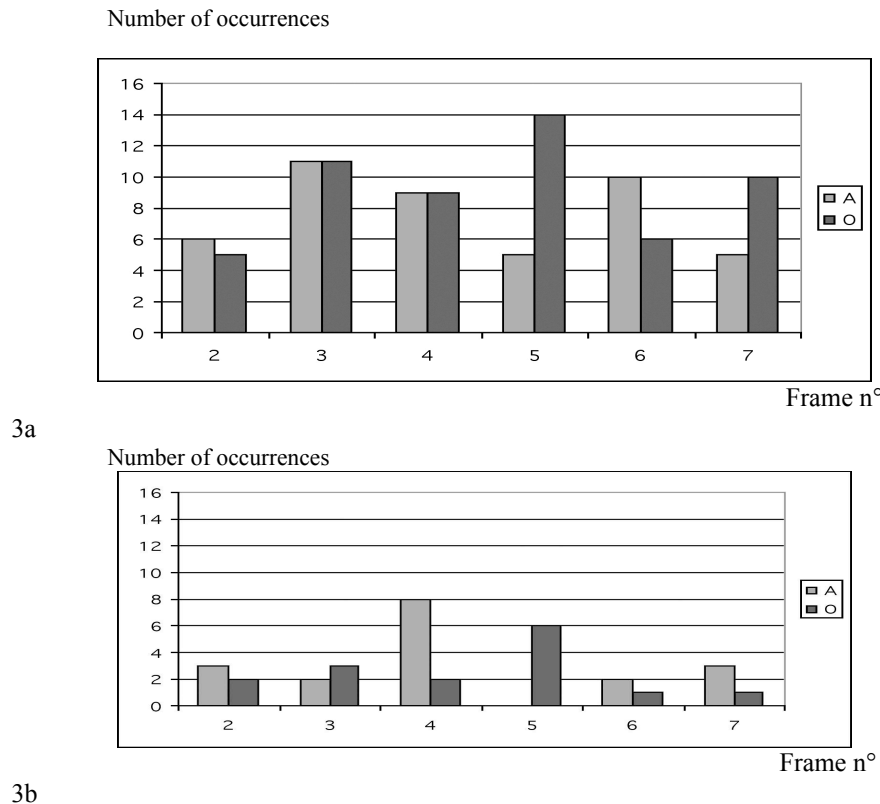
Whereas the number of conclusive patterns perceived in the productions of the children who could see all the pictures did not vary significantly across event sequence types, the number of conclusive patterns perceived in the productions of the children who discovered the pictures one by one was higher for the ordered sequence than for the arbitrary sequence. Unlike the results of the continuative unit analysis, which provided a solid developmental indication of the speakers' behavior, these results mainly confirmed the idea of an impact of the production conditions on the generation of conclusive units. Discovering the ordered pictures one by one led the speakers to generate a high number of content units associated with a prosodic pattern that created a closure effect.

### 3.3.4. Prosodic structure

A last series of analyses dealt with the recognition of continuative and conclusive patterns in the narrations of Frames 2 to 7. This was based on the consideration that the prosodic organization of this portion of the production, unlike Frame 8, was not constrained to the use of conclusive segments due to the need to stop speaking. The analyses carried out with *the number of continuative patterns* as the dependent variable confirmed the age effect described above (majority: age 7 = 5.8, age 9 = 5.5, and age 11 = 7.5; [ $F(2,72) = 3.75$ ,  $p < 0.282$ ]; unanimity: age 7 = 2.9, age 9 = 3.3, and age 11 = 4.3 [ $F(2,72) = 3.04$ ,  $p < 0.541$ ]). The analysis with *the number of conclusive patterns* (majority) confirmed the effect of the display mode described above [ $F(1,72) = 26.37$ ,  $p < 0.0001$ ]. The analysis on the number of unanimous conclusive patterns also confirmed the display mode effect [ $F(1,72) = 29.21$ ,  $p < 0.00001$ ] and the interaction between the display mode and the type of event sequence [ $F(1,72) = 4.07$ ,  $p < 0.0475$ ].

The effect of the type of event sequence was observed only when the pictures were displayed consecutively. In this display mode, the number of conclusive patterns identified by the listeners in ordered-sequence narrations was larger than for arbitrary-sequence narrations. At first sight this result seemed paradoxical. Indeed, one could think that the type of event sequence would have an effect in simultaneous display (where information on the nature of the event sequence was available from the start) rather than in consecutive display (where it was discovered gradually). Moreover, one could expect conclusive units to be more numerous in the arbitrary-sequence productions (which favored the description of independent actions) than in ordered-sequence ones (which favored a narrative process).

To interpret this interaction, the conclusive units identified by the listeners in each production were examined by taking into account the portion of the production (referring each to a picture-content) where each conclusive unit occurred. This was based on the assumption that the higher number of perceived conclusive units perceived when the ordered pictures were discovered one at a time might be related to the discovery of the surprise event during the narration (the boy catches a shoe in Frame 5). The portions of the production referring to the content of each picture (Frames 2 to 7) were thus checked by counting the number of productions in which at least one conclusive pattern was unanimously identified. Figure 3 presents these counts for each type of event sequence and display mode.



*Figure 3:* Frames 2 to 7, number of productions in which one conclusive pattern was unanimously identified (distributions: max. 21 per picture – 3 ages x 7 speakers )(4a: consecutive display mode, 4b: simultaneous display mode).

Legend.

A: arbitrary sequence, O: ordered sequence, S: simultaneous display, C: consecutive display.

In the consecutive display mode (Figure 4a), the number of productions in which at least one conclusive unit was identified was similar for Frames 2, 3 and 4. Differences between arbitrary and ordered sequences started on frame 5. As expected, the mode of the distribution was located at Frame 5: two thirds of the productions (66.7%) were such that at least one conclusive unit was identified, versus 23.8% at this point for the arbitrary sequence (the lady is putting on her makeup). Thus, the greater number of conclusive units identified by the listeners in the productions of ordered pictures discovered one at a time resulted largely from the surprise event deliberately introduced by the authors. Note that this explanation was supported by a similar phenomena observed when all the pictures were available before the narration (figure 4b), where conclusive patterns were scarce. In this case, the mode

of the ordered-sequence distribution was also located at Frame 5 (28.6%). Furthermore, the mode of the arbitrary-sequence distribution was located at Frame 4, where the picture showed the lady partially undressed. At this point, approximately a third of the productions (38%) were such that at least one conclusive pattern was identified. Thus, the simultaneous display of the pictures seemed to allow the children to cause a particular event in the sequence stand out. This event (not voluntarily selected for that purpose by the authors) led certain narrators to highlight it using prosody.

#### **4. Discussion**

The present research took a first step in the study of what might determine how often and when speakers insert breaks in their discourse. Two factors were examined here: (1) the speaker's communicative intention, and (2) the constraints imposed on discourse planning. Narrative productions of 7- to 11-year-old children were evaluated perceptually by a group of ordinary listeners.

The methodology relied partly on the fact that hearers are generally able to adapt to the characteristics of any speaker in order to interpret what s/he is saying, and partly on the idea that listeners can consciously detect the most salient intonation markers and categorize them in terms of discursive continuation or finality when perceptual evaluation is requested.

Setting up a study on narrative discourse provided us with a way of standardizing content across speakers by asking them to tell picture stories displayed in comic strip format. It also allowed us to benefit from the findings of the numerous studies on narrative development conducted over the past few decades.

The auditory evaluation of children's narration showed, first of all, that the listeners always agreed at a rate of at least 60% when identifying prosodic breaks and their patterns, which means that listeners were able to adapt to the various characteristics of the speech of the storytellers. Studies conducted so far have relied mainly on the perceptual evaluation by expert listeners of the productions of very few speakers. The present results encourage us to continue our experimental research into the determinants of prosodic segmentation of discourse by relying on naive listeners and their analysis of the productions of multiple speakers. This kind of study obviously provides coarse results, but they are robust.

The analysis of the listeners' judgments gave us information about the children's prosodic phrasing, but it should be stressed that to arrive at the results as summarized below, a number of transformations, reductions and refinements of the original data were necessary, even though they were collected on the basis of standardized support. All these operations required making decisions that were theoretically justified from the linguistic and psychological points of view. This points out the need for an interdisciplinary approach to create better conditions for studying the characteristic features of spontaneous discourse and the multiple determinants of its prosodic structure.

The present experiment highlighted two facts, one concerning the use of a continuative prosodic unit, and the other, the use of a conclusive unit.

First, the intentional control manipulated via the age variable seems to have mainly played a role in the use of continuative units. Earlier studies on the development of narrative skills showed that the ability to understand and mentally represent a series of events as an integrated (and tellable) whole emerges and solidifies gradually in the course of development (van den Broek, 1997). As expected, the listeners identified more continuative patterns in the narrations of the eleven-year-olds (more involved in controlling a narrative process) than in those of the seven-year-olds (more involved in controlling a descriptive process). But the scope of this result must be specified. Few children constructed stories in which they assigned goals and motivations to the characters. The speakers usually focused on actions. As requested in the instructions, they faithfully reported the events shown in the pictures, only

sometimes mentioning the characters' thoughts and affective reactions. One can assume that the increased use of continuative units with the narrator's age represents an elementary way of expressing that the described events constitute a story.

Second, the planning constraints, manipulated in the experiment by way of the display mode and type of event sequence variables were determining factors in the use of conclusive patterns. Manipulating the frame display mode allowed us to control whether the speakers could (simultaneous display mode) or could not (consecutive display mode) make verifications that might help them create or understand the temporal and causal structure of the events to be related. Thus, the frame display mode was used to vary the constraints imposed on the organization and planning of story content (Levelt, 1989; Bock, 1995). On the one hand, the consecutive display mode reduced the span of available information, and since looking forward or backward in the material was not allowed, only mental backtracking was possible. The simultaneous display mode, on the other hand, made the events to be narrated available at all times. The order of the pictures materialized a process whose beginning, middle, and end were indicated by the experimenter. As expected, in the productions the listeners judged, they identified more conclusive patterns when the children told the story as they discovered the pictures than when the children could see the entire comic strip before narrating. Manipulating the type of event sequence provided a way of varying the amount of information available for ascertaining a thematic unit, establishing links between the depicted events, and assigning thoughts, plans, or affective reactions to the character in order to make a story. In the present experiment, it was not possible to show that the links between the events to be reported (less constrained in the arbitrary sequence than in the ordered sequence) had an impact on prosodic phrasing.

A final result regarding intentional control remains to be mentioned. When the ordered sequence was discovered gradually, the listeners identified the greatest number of conclusive patterns in the narrations. This result is not related to an interaction between the type of event sequence (as characterized above) and the display mode, but rather to the discovery in the pictures of a particular salient event by the speaker. The conclusive units identified by the listeners in this production condition were mainly produced at the point where speakers discovered the event that broke the ordinary sequence of actions (the boy catches a shoe). This interpretation was reinforced by the examination of the small number of conclusive patterns identified in the simultaneously displayed arbitrary sequence. Conclusive patterns were more often identified when the children described the picture in which the lady was getting undressed.

This last result illustrates the complexity of the links between the conceptual elaboration of the messages and prosodic phrasing. It illustrates the broad span of meaning that prosody is able to convey in spontaneous speech. Clearly, the issue of what determines prosodic phrasing while speaking needs many more interdisciplinary investigations (1) to clearly differentiate between the various dimensions that affect how prosody "displays" the structure of a text, and (2) to specify how and in what production settings speakers rely on prosody to indicate conceptual relationships.

## References

- AURAN, C., PORTES, C., AURAN, C., PORTES, C., RAMI, E. & RIGAUD, N. (2001), La distinction entre frontières conclusives et continuatives est-elle pertinente dans le discours spontané ? Paper presented at the *Journées Prosodie 2001*, Grenoble.
- BOCK, K. (1995), Sentence production: from mind to mouth, in Miller, J.L. & Eimas, P.D. (eds.), *Speech Language and Communication*, Academic Press, London, 181-216.

- BOTINIS, A., GRANSTRÖM, B. & MÖBIUS, B. (2001), Developments and paradigms in intonation research, *Speech Communication* 33, 263-296.
- BRONCKART, J. P. (1985), Pour un modèle de production du discours, in Bronckart, J-P (ed.), *Le fonctionnement des discours*, Delachaux et Niestlé, Neuchâtel, 3-58.
- CHAFE, W. (1986), Cognitive constraints on information flow, in Tomlin, R. (ed.), *Coherence and grounding in discourse. Typological studies in language*, Vol. 11, John Benjamins Publishing Company, Amsterdam, 21-51.
- COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993), An interactive graphical system for designing and controlling experiments in the psychological laboratory using Macintosh computers, *Behavior Methods, Research, Instruments and Computers* 25, 257-271.
- CUTLER, A., DAHAN, D., & DONSELAAR, W. (1997), Prosody in the comprehension of spoken language: A literature review, *Language and Speech* 40, 141-201.
- DI CRISTO, A. (2000), Interpréter la prosodie, Paper presented at the *XXIIIèmes Journées d'Etude sur la Parole, France*, Aussois, 13-29.
- FON, J. J., K. (2002), A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech. Ohio State University.
- GROSZ, B. & HIRSCHBERG, J. (1992), Some intonational characteristics of discourse structure Paper presented at the *International Conference on Spoken Language Processing*, Banff, Vol. 1, 429-432.
- GROSZ, B., HIRSCHBERG, J. & NAKATANI, G. H. (1994), A study of intonation and discourse structure in directions Paper presented at the *Workshop on the Integration of Natural Language and Speech Processing American Association for Artificial Intelligence*, 124-131.
- HIRSCHBERG, J. & PIERREHUMBERT, J. (1986), The intonational structuring of discourse', Paper presented at the *Twenty-four Annual Meeting of the Association for Computational Linguistics*, 134-144.
- HIRST, D. & DI CRISTO, A. (1998), A survey of intonation systems', in Hirst, D., & Di Cristo, A. (eds.), *Intonation systems: A survey of twenty languages*, Cambridge University Press, Cambridge, 1-44.
- LEVELT, J. M. (1989), *Speaking. From intention to articulation*, MIT Press, London.
- MACWHINNEY, B. (1991), *The Childes project. Tool for analyzing talk*, Lawrence Erlbaum, Hillsdale, NJ.
- PIERREHUMBERT, J. & HIRSCHBERG, J. (1990), The meaning of intonational contours in the interpretation of discourse, in Cohen, P.P., Morgan, J., & Pollack, M. E.(eds.), *Intentions in communication*, MIT Press, Cambridge, MA, 271-323.
- PORTES, C. (2000), *Approche du rôle de la prosodie dans la structuration du discours oral en Français*, Unpublished research note. Université de Provence, Aix -en-Provence.

- PORTES, C. (2002), Approche instrumentale et cognitive de la prosodie du discours en français, *Travaux Interdisciplinaires Parole et Langage* 21, 101-119.
- ROTONDO, J. A. (1984), Clustering analyses of subjective partitions of text, *Discourse Processes* 7, 69-88.
- SILVERMAN, K., BECKMAN, M., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C. M., PRICE, P., PIERREHUMBERT, J. & HIRSCHBERG, J. (1992), ToBI: a standard for labeling English prosody, Paper presented at the *International Conference on Spoken Language Processing*, Vol. 2, 867-870.
- SWERTS, M. (1997), Prosodic features at discourse boundaries of different strength, *The Journal of the Acoustical Society of America* 101, 514-521.
- SWERTS, M., BOUWHUIS, D. & COLLIER, R. (1994), Melodic cues to the perceived « finality » of utterances, *Journal of Acoustical Society of America* 96(4), 2064-2075.
- SWERTS, M. & GELUYKENS, R. (1994), Prosody as a marker of information flow in spoken discourse, *Language and Speech* 37, 21-43.
- VAN DEN BRÆK, P. (1997), Discovering the cement of the universe: the development of event comprehension from childhood to adulthood, in vandenBræk, P., Bauer, P., & Bourg, T. (eds.), *Developmental spans in event comprehension and representation, Bridging fictional and actual events*, Lawrence Erlbaum, Mahwah, NJ, 321-342.
- VION, M. & COLAS, A. (1998), L'introduction des référents dans le discours en français : contraintes cognitives et développement des compétences narratives, *L'Année Psychologique* 98, 37-59.
- VION, M. & COLAS, A. (1999a), Expressing coreference in French: Cognitive constraints and development of narrative skills, *Journal of Psycholinguistic Research* 28, 261-291.
- VION, M. & COLAS, A. (1999b), Maintaining and reintroducing referents in French: Cognitive constraints and development of narrative skills, *Journal of Experimental Child Psychology* 72, 32-50.
- VION, M. & COLAS, A. (2000), Mode de recueil et outil d'analyse d'un corpus de parole spontanée étudié d'un point de vue psycholinguistique, *Travaux Interdisciplinaires Parole et Langage* 11, 155-167.
- VION, M. & COLAS, A. (2004), On the use of the connective "and" in oral narration: A study of French-speaking elementary school children, *Journal of Child Language* 31, 399-419.
- VION, M. & COLAS, A. (2005), Using connectives in oral French narratives: Cognitive constraints and development of narrative skills, *First Language* 25, 39-66.