This is a guide to CHILDES data on narrations by children and adults. For a general introduction to the CHILDES database, please consult [intro.pdf](#). The links in the table below are clickable, as are the thumbnails to the left.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Age Range</th>
<th>N</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetries on page 2</td>
<td>child, adult, elderly</td>
<td>30,20,20</td>
<td>Picture descriptions in Dutch from a control population</td>
</tr>
<tr>
<td>ENNI on page 7</td>
<td>4-9</td>
<td>400</td>
<td>Reference database for comparing TD and SLI children’s narratives</td>
</tr>
<tr>
<td>Frog stories on page 11</td>
<td>3–adult</td>
<td>600</td>
<td>Narratives from 8 languages elicited using a picture book describing the adventures of a boy, a dog, and a frog</td>
</tr>
<tr>
<td>Gopnik on page 29</td>
<td>2–5 years</td>
<td>61</td>
<td>Cross-sectional study based on story book descriptions from a population of normal children</td>
</tr>
<tr>
<td>Hicks on page 31</td>
<td>1st grade 2nd grade 5th grade</td>
<td>20 18 5</td>
<td>Cross-sectional study of primary school children’s narrative genre skills. Data focus on their ability to produce a range of narratives</td>
</tr>
<tr>
<td>MacWhinney / Bates 1 on page 36</td>
<td>3 years 4 years 5 years adult</td>
<td>30 30 30</td>
<td>Cross-sectional study using a picture description task in English, Italian, and Hungarian</td>
</tr>
<tr>
<td>MacWhinney / Bates 2 on page 37</td>
<td>3 years 6 years 10 years adult</td>
<td>76 53 44 103</td>
<td>Cross-sectional study using a film description task and a cartoon-description task with Hungarian and American children and adults</td>
</tr>
<tr>
<td>VionColas on page 57</td>
<td>7, 9, 11</td>
<td></td>
<td>Picture descriptions</td>
</tr>
</tbody>
</table>


1. Asymmetries – Dutch

Petra Hendriks
Charlotte Koster
Sanne Kuijper
University of Groningen
P.O. Box 716
9700 AS Groningen, Netherlands
p.hendriks@rug.nl
www.let.rug.nl/~hendriks/vici.htm

1. Background
The Asymmetries Project collection contains Dutch language productions gathered in Groningen and neighboring towns in the northern Netherlands, between 2007 and 2012. The research was carried out by members of the NWO/Vici project “Asymmetries in Grammar” at the University of Groningen. This project investigates asymmetries between production and comprehension in unimpaired children, in young and elderly adults, and in autistic and ADHD children and adolescents. It is funded by a grant from the Netherlands Organization for Scientific Research (NWO) awarded to Petra Hendriks (grant no. 277-70-005).

2. Participants
All participants are native Dutch speakers. The participants in the CK sub-corpus have no history of language problems. The CK sub-corpus includes 31 typically developing children (4;3-6;5, mean 5;6), 20 young adults (18-35, mean 26;2), and 20 elderly adults (69-87, mean 78;8). The groups are balanced for sex. At a later date (2012), transcripts from the SK sub-corpus will also be submitted to CHILDES, including 40 children with Autistic Spectrum Disorder (ASD), 40 children with Attention Deficit Hyperactivity Disorder (ADHD), 40 typically developing children, between 6-12 yrs old. To date (spring, 2011), the majority of the SK-children have been tested and transcribed.

3. Procedure
The children in the present study (CK-sub-corpus) were individually tested in a quiet room at their primary school in Stadskanaal (45 km southeast of Groningen). Testing of the children took place in the winter of 2007-2008. In the spring of 2010, the two adult groups participated in the same experiment. The young adults, both students and non-students, were tested in their homes or at the university. The elderly adults were all tested in their homes. The elderly were socioeconomically representative of their generation’s middle class and all still lived independently, with a minimum of assistance. All the adults lived in the greater Groningen area.

The language productions consisted of structured storytelling, which was the first part of a larger experimental session including a memory test and a language comprehension test. A child was first shown one introductory page including pictures of all the storybook figures, such as a princess, witch, ballerina, nurse, pirate, knight,
cowboy, indian, etc. The figures were depicted in stereotypical color drawings. The child was asked to name the figures and was helped if needed. Then the child was told that she would see picture books and should tell what was happening on the pages of the books. Investigator-1 explained that investigator-2 wanted to listen in, but because she was sitting further away behind a computer screen, she couldn’t see the pictures. So the child should explain as clearly as possible what was happening in the picture books. After a practice session, the child saw a picture book with six pictures, one per page (see storybook pictures below). She described the activity on each picture-page as she looked at it and could see only one picture-page at a time. The child saw four picture books in total. After each picture book was completed, the child was rewarded with a sticker and intermittently reminded that investigator-2 couldn’t see the pictures. The child’s descriptions were basically monologues. If necessary, investigator-1 prompted the child. Investigator-1 often gave slight encouragements while turning the page, such as “yes” or “good job” or some other short, “empty” supporting comment. The total production session usually took between 7 - 15 minutes. This time includes the instructions, introductory page, practice stories, four test stories and rewards between stories. The children told all four stories within about 3 – 7 minutes (including reward-time between stories). Adults were quite efficient, with a total storytelling time of about 3 – 5 minutes. The elderly talked more, or more slowly and completed the four stories in about 5 – 10 minutes.

Experimentation with the two groups of adult participants followed the same procedure, with a few minor differences. The adults got no “rewards” between storybooks, although they were given a small present of chocolates at the end of their session. They were tested by only one investigator, who explained that someone else would process the tapes further. They were told that this second investigator would not see the pictures, so the participants must be clear about what was happening. The adults were also asked to limit the length of their answers.

4. Transcriptions, recordings, and use of data

The recorded sessions were first transcribed in Word by the investigator who was present at the test session. All participant productions and investigator productions (above a whisper) were transcribed. For the children, their non-storytelling productions during the reward moments between the four picture books were not transcribed. Then the transcripts were coded in CHAT. A different investigator again controlled the CHAT files and sound files for coding consistency. Productions relating to each picture of each story were transcribed orthographically on the main tier and separated by gems (@G). Occasional use of a second tier was limited to %com: for exceptional situations. Pauses were noted as (.) for a clear pause shorter than 2 seconds, (..) for 2-4 seconds, and (…) for more than 4 seconds. Researchers specifically interested in pauses, stuttering, mispronunciations, contrastive stress should consult the sound files for exact measurements.

The participants are registered anonymously under number, as can be seen on the first initial header line (see example below). The 31 children are coded with CKc01 – CKc33. Participants CKc24 and CKc25 have been excluded from the set. The 20 young adults are coded as CKa01 – CKa21. Young adult CKa15 has been excluded. The 20 elderly adults are coded as CKe26 – CKe47. Elderly participants CKe29 and CKe31 have
been excluded. Several investigators are identified by first name (coincidentally, they often have the same first name!). On the second line, the language, participant’s age, sex and group can be found. The groups are identified as SAchildren, SAadults and SAelderly, with SA standing for Subject Anaphora (see Research Goals below). Example of the relevant initial header lines:

@Participants: CHI CKc11 Target_Child, INV Sanne3 Investigator
@ID: nld|asymmetry|CHI|6;2.|female|SAchildren| [Target_Child] |
@ID: nld|asymmetry|INV| [female|SAchildren| [Investigator] |

The sound files were recorded on an Olympus voice recorder as Windows Media Audio (WMA) files. The names of the sound files match the transcript file names.

No restrictions are placed on the use of either the transcribed data or the sound files of the children, young adults and elderly adults in the CK-subcorpus.

5. Research Goals
These productions are being used in several ongoing investigations. The main research goal for which the picture books were designed was to study subject anaphora in relation to discourse topic and topic shift: specifically, when does a speaker use a full noun phrase and when does she use a pronoun.

6. Acknowledgements
A journal article is in preparation. Until further notice, researchers wishing to use this data should cite the project website:
www.let.rug.nl/~hendriks/vici.htm

7. Picture Storybooks
The original picture books were six A4 pages long, with one picture on each page, labeled as below. If anyone wishes to use this material, they must first contact Petra Hendriks (p.hendriks@rug.nl) for permission and the original files.
Ballerina story (@G: ballerina 1-6)

Pirate story (@G: piraat 1-6)

Princess story (@G: prinses 1-6)
Indian story (@G: indiaan1 1-6)
2. ENNI

Phyllis Schneider
University of Edmonton
Edmonton, CA

The Edmonton Narrative Norms Instrument (ENNI) database was collected to develop measures and norms for storytelling using stimuli designed for the purpose. We used the ENNI to collect local norms from children in the city of Edmonton, Alberta, Canada. The data collection was funded by Children's Health Foundation of Northern Alberta. The ENNI website is: http://www.rehabresearch.ualberta.ca/enni

Acknowledgements. Please cite the following in publications and presentations of work that used the ENNI. Contact information for the authors is available on the website.


Restrictions. The authors ask that researchers who use the data send copies of publications based on the data.

Description of the Normative Study

Participants. The ENNI sample consisted of two subgroups within every age group: a wide range of typically developing children and children previously identified as having a language impairment. Because the norms will be particularly useful for professionals interested in language impairment, special care was taken to include a representative sample of children previously identified as having a specific language impairment. However, since we did not obtain IQ test information for the children, the group is best described as having language impairment.

Sample size for the group with typically developing language was 50 children per age group (one-year intervals), with equal numbers of boys and girls. The group with language impairment varies from 10 to 17 children per age group. Gender was left to vary in this group; as expected, there were more boys than girls (48 of 77 – 62%) in the group with language impairments. Stories were collected from children ages 4 through 9;11, for a total of 377 children.

Number, Age, and Language Status Information for the Normative Sample

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Language Group</th>
<th>Total N</th>
<th>N Boys</th>
<th>Mean Age</th>
<th>Age SD</th>
<th>Age Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>TD</td>
<td>50</td>
<td>25</td>
<td>4.60</td>
<td>.24</td>
<td>4.04-4.97</td>
</tr>
</tbody>
</table>
Schools, preschools and daycares were randomly selected from areas all across Edmonton to assure a sample that was representative of the Edmonton population. In all, 34 elementary schools and 13 daycares, preschools and independent Kindergarten programs were visited to collect the data.

Socioeconomic and ethnic characteristics of the sample are reported in the ENNI manual, [http://www.rehabresearch.ualberta.ca/en/](http://www.rehabresearch.ualberta.ca/en/).

**Materials**

Six original picture stories with animal characters were used to elicit stories, organized into two sets with 3 levels of complexity each. The stories were controlled in pairs and systematically varied across levels for number of pictures, amount of story information, and number and gender of characters. The table below provides a summary of the characteristics of the story sets. The pictures for each story were placed in page protectors in a binder, each story in its own binder. Information about the development of the pictures is available on the ENNI website manual, as are the pictures themselves.

**Characteristics of the Story Sets**

<table>
<thead>
<tr>
<th>Story</th>
<th>Number of Episodes</th>
<th>Setting</th>
<th>Number of Characters</th>
<th>Character Description</th>
<th>No. of Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>Swimming pool</td>
<td>2</td>
<td>young female elephant young male giraffe</td>
<td>5</td>
</tr>
<tr>
<td>A2</td>
<td>2</td>
<td>same</td>
<td>3</td>
<td>same as A1 plus adult male elephant lifeguard</td>
<td>8</td>
</tr>
<tr>
<td>A3</td>
<td>3</td>
<td>same</td>
<td>4</td>
<td>same as A2 plus adult female elephant</td>
<td>13</td>
</tr>
<tr>
<td>B1</td>
<td>1</td>
<td>Park</td>
<td>2</td>
<td>young male rabbit young female dog</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>2</td>
<td>same</td>
<td>3</td>
<td>same as B1 plus adult female rabbit doctor</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>3</td>
<td>same</td>
<td>4</td>
<td>same as B2 plus adult male rabbit balloon-seller</td>
<td></td>
</tr>
</tbody>
</table>

**Procedure**

Each child was seen individually in the child's school, preschool, or daycare. The child was first given a training story, which was similar to the simple stories in the two story sets in terms of length (5 pictures, 1 episode) and number of characters (2). The purpose of the training story was to familiarize the child with the procedure and to allow the examiner to give more explicit prompts if the child was having difficulty with the task. For the sets A and B stories, the examiner was restricted to less explicit assistance such as general encouragement, repetition of the child’s previous utterance, or if the child did not say anything, a request to tell what was happening in the story.

After the training story, the child then viewed the pictures for each story in turn and was asked to tell the story to the examiner. When presenting the stories, the examiner held the binder in such a way that she could not see the pictures as the child told the story, which meant that the child needed to use language rather than pointing or gesturing if the examiner was to understand the story. The instructions emphasized that the examiner would not be able to see the pictures, so the child would have to tell a really good story so the examiner could understand it.

The examiner first went through all the pages so that the child could preview the story, after which the examiner turned the pages again as the child told the story. Administration of the story sets was counterbalanced.

Note that this task is story generation from pictures, not a retell task. The child was not told a story by the examiner.

Children's story retellings were audiotaped and later transcribed orthographically in full. The transcripts were divided into communication units (C-units), each of which consisted either of one independent clause plus any dependent clauses associated with it or of a sentence fragment. Contractions were spelled out using parentheses, for example, he (i)s for he's. No attempt was made to transcribe speech errors accurately. Children's names do not appear in the transcripts. More information on transcription is available on the website, [http://www.rehabresearch.ualberta.ca/enni/manual/transcription](http://www.rehabresearch.ualberta.ca/enni/manual/transcription).

Transcripts were checked against the recordings by the primary investigator before being analysed. A research assistant transcribed 5% of the stories for reliability purposes; word-by-word reliability was calculated to be 97%.

**Measures.** Measures were developed to assess story information included in children's tellings (Story Grammar), introduction of referents (First Mentions), and
syntactic complexity (Complexity Index). Story Grammar is scored using two stories, A1 and A3; the other scores are based on all 6 stories. Although CLAN was used to code these measures, the codes are not included in the files donated to CHILDES. There are norms for these measures as well as for Mean Length of Communication Unit, Total Number of Words, and Number of Different Words. These measures and norms are available on the website, http://www.rehabresearch.ualberta.ca/enni/analyses.

File information. The first number of each file indicates age (e.g., 4xx indicates age 4). Each file contains a header with information about the child's age, gender, and group (TD = typically development, SLI = language impairment). Stories and sets are separated by 'gem' codes; for example:

@BG: SetA  beginning of Set A stories
@BG: A1    beginning of story A1
@EG: A1    end of story A1
3. Frog Stories

Researchers in many countries have used Mercer Mayer’s wordless “frog story” picture book entitled “Frog, where are you?” as a tool for eliciting narrative descriptions. The book tells a story without words in 24 pictures. The principle source for documentation of this work, its rationale, and the various data analysis procedures is the book by Berman and Slobin (1994). Because that book provides such complete documentation for this project, the current documentation will only cover the general issues in the research. Researchers can also consult that book for a complete listing of research in additional languages and with second-language learners using the Frog Story framework. CHILDES currently has data for seven languages. The following table summarizes the data available.

Table 1: Frog Story Corpora

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Languages</th>
<th>Ages</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarssen, Jeroen &amp; Bos, Petra - page 17</td>
<td>Dutch, Arabic, Turkish</td>
<td>4-10</td>
<td>175</td>
</tr>
<tr>
<td>Aguilar, Cesar</td>
<td>Spanish</td>
<td>6, 12</td>
<td>20</td>
</tr>
<tr>
<td>Aksu-Koç, Ayhan</td>
<td>Turkish</td>
<td>3, 5, 9, 20</td>
<td>40</td>
</tr>
<tr>
<td>Bamberg, Michael</td>
<td>German</td>
<td>3, 5, 9, 20</td>
<td>43</td>
</tr>
<tr>
<td>Berman, Ruth</td>
<td>Hebrew</td>
<td>3, 4, 5, 7, 9, 11, 20</td>
<td>91</td>
</tr>
<tr>
<td>Cipriani, Paola</td>
<td>Italian</td>
<td>6, 8, 9</td>
<td>59</td>
</tr>
<tr>
<td>Hemphill, Lowry – page 17</td>
<td>English</td>
<td>6, 7, 8</td>
<td>30</td>
</tr>
<tr>
<td>Lopez-Ornat, Susana</td>
<td>Spanish</td>
<td>unknown</td>
<td>50</td>
</tr>
<tr>
<td>Marchman, Virginia</td>
<td>English</td>
<td>3, 4, 5, 9, 20</td>
<td>60</td>
</tr>
<tr>
<td>Orsolini, Margherita</td>
<td>Italian</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pearson, Barbara – page 20</td>
<td>Spanish, English</td>
<td>various</td>
<td>various</td>
</tr>
<tr>
<td>Sebastián, Eugenia</td>
<td>Spanish</td>
<td>3, 4, 5, 9, 20</td>
<td>59</td>
</tr>
<tr>
<td>Slobin, Dan Isaac</td>
<td>Russian</td>
<td>various</td>
<td>41</td>
</tr>
<tr>
<td>Zlatev, Jordan &amp; Yangklang, Peerapat – page 22</td>
<td>Thai</td>
<td>4, 6, 9, 11, 20</td>
<td>50</td>
</tr>
</tbody>
</table>

The addresses of the contributors of Frog Story data are:

Jeroen Aarssen  
Tilburg University – BABYLON,  
Center for Studies on Multilingualism in the Multicultural Society  
P.O. Box 90153  
5000 LE Tilburg  
The Netherlands  
J.Aarssen@kub.nl

Aguilar, Cesar
Language Engineering Group, UNAM
CAguilar@iingen.unam.mx

Aksu-Koç, Ayhan
Department of Psychology
Bogaziçi University
80815 Bebek
Istanbul, Turkey
koc@boun.edu.tr

Bamberg, Michael
Department of Psychology
Clark University
Worcester, MA 01610
mbamberg@vax.clarku.edu

Berman, Ruth
Department of Linguistics
Tel-Aviv University
Ramat Aviv, Tel-Aviv 69978, Israel
rberman@ccsg.tau.ac.il

Bos, Petra
Tilburg University – BABYLON,
Center for Studies on Multilingualism in the Multicultural Society
P.O. Box 90153
5000 LE Tilburg
The Netherlands
P.H.F.Bos@kub.nl

Cipriani, Paola
IRCCS “Stella Maris”
INPE-Università di Pisa
Viale del Tirreno, 331
Calambrone (Pisa), Italy

Hemphill, Lowry
Harvard Graduate School of Education
703 Larsen Hall
Cambridge, MA 02125 USA
hemphilo@hugse1.harvard.edu

Lopez-Ornat, Susana
Departamento de Procesos Cognitivos
Facultad de Psicología
Universidad Complutense de Madrid
The same procedures were followed across different age groups. Each participant was interviewed individually and was given the same instructions (with slight variations for adults, preschool children, and older children). A deliberate effort was made to minimize
the burden on memory, and to make children aware in advance that they were being asked to tell a story. To this end, children were first asked to look through the entire booklet, and then to tell the story again, while looking at the pictures. They were explicitly oriented to the booklet as presenting a “story” in the initial instructions: “Here is a book. This book tells a story about a boy [point to picture on cover], a dog [point], and a frog [point]. First, I want you to look at all the pictures. Pay attention to each picture that you see and afterwards you will tell the story.”

Because the goal was to leave the burden of narration on the child, without scaffolding by the adult, the various adult interviewers were instructed to minimize their verbal feedback to neutral comments that would not influence the form of expression chosen by the child. It was especially important that the interviewer avoid prompts that would lead to a particular choice of verb tense, aspectual marking, or perspective on the part of the child. The following prompt types were used, presented below in English, in order of preference (neutrality): (1) silence or nod of head, (2) “uh-huh,” “okay,” “yes,” (3) “Anything else?” (4) “and...?” (5) “Go on.”

For most of the languages, groups of 12 participants were recorded at several different age levels. The file names give the age level of the participant along with a letter. The files were originally transcribed in a format specified in the Berman and Slobin manual and then converted to CHAT in 1995. The conversion to CHAT was straightforward except for the English Berkeley data which will need some additional double-checking to eliminate inaccuracies. During the conversion, pictures were marked with @g headers. Researchers have used two different systems for marking picture or page numbers. One system uses 1a and 1b for left and right pages. The other system numbers pages without regard to left-right position. Here are the correspondences between the two systems:

<table>
<thead>
<tr>
<th>Actual Page</th>
<th>Left Page</th>
<th>Right Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6a, 9</td>
</tr>
<tr>
<td>2a</td>
<td>2</td>
<td>6b, 10</td>
</tr>
<tr>
<td>2b</td>
<td>3</td>
<td>7, 11, 12b</td>
</tr>
<tr>
<td>3a</td>
<td>4</td>
<td>8, 12a, 18</td>
</tr>
<tr>
<td>3b</td>
<td>5</td>
<td>9a, 13, 13b</td>
</tr>
<tr>
<td>4a</td>
<td>6</td>
<td>9b, 14, 14a</td>
</tr>
<tr>
<td>4b</td>
<td>7</td>
<td>10a, 15, 14b</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10b, 16, 15</td>
</tr>
</tbody>
</table>

Here is a brief page-by-page description of the pictures in *Frog, Where Are You?* by Mercer Mayer (1969). The pictures from the book and additional guidelines from Berman and Slobin can be found in [http://childes.psy.cmu.edu/manuals/frog.pdf](http://childes.psy.cmu.edu/manuals/frog.pdf)

Page 1: A boy and his dog are in the boy’s bedroom admiring a smiling frog in a glass jar. The moon can be seen through an open window. The boy is in his pajamas, his boots are at the foot of the bed and his clothes are on the floor.
Page 2: The boy and the dog are asleep in the boy’s bed. The frog is climbing out of the jar.
Page 3: It is now morning. The boy and the dog are awake and have observed that the
frog is missing.
Page 4: The boy is looking in one of his boots for the frog while the dog has stuck his head in the frog’s jar.
Page 5: The boy and the dog are looking out the window (the reader sees the building exterior). The boy looks like he is calling out something (i.e., both hands are by his mouth, which is open). The glass jar is stuck on the dog’s head.
Page 6: The dog is falling out the window and the boy looks puzzled.
Page 7: The boy has come outside and is holding the dog. The jar has broken and pieces are lying on the ground. The boy has a scowl on his face and the dog is licking the boy’s cheek.
Pages 8 & 9 (one picture): The boy is calling (i.e., both hands are by his mouth, which is open) and the dog is sniffing with his nose in the air. In the distance is a forest. A beehive is hanging in a tree by the edge of the forest and bees can be seen flying from it.
Page 10: The boy is calling (i.e., one hand is by his mouth, which is open) down a hole in the ground while the dog is jumping up toward the beehive.
Page 11: A small ground rodent, such as a ground squirrel or gopher, has popped out of the hole. The boy is holding his nose and looking unhappy. The dog is still jumping up toward the beehive.
Page 12 & 13 (one picture): The beehive has fallen out of the tree and angry bees are swarming. The boy is sitting on a branch of a large tree exploring a hole in the tree.
Page 14 & 15 (one picture): An owl, with open wings, has come out of the hole and the boy has fallen on the ground. The bees are chasing the dog.
Page 16: The boy is running away from the owl. In the background is a large boulder. Branches of trees can be seen behind it.
Page 17: The boy has climbed to the top of the boulder and is calling (i.e., one hand is by his mouth, which is open). He is holding a branch of a tree. The dog can be seen slinking toward the boulder. His tail is between his legs.
Page 18: What appeared to be branches are, in fact, the antlers of a deer. The boy can be seen draped over the deer’s head.
Page 19: The deer is walking, with the boy on his head, toward a cliff. The dog is chasing the deer.
Page 20: The deer has tipped the boy over the edge of the cliff and the dog has apparently fallen off the cliff. Both the boy and the dog are in the midst of falling into a marshy pond.
Page 21: The boy and the dog have fallen head first into the water with a splash. Only their legs are visible.
Page 22: The boy is sitting in the water and the dog is sitting on the boy’s shoulder looking over his head. The boy is holding his hand to his ear and smiling, as if he has heard something.
Page 23: The boy is kneeling beside a large log. The dog is swimming toward him. The boy is holding one finger to his mouth (i.e., gesture indicating a need for silence).
Page 24: The boy and the dog are looking over the log. The reader observes them from the back and does not know what they are seeing.
Page 25: The boy and the dog are sitting on the log and are looking at a mother and father frog – one or the other may be the frog that escaped. The frogs are snuggled together and smiling.
Page 26: The frogs’ children emerge from tall grasses on the right. The adult frogs have proud smiles on their faces as they look at their children. The boy and the dog are sitting on the log. The boy is smiling.

Page 27s & 28: The boy and the dog are leaving. The boy has a small frog in his hand and is waving at the frog family, which is sitting on the big log.

The data from Berkeley, Israel, Germany, and Rome were collected in the context of a project directed by Dan I. Slobin and Ruth A. Berman with support from the United States-Israel Binational Science Foundation (Grant 2732/82), the Linguistics Program of the National Science Foundation (Grant BNS-8520008), the Sloan Foundation Program in Cognitive Science, the Institute of Human Development at UC Berkeley, the Committee on Research of the Academic Senate at UC Berkeley, and the Max-Planck Institute for Psycholinguistics. The Turkish data were gathered with support from the Bogazici University Research Fund (Project No. 86 B 0724). The Swedish project was supported by a Swedish Tercentenary Foundation grant to Sven Strömqvist (Riksbankens Jubileumsfond, grant 91-231:01). The collection of the English corpus from Wolf and Hemphill was supported by a Program Project grant from NIH on the “Foundations of Language Assessment.” The data from Miami were collected in conjunction with the Bilingualism Study Group Literacy Grant, supported by NIH Grant #1R01 HD 30762-01 to D. Kimbrough Oller and Rebecca Eilers.

For a full description of the data collection methods, codes, and analyses followed in most of these studies, please consult this basic work that should also be cited in publications using these data:


Additional details are available for a few of the corpora. These are given next.
3.1 Aarsen/Bos Corpus

In the description of the bilingual corpora for Aarsen/Bos, you can find a full description of this corpus.


3.2 Chinese-Guo


3.3 Hemphill Corpus

This corpus, donated by Lowry Hemphill and Dennis Palmer Wolf, includes transcripts from 30 children whose discourse development was studied from ages 6 to 8. The work was funded by a larger project, “Foundations for Language Assessment in Spontaneous Speech,” funded by the National Institutes of Health.

Participants were selected at age 1 from a larger sample of 100 children participating in the MacArthur Individual Differences Project. Information about participant recruitment and characteristics of the original sample can be found in Snow (1989) and Dale, Bates, Reznick, and Morisset (1989). The present sample of 30 children is 50% girls and 50% boys; all are white English-speakers. Fourteen of the children are from working class families; sixteen are from middle class families. All attained milestones for early language development (e.g., MLU) at appropriate ages. Children were videotaped in their home each year at ages 6, 7, and 8, participating in a range of narrative and other discourse tasks. This corpus includes data only for the wordless picture book narration task. Procedures for eliciting the narratives were to have the child look through Mercer Mayer’s wordless “frog story” picture book entitled “A Boy, a Dog, A Frog” to develop a sense of the entire story depicted. This book is similar in format to the “Frog, Where are you?” book used in the other frog story research, but the actual events in the pictures are all different. Then the experimenter asked, “Can you tell me the story, looking through the book?” If the child seemed to have trouble producing narration at any point, the experimenter asked, “What happened next?”

Transcribers trained in CHAT conventions prepared the transcripts, using the videotaped frog story narrations. Utterance boundary decisions were based on intonation contours and pauses. Utterances are broken into grammatical clauses using [c] as a marker of clause boundaries. Each clause is coded for narrative function (e.g., event, reported speech, durative/descriptive), for verb forms, and for use of connectives.

Publications using these data should cite:


Additional relevant references include:

3.4 Pearson Corpus

This directory contains frog story narratives collected in Miami, Florida by Barbara Zuker Pearson with the help of Ana Maria Ferrer, Patricia Ortega, Mayrela Palau, Samantha Pearson, Esperanza Rodriguez, and Yael Wiesner. This is the second set of frog stories collected in conjunction with the Bilingualism Study Group Literacy Grant, supported by NIH Grant #IR01 HD 30762-01 to D. Kimbrough Oller and Rebecca Eilers, with Barbara Pearson and Vivian Umbel.

There are 447 files: 269 in English and 178 in Spanish. They complete the 20 cells of a nested factorial with the factors explained later in the description of the ID numbers. There are 16 cells of 10 Spanish-English bilinguals, with two stories each (for the most part), and 4 cells of monolinguals, 20 children in each with only one story per child. All of the children were born in the United States; they were enrolled in three different instructional programs in Dade County Public Schools in Miami: 1) English immersion for Hispanic students, 2) so-called “two-way” bilingual programs for Hispanic students with 50% Spanish and 50% English instruction, 3) regular monolingual English classrooms for non-Hispanic students, and 4) monolingual English children in schools with primarily Hispanic populations. (Groups 1, 3, and 4 have essentially the same instructional program, but the relation between the student’s own language and the language of the peer population is different.) The stories are on audiotape and 15% of the tapes have been independently transcribed twice for reliability; another 60% have had “second listenings” (where the second transcriber worked from the first listener’s transcription). There are six directories of files:

- BLENG2: Bilingual 2nd graders speaking English
- BLENG5: Bilingual 5th graders speaking English
- BLSpan2: Bilingual 2nd graders speaking Spanish
- BLSpan5: Bilingual 5th graders speaking Spanish
- MLENG2: Monolingual English children 2nd graders
- MLENG5: Monolingual English 5th graders.

Most, but not all, of the bilingual children have both an English and a Spanish story, which can be located by matching the ID number and file names in the English and Spanish directories. Those who wish to work with the matching files would be advised to modify the file names to indicate the language of the story, but should be aware that the @ID line matches the current file name and does not distinguish language. Whether the Spanish or English was told first (on different days) is indicated in the header.

ID Numbers: Files are arranged in English and Spanish directories by ID number, which gives information about group status: digit 1 is school type (above), digit 2 is SES (1=mid, 2=low), digit 3 is language of the home (1 = mostly Spanish, 2 = English and Spanish equally, 3 = only English), and digit 4 is grade (2 = 2nd, 3 = 5th), followed by a 4-digit unique identifier. (For example, 21131489.cha are the stories from participant #1489: she is a bilingual in a two-way school, mid-SES, with mostly Spanish in the
home, in fifth grade at the time of the story.) Within the header, gender is indicated as M or F; the approximate age is in parentheses alongside the grade, 7 or 8 years old for second grade, 10 or 11 for fifth grade. The project records also have birthdays for each child and information about the country of the parents’ origin.

The transcribing conventions were derived loosely from the guidelines found in Ber- man and Slobin (1994) and then converted to CHAT with extensions as noted in the 00depadd file. Comments in the text marked by %exc indicate nonnarrative comments and %pro indicates a pronunciation that is not predictable from the standard orthography. Each verbed clause is marked by a [c]. Verbed clauses need not have a finite verb and in some cases the verb will be absent, as in ellipsis. Modals and aspectual serial verbs are considered as a single verb, as long as the subject does not change. Morphological errors or omissions are marked with %err coding, although users should be aware that this coding has not been found reliable and is used only as a guide by the original researchers.

Publications using these data should cite:

3.5 **Slobin Corpus**

These data were gathered in Berkeley, California, by Tanya Renner (child stories) and Virginia Marchman (adult stories), and these researchers should be credited for the use of the data.

3.6 Thai Corpus

This data was collected as part of the First Language Acquisition of Thai project, funded by The Swedish Foundation for International Cooperation in Research and Higher Education (STINT) and hosted by the Department of Linguistics, Chulalongkorn, Thailand during 2000. Though our project focused on the development of spatial expressions in Thai, we made a serious effort to make the data as consistent and general as possible so that it could be used for other studies as well. We would also like to thank everyone who helped us carry out the collection and transcription of this data: Janich Feangfu, Maneeya Sangjan, Mingmit Sriprasit, Soraya Osathanonda, Martha Karrebaek Hentze and Katarina Lindblom.

The child data was collected in three Bangkok schools and the adult data was collected from students of Chulalongkorn University. The interviewer, always a native Thai speaker, first showed the Frog Story book to the subject and let him scan through it by himself for about 5 minutes. For the children, the instruction were approximately as follows:

This story is about a boy, his dog, and a frog.
I’ll let you take a look at the pictures of the story, first.
Then, I will ask you to tell me the story, picture by picture.

The interviewer sometimes encouraged the child to proceed with the story. These utterances of the interviewer have not been transcribed. Even though we tried to keep the elicitation conditions as uniform as possible, there were inevitable differences due to the fact that five different interviewers collected the data. (The name of the interviewer appears first in the @Transcriber list.)

3.6.1 Transcription

Each recorded narrative was transcribed in standard Thai orthography, in almost all cases by the person, who performed the interview. The Thai transcription was then converted into a phonemic notation via the semi-automatic Thai Transcription program, developed at the Department of Linguistics, Chulalongkorn University. The consonants are as follows:

<table>
<thead>
<tr>
<th>Consonant</th>
<th>labial</th>
<th>postdental</th>
<th>palatal</th>
<th>velar</th>
<th>glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>stop, +voice -asp</td>
<td>b</td>
<td>d</td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>stop, -voice -asp</td>
<td>p</td>
<td>t</td>
<td>c</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>stop, -voice +asp</td>
<td>ph</td>
<td>th</td>
<td>ch</td>
<td>kh</td>
<td></td>
</tr>
<tr>
<td>fricative</td>
<td>f</td>
<td>s</td>
<td></td>
<td></td>
<td>h</td>
</tr>
<tr>
<td>semivowel</td>
<td>w</td>
<td></td>
<td>j</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>lateral</td>
<td></td>
<td></td>
<td>l</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The vowels are as follows

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td>i</td>
<td>U</td>
<td>u</td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td>q</td>
<td>o</td>
</tr>
<tr>
<td>Open</td>
<td>x</td>
<td>a</td>
<td>O</td>
</tr>
</tbody>
</table>

Tones were marked as:  Mid: 0, Low: 1, Falling: 2, High: 3, Rising: 4. Due to requirements of CHAT, the ? for glottal stop was omitted. The presence of the glottal stop is nevertheless derivable from the data since Thai syllables can not begin with a vowel or end with a short vowel. Whenever that seems to be the case in the data, there is an “invisible” glottal stop before the initial vowel or after the final short vowel.

3.6.2 Segmentation

Thai orthography does not place spaces between words: the space within the utterance corresponds to a hesitation pause. The transliteration program does not perform word segmentation either. Therefore, in order to allow the CLAN programs to perform automatic analyses (mlu, frequency counts etc.) the phonemic transcription needed to be segmented into words manually. This was straightforward in most cases since the vast majority of Thai words, especially in the colloquial register, are monosyllabic. However, it is not always clear if certain multi-syllabic expressions should be treated as (a) mono-morphemic words, (b) multi-morphemic words including lexical compounds or (c) phrases consisting of one or more words. In deciding how to analyze particular examples, we used the following criteria, disregarding diachronic evidence:

1. Mono-morphemic word IFF at least one of the syllables in the expression does not have a transparent separate meaning, e.g. naa2taaN1 (‘window’). Even though this expression is probably a compound diachronically, the compounding is not transparent for present-day speakers, so we decided to treat it as mono-morphemic.

2. Multi-morphemic word (“+” between the syllables) IFF all the syllables have transparent separate meanings, but the meaning of the whole is not derivable by combining that of the parts, e.g. phuu2+jaj1 (‘person’+‘big’ = ‘adult’). Lexical compounds are one subclass of this category. Derivations such as khwaam0+suk1 (PROPERTY + ‘happy’ = ‘happiness’) are also included in this category, even though their derivation is semantically regular.

3. Phrase (SPACE between the syllables) IFF the syllables have separate meaning, and combine systematically to give the meaning of the whole, e.g. maa4 (‘dog’) noj4 (‘little’) = ‘little dog’; raN4 (‘nest’) phUN2 (‘bee’) = ‘beehive’.

In the same group as the second category, and thus marked in the same way (with a “+” connecting the parts) were expressions that appeared to be formulaic, e.g. may0+pen0+raj0 (‘never mind’). As for reduplications (cf. Luksaneeyanawin 1984),
these were marked by connecting the parts with a double plus sign “++”, e.g. dek1++dek1 (‘children’). Using this notation compounds and other multi-morphemic words, formulaic expressions and reduplications can be treated as single lexical items (which is intuitively correct) by the CLAN programs, at the same time as analysis can easily be performed on their parts if required. For example, by adding the switch +b+ in the command line of the program MLU the constituent morphemes will be counted separately.

3.6.3 CHAT Formatting

The rough phonemic transcription was then checked against the original tape recordings and corrections were made. Deviations from standard pronunciation were included, using the convention offered by CHAT, placing the standard in square brackets behind pronounced form, e.g. laN0 [: raN0]. We then listened through the tape once more in order to mark all pauses: short (#) and long (###) and extra-long vowels, e.g. maa:4. Repetitions and re-tracings were marked using the CHAT conventions, i.e. the repeated material was surrounded by <> and followed by [/], [//] or [///].

Following the CHAT convention, each main line was made to include only one utterance – defined with a combination of phonetic and grammatical criteria. Thus, a line/utterance ends when both conditions are met:
1. There is short pause (#), a long pause (###), or a “vowel lengthening”, and
2. This coincides with the end of a clause, marked as [c].

If only (1) is met, the pause is marked within the utterance/line. If only (2) is met, [c] marks the end of the clause but not the utterance/line. However, we sometimes allow a line/utterance to end even if there is a word between the pause and clause boundary.

The operational definition of a clause provided by Berman and Slobin (1994:660) (“a unit that contains … a predicate that expresses a single situation (activity, event, state). Predicates include finite and non-finite verbs, as well as predicative adjectives.”) could not be used since serial verb constructions, which are ubiquitous in Thai, can involve up to six verbs in what is arguably a representation of a “single situation”. While knowing that our decisions will not satisfy all scholars of Thai linguistics, we have tried hard to provide a clear set of criteria for identifying clause boundaries in the corpus, and by marking them with the conventional CHAT symbol [c], allow a first-order surface representation of grammatical complexity. According to our criteria [c] was used:
1. Before the introduction of a new explicit or implicit subject.
2. Before the relative-clause markers (RCMs) thii2 and sUN2 (‘which’). If there is only a noun phrase between the previous [c] and the RCM, the clause boundary [c] is instead placed at the end of the relative clause. It is also used in other places where a relative-clause marker may be inserted.
3. Where clause boundaries are indicated by the presence of conjunctions such as lx3, lxxw3 (‘and’), lxxw3 kO2 (‘and then’), kO2 (‘then’), thxx1 (‘but’), phrO3 (‘because’), mUa2, phOO0 (‘when’), con0 (‘until’), mxx3 (‘though’) or in other places where a conjunction may be inserted.
4. After wa2 (‘that’), if it is both preceded and followed by text segments with main verbs (excluding cases where wa2 is a main verb, and where it has nouns and other non-verb expressions as complements);

Given this way of marking clauses, we were faced with a dilemma as to how to satisfy the CHAT convention of having only one utterance per line (tier). If we chose only phonetic criteria, i.e. pauses and intonation contours, to define utterance boundaries, we would have to break up clauses – as previously defined – into many lines, and thus decrease readability and analyzability. On the other hand, if we neglected prosodic criteria and only segmented the text into clauses, we would miss the information that some clause boundaries coincided with pauses, and thus seemed to constitute processing units, while others did not. We resolved this dilemma with a compromise, operationally defining an “utterance”, i.e. the unit to be placed on a single line/tier, through a combination of phonetic and grammatical criteria:

U1. An utterance boundary (.) occurs when there is BOTH a phonetic indication of utterance closure – a short pause (#), a long pause (##) or a vowel lengthening (:) – AND a clause boundary, marked with [c].

This means that if there is only a pause but no clause boundary, then the pause is marked within the utterance. If the utterance seems to terminate without a clause being completed – the speaker “trails off” – this is marked by ending the line with the symbol “+…” instead of the utterance delimiter “.”. Likewise if there is a clause boundary, but no pause of vowel lengthening, the utterance is assumed to continue until a clause boundary and an utterance boundary coincide. An exception of the condition U1 was made when there was only a single word (or short phrase) serving as a “filler” between the clause boundary and the actual pause, and in this case the utterance was terminated after this filler, as in the following authentic example:
lxxw3 dek1 tUUn1 khU2 maa0 # phroOm3 maa4 [c] lxxw3 #. and child wake up come together dog and ‘And the child woke up, with the dog and.’

Finally, each of the 50 narratives was read though once again by at least two different checkers, correcting for any inconsistencies. Furthermore, a listing of all the words in the corpus was produced using the CLAN command freq +k *.cha +u +r6, and we went through this list word by word, making sure that each word is transcribed consistently throughout the corpus. In addition to standard CHAT codes, we used the %tai dependent tier for the Thai transcription and a double ++ to indicate reduplication as in “luuk2++luuk2.”

3.6.4 Files

The files are summarized in the following table. The subjects’ names do not appear in the transcripts.

Table 4: Thai Frog Files
<table>
<thead>
<tr>
<th>File</th>
<th>Age</th>
<th>Sex</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4a</td>
<td>4;3.8</td>
<td>f</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>4b</td>
<td>3;11.20</td>
<td>f</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>4c</td>
<td>4;4.4</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>4d</td>
<td>3;10.12</td>
<td>f</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>4e</td>
<td>3;11.16</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>4f</td>
<td>4;0.2</td>
<td>f</td>
<td>8-SEP-2000</td>
</tr>
<tr>
<td>4g</td>
<td>3;11.19</td>
<td>m</td>
<td>8-SEP-2000</td>
</tr>
<tr>
<td>4h</td>
<td>3;6.15</td>
<td>m</td>
<td>8-SEP-2000</td>
</tr>
<tr>
<td>4i</td>
<td>3;11.2</td>
<td>m</td>
<td>8-SEP-2000</td>
</tr>
<tr>
<td>4j</td>
<td>3;11.22</td>
<td>f</td>
<td>8-SEP-2000</td>
</tr>
<tr>
<td>6a</td>
<td>6;02.15</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6b</td>
<td>5;11.23</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6c</td>
<td>5;6.01</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6d</td>
<td>5;06.11</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6e</td>
<td>5;08.18</td>
<td>f</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6f</td>
<td>6;04.1</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6g</td>
<td>5;10.10</td>
<td>f</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6h</td>
<td>5;06.25</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6i</td>
<td>5;11.21</td>
<td>m</td>
<td>18-FEB-2000</td>
</tr>
<tr>
<td>6j</td>
<td>6;01.17</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9a</td>
<td>8;7.8</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9b</td>
<td>9;0.21</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9c</td>
<td>8;10.27</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9d</td>
<td>9;0.11</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9e</td>
<td>9;2.0</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9f</td>
<td>9;3.2</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9g</td>
<td>8;10.8</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9h</td>
<td>8;7.0</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9i</td>
<td>9;0.3</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>9j</td>
<td>9;7.2</td>
<td>f</td>
<td>2-FEB-2000</td>
</tr>
<tr>
<td>11a</td>
<td>10;10.16</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11b</td>
<td>10;05.22</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11c</td>
<td>11;1.5</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11d</td>
<td>11;4.20</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11e</td>
<td>11;2.24</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11f</td>
<td>10;8.27</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11g</td>
<td>10;3.22</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11h</td>
<td>11;0.16</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11i</td>
<td>10;4.24</td>
<td>f</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>11j</td>
<td>11;3.18</td>
<td>m</td>
<td>16-FEB-2000</td>
</tr>
<tr>
<td>20a</td>
<td>22;8.0</td>
<td>f</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20b</td>
<td>35;06.0</td>
<td>f</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20c</td>
<td>21;7.0</td>
<td>f</td>
<td>15-APR-2000</td>
</tr>
</tbody>
</table>
If you publish any paper based on this data, please send an MS-Word or PDF-formatted version of your paper as an attachment to jordan_zlatev@lucs.lu.se. Users of these data should cite:


<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20d</td>
<td>22;11.0</td>
<td>f</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20e</td>
<td>22;8.0</td>
<td>f</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20f</td>
<td>22;11.30</td>
<td>m</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20g</td>
<td>29;01.20</td>
<td>m</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20h</td>
<td>20;07.11</td>
<td>m</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20i</td>
<td>23;10.27</td>
<td>m</td>
<td>15-APR-2000</td>
</tr>
<tr>
<td>20j</td>
<td>19;06.06</td>
<td>m</td>
<td>15-APR-2000</td>
</tr>
</tbody>
</table>
4. Gopnik

Myrna Gopnik
Department of Linguistics
McGill University
1001 Sherbrooke W.
Montréal, PQ H3A 1G5 Canada

This directory contains data that were contributed by Myrna Gopnik of McGill University to the CHILDES in August of 1988. They include story book descriptions from normal children between the ages of 2 and 5.

The file names use this syntax:

- **Storytype**: f (free) or p (prompted) or q (questionnaire) or g (game)
- **StudentID**: 3-digit number
- **Booktype**: 1 (free) 2-5 (one of four books)
- **Session**: 1, 2, or 3 (which test session)

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Birthdate</th>
<th>Teacher</th>
<th>Verified</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>F</td>
<td>27-10-78</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>02</td>
<td>F</td>
<td>12-08-78</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>03</td>
<td>F</td>
<td>01-07-79</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>04</td>
<td>F</td>
<td>10-02-79</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>05</td>
<td>M</td>
<td>29-11-78</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>06</td>
<td>M</td>
<td>23-10-78</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>07</td>
<td>M</td>
<td>31-08-78</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>08</td>
<td>F</td>
<td>17-08-79</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>09</td>
<td>M</td>
<td>06-12-79</td>
<td>BL</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>13-12-79</td>
<td>BL</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>14-11-80</td>
<td>LD</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>19-01-81</td>
<td>LD</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>26-05-81</td>
<td>BL</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>25-07-80</td>
<td>BL</td>
<td>Y</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>05-01-80</td>
<td>BL</td>
<td>Y</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>20-08-79</td>
<td>LS</td>
<td>Y</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>29-03-79</td>
<td>LS</td>
<td>N</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>19-02-80</td>
<td>BL</td>
<td>Y</td>
</tr>
<tr>
<td>46</td>
<td>F</td>
<td>21-01-81</td>
<td>LD</td>
<td>Y</td>
</tr>
<tr>
<td>47</td>
<td>M</td>
<td>02-04-80</td>
<td>HD</td>
<td>N</td>
</tr>
<tr>
<td>48</td>
<td>M</td>
<td>01-03-81</td>
<td>LD</td>
<td>N</td>
</tr>
<tr>
<td>49</td>
<td>F</td>
<td>17-02-82</td>
<td>LG</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>------</td>
<td>----</td>
</tr>
<tr>
<td>50</td>
<td>F</td>
<td>01-06-80</td>
<td>XY</td>
<td>N</td>
</tr>
<tr>
<td>51</td>
<td>F</td>
<td>09-10-79</td>
<td>BL</td>
<td>N</td>
</tr>
<tr>
<td>52</td>
<td>M</td>
<td>18-02-82</td>
<td>SG</td>
<td>N</td>
</tr>
<tr>
<td>53</td>
<td>F</td>
<td>20-04-82</td>
<td>SG</td>
<td>N</td>
</tr>
<tr>
<td>54</td>
<td>M</td>
<td>28-01-80</td>
<td>BL</td>
<td>N</td>
</tr>
<tr>
<td>55</td>
<td>F</td>
<td>01-01-80</td>
<td>XY</td>
<td>N</td>
</tr>
<tr>
<td>56</td>
<td>M</td>
<td>12-10-80</td>
<td>BL</td>
<td>N</td>
</tr>
<tr>
<td>57</td>
<td>F</td>
<td>19-03-82</td>
<td>LG</td>
<td>N</td>
</tr>
<tr>
<td>58</td>
<td>M</td>
<td>28-05-80</td>
<td>BL</td>
<td>N</td>
</tr>
<tr>
<td>59</td>
<td>F</td>
<td>06-03-81</td>
<td>LD</td>
<td>N</td>
</tr>
<tr>
<td>60</td>
<td>M</td>
<td>31-08-82</td>
<td>LG</td>
<td>Y</td>
</tr>
<tr>
<td>61</td>
<td>F</td>
<td>11-03-82</td>
<td>XY</td>
<td>N</td>
</tr>
</tbody>
</table>

Publications using these data should cite:

5. Hicks

Deborah Hicks  
Department of Educational Development  
College of Education  
University of Delaware  
Newark, DE 19716 USA  
hicks@brahms.udel.edu

The narratives in this directory were collected by Deborah Hicks in the context of a study of primary school children’s narrative genre skills, focusing on their ability to produce a range of kinds of narratives. In the study, children from three primary grade levels — first, second, and fifth — were shown a shortened version of the silent film, “The Red Balloon.” After viewing the film, children were asked to tell the film’s events in three different ways: as a factual news report, an ongoing event case, and as a more embellished story. These three narrative genres are representative of what Heath (1983) terms “key” narratives, or narratives that are found crossculturally in children’s language learning environments. The narrative data were coded by utterances for linguistic forms that might mark genre differences.

This directory contains four subdirectories: 1st, 2nd, 5th, and del. The first three are taken from first, second, and fifth graders in Cambridge, Massachusetts. The fourth is taken from a lower-class SES group in Delaware. For comparison with the Delaware children, these 12 files in the 1st grade directory were used: 4, 5, 12, 16, 27, 29, 30, 35, 38, 39, 40, and 42.

The children with files in the subdirectories called 1st, 2nd, and 5th were first grade, second grade, and fifth grade students in a private elementary school in Cambridge, Massachusetts. The majority of students attending this school were members of middle class families in which one or both parents were working professionals, so that these children could be considered members of mainstream culture. The classrooms were somewhat progressive in nature, so that children were free to choose from a range of activities those that they would work on. Many of the activities that children performed regularly were language activities, such as recounting on tape a story of how the world was created, writing about “what we did in science class,” and recounting personal experiences during sharing time episodes. The narrative genre tasks were thus presented to the children as one of the many options available, and in all but a few cases, children were willing and eager to leave the room for the tasks.

Before performing any of the narrative tasks, children were told that they would watch a film and would then tell what happened in the film in three different ways. In the case of the online narration task, children listened to the experimenter saying “This is [child’s name] and Deborah, sportscasters, and we’re gonna say everything we see happening in the film. I’m gonna start off and then [child’s name] is gonna take over.” The child then watched the 3-minute segment of the film and then the experimenter
started the narration by saying “The little boy and the red balloon are going past a church steeple. And they’re coming to a bakery shop. The little boy is looking inside the bakery shop. Now he’s checking in his pocket to see if he has enough money to buy something to eat. Looks good. Now he’s walking into the bakery shop.” Then the experimenter turned to the child and asked, “Can you take over now and be the sportscaster?” The order of the report and event cast tasks was randomly selected within grade levels.

The storytelling task was performed separately from the report and event cast, in a session that took place approximately 1 hour after the completion of the first two tasks. This particular research design was chosen on the grounds that performance of three consecutive tasks would be too demanding for many of the children in the study, particularly the 5-year-old children. For the storytelling task, the leading given by the experimenter was “This is [child’s name] and Deborah, and we’re gonna be storytellers and tell the story of The Red Balloon. I’m gonna start off and then [child’s name] is gonna take over.” During this, the experimenter holds a “storybook” which has on the front cover a picture from the film but which has neither words nor pictures inside. The experimenter then says “The Red Balloon. Once upon a time there was a little boy who lived in Paris, France. One day, on his way to the bus stop, he found this big beautiful red balloon. He wanted the balloon to be his friend.” Then the experimenter turns to the child and says, “Can you take over now and be the storyteller?” At this point, the experimenter passes the storybook to the child.

In an attempt to create some degree of homogeneity in the data, in addition to providing an interaction with the highest possible degree of ecological validity, children were provided with a great deal of contextual support for the tasks. As was noted in the introductory section to this chapter, children were reminded before each task of the particular narrative “voice” they were to assume: that of a news reporter, a sportscaster, or a storyteller. For the storytelling task, children were also given a storybook containing only a single picture on the outside cover, which they were encouraged to hold during the story narration.

The data obtained from the study were transcribed in CHAT and analyzed using the CLAN programs for child language analysis. The entire narration was divided into clauses with one clause on each line of the CHAT transcript. The segmentation of the narrative data was done on the basis of clause units of analysis, following Berman and Slobin (1994). Clause units were defined as any linguistic utterance containing a predicate, so that the following would all be considered separate units of analysis: “he climbed up the stairs”, “when the boy was inside the bakery shop,” and “the boys who stole the balloon.” Segmentation of complement clauses was done so that the utterance “he saw the balloon floating by the door” was segmented into “he saw” and “the balloon floating by the door.” Utterances containing verbs with three arguments such as “he told the balloon to stay by the door” were segmented as “he told the balloon” and “to stay by the door.”

The narrative data were examined in terms of children’s use of specific linguistic forms representing three basic subsystems: a) syntactic constructions, b) temporal and
event expressions, and c) indexical clauses. The analysis of syntactic constructions was designed to assess possible genre differences in the syntactic complexity of event casting, reportative, and story narratives. The analysis of expressions of temporality and event relations was designed to examine genre differences in how temporal and logical relations between events were expressed in the discourse. Finally, the analysis of indexical clauses was an attempt to assess genre differences in how children went beyond the basic narration tasks to provide evaluative or descriptive information about events in the narrative.
5.1 Coding

The codes for assessing genre differences in children’s narratives consists of three basic sections representing different linguistic subsystems. The codes for syntax are drawn from Berman and Slobin (1994) and Quirk and Greenbaum (1972). The codes are designed to assess differences in syntactic complexity among the three narrative genres. The codes for temporality and event relationships are drawn primarily from Berman and Slobin (1994) and are designed to assess genre differences in the use of verb forms, aspectual markers, timemarkers and logical connectors. Finally, the measures for indices and intensifiers are designed to assess differences in the marking of nonmainline event clauses or in the embellishment of narrative clauses through the use of highlighters and intensifiers.

5.1.1 Syntactic Codes

IND independent clause  
SADV subordinate adverbial clause “holding his balloon”  
CMP complement clause  
REL relative clause  
QES questions  
DIA direct dialogue, quoted from characters  
NEG negation

5.1.2 Event Codes

T:SQ temporal sequential marker  
T:CN temporal connective  
T:ADV temporal adverb  
MOD modals  
MODV modal verbs  
HYP hypothetical statement  
FUT future constructions  
L:CONT one event is contingent on another, as in “when X, they Y”  
L:TR one event is contrasted with another  
CES concessive  
PUR purpose  
REA reason  
RES result  
C:TEM clauses in temporal sequence  
C:AD clauses in adversive relationship  
C:REA clauses in reason relationship
### 5.1.3 Indexical Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I:DES</td>
<td>indicator of description</td>
</tr>
<tr>
<td>I:COM</td>
<td>indicator of comparison</td>
</tr>
<tr>
<td>I:S</td>
<td>indicator of internal state</td>
</tr>
<tr>
<td>I:S:EMO</td>
<td>indicator of emotional state</td>
</tr>
<tr>
<td>I:S:PHY</td>
<td>indicator of physical state</td>
</tr>
<tr>
<td>I:S:MEN</td>
<td>indicator of mental state verb</td>
</tr>
<tr>
<td>I:E</td>
<td>indicator of evaluation</td>
</tr>
<tr>
<td>I:O</td>
<td>indicator of orientation</td>
</tr>
<tr>
<td>I:MET</td>
<td>indicator of metaphor or simile</td>
</tr>
<tr>
<td>I:META</td>
<td>indicator of metacomments</td>
</tr>
<tr>
<td>I:TNG</td>
<td>indicator of tangential remarks</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E:DES</td>
<td>event descriptions</td>
</tr>
<tr>
<td>E:HAB</td>
<td>habitual events</td>
</tr>
<tr>
<td>E:I:EMO</td>
<td>internal state events</td>
</tr>
<tr>
<td>E:I:MEN</td>
<td>internal mental states</td>
</tr>
<tr>
<td>E:I:PHY</td>
<td>internal physical states</td>
</tr>
<tr>
<td>C:MET</td>
<td>commentary involving metaphors or similes</td>
</tr>
<tr>
<td>C:COM</td>
<td>commentary on narration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EX:ADJ</td>
<td>adjectival expansions</td>
</tr>
<tr>
<td>EX:ADV</td>
<td>adverbial expansions</td>
</tr>
<tr>
<td>INTN</td>
<td>intensifiers</td>
</tr>
<tr>
<td>STR</td>
<td>stress</td>
</tr>
<tr>
<td>P</td>
<td>pitch marking</td>
</tr>
</tbody>
</table>

Publications using these data should cite:


Additional references include:


These data were gathered in English, Italian, and Hungarian with adults and children aged 3, 4, and 5 years using the picture description task of MacWhinney and Bates (1978). In the 1978 experiment, there were 120 participants: 40 Americans, 40 Hungarians, and 40 Italians. Within each language community, there were ten 3-year-olds, ten 4-year-olds, ten 5-year-olds, and ten adults. The chief focus of attention was on the development in the 3–6 year period. The adult participants were included as controls to see if any further developmental changes might be present after age 6 in use of these devices. Each group of 10 participants included five females and five males. The children were enrolled in nursery schools in Denver, Budapest, and Rome. There is every reason to believe that the children at each age were generally equal in terms of overall linguistic ability, because they were all normal, middle-class members of the majority culture and all resided in large metropolitan areas within what is commonly known as Western culture. Unfortunately, no cross-culturally valid measure of general linguistic ability is yet available.

Before a participant was tested, the pictures were placed into the order in which they were to be administered. The order of the nine series of pictures within each series was also randomized. Following each series, a picture of a common object such as a bottle or a sailboat was inserted. This was done to break up any set (Einstellung) effects. Participants were examined individually. Each participant was first seated next to the experimenter at a table. The participant was told that he or she would be asked to describe some pictures. Adults were told to describe the pictures in a simple direct fashion. The experimenter showed the pictures to each participant one at a time in the sequence determined by the randomization procedure. Two probes were used: “Tell me about this picture” and “What’s happening in this picture?” Use of the two probes was also randomized. Each session was taperecorded in its entirety.

Publications using these data should cite:
7. MacWhinney / Bates 2

This directory contains the Hungarian and English data from the film description study that Brian MacWhinney and Elizabeth Bates conducted during 1980 to 1982. The results of this work were never published in full. The Italian data from the study are not yet included. There were two orderings of the film — order A and order B. There was also a single version of a cartoon film. Participants came from four age groups: 3, 6, 10, and adult. The file names give the age of the participant, which version of the film they saw, and their participant number.
### 7.1 Script for Pixolation Film

* = line for coding summary data

<table>
<thead>
<tr>
<th>Scene</th>
<th>Seg</th>
<th>Script A</th>
<th>Script B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>1</td>
<td>Hippo turns. (zoom and cut)</td>
<td>Kangaroo turns. (zoom and cut)</td>
</tr>
<tr>
<td>2*</td>
<td></td>
<td>Hippo, ladder, and lock on stage.</td>
<td>Kangaroo, orange, and pencil on stage.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Hippo moves.</td>
<td>Orange moves.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Hippo hits ladder.</td>
<td>Orange hits kangaroo.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Ladder moves.</td>
<td>Kangaroo moves.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Ladder hits lock.</td>
<td>Kangaroo hits pencil.</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Lock moves.</td>
<td>Pencil moves.</td>
</tr>
<tr>
<td>1b</td>
<td>1</td>
<td>Lion turns. (zoom and cut)</td>
<td>same as A</td>
</tr>
<tr>
<td>2*</td>
<td></td>
<td>Lion, spool, and ball on stage.</td>
<td>same as A</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Lion moves.</td>
<td>same as A</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Lion hits spool.</td>
<td>same as A</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Spool moves.</td>
<td>same as A</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Spool hits ball.</td>
<td>same as A</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Ball moves.</td>
<td>same as A</td>
</tr>
<tr>
<td>1c</td>
<td>1</td>
<td>Kangaroo turns. (zoom and cut)</td>
<td>Hippo turns. (zoom and cut)</td>
</tr>
<tr>
<td>2*</td>
<td></td>
<td>Kangaroo, orange, pencil on stage.</td>
<td>Hippo, ladder, and lock on stage.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Orange moves.</td>
<td>Hippo moves.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Orange hits kangaroo.</td>
<td>Hippo hits ladder.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Kangaroo moves.</td>
<td>Ladder moves.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Kangaroo hits pencil.</td>
<td>Ladder hits lock.</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Pencil moves.</td>
<td>Lock moves.</td>
</tr>
<tr>
<td>2a</td>
<td>1*</td>
<td>Bottle, goat, and lock on stage.</td>
<td>Block, pipe, and dog on stage.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Bottle moves.</td>
<td>Block moves.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Bottle hits goat.</td>
<td>Block hits pipe.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Goat moves.</td>
<td>Pipe moves.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Goat hits lock.</td>
<td>Pipe hits dog.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Lock moves.</td>
<td>Dog moves.</td>
</tr>
<tr>
<td>2b</td>
<td>1*</td>
<td>Block, pipe, and dog on stage.</td>
<td>Bottle, goat and lock on stage.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Block moves.</td>
<td>Bottle moves.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Block hits pipe.</td>
<td>Bottle hits goat.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Pipe moves.</td>
<td>Goat moves.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Pipe hits dog.</td>
<td>Goat hits lock.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Dog moves.</td>
<td>Lock moves.</td>
</tr>
<tr>
<td>2c</td>
<td>1*</td>
<td>Giraffe, stool, and tree on stage.</td>
<td>same as A</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Giraffe moves.</td>
<td>same as A</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Giraffe hits stool.</td>
<td>same as A</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Stool moves.</td>
<td>same as A</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Stool hits tree.</td>
<td>same as A</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Tree moves.</td>
<td>same as A</td>
</tr>
</tbody>
</table>
3a 1* Cow, basket, and table on stage. Chair, fish, and TV on stage.
2 Cow moves. Chair moves.
3 Cow hits basket. Chair hits fish.
4 Basket moves. Fish moves.
5 Cow moves. Chair moves.
6 Cow hits (pushes) table. Chair hits TV.
7 Table moves. TV moves.

3b 1* Chair, fish, and TV on stage. Flowerpot, ball, and pig on stage.
2 Chair moves. Flowerpot moves.
3 Chair hits fish. Flowerpot hits ball.
4 Fish moves. Ball moves.
5 Chair moves. Flowerpot moves.
6 Chair hits TV. Flowerpot hits pig.
7 TV moves. Pig moves.

3c 1* Flowerpot, ball, pig on stage. Cow, basket, and table on stage.
2 Flowerpot moves. Cow moves.
3 Flowerpot hits ball. Cow hits basket.
4 Ball moves. Basket moves.
5 Flowerpot moves. Cow moves.
6 Flowerpot hits pig. Cow hits table.
7 Pig moves. Table moves.

4a 1 Cylinder on stage. scene only appears in A
2 Ball moves.
3 Ball hits cylinder.
4 Cylinder falls.

4b 1 Cylinder on stage. same as A
2 Ball moves.
3 Ball hits cylinder.
4 (Cylinder falls.)

4c 1 Cylinder on stage. same as A
2 Ball enters.
3 Ball misses cylinder.
4 Cylinder doesn’t fall.
5 Ball exits.

4d 1 (Ball enters.) same as A
2 (Ball hits cylinder.)
3 (Cylinder falls.)
4 Cylinder lying down.
5 Ball exits.

4e 1 (Ball misses cylinder.) same as A
2 (Cylinder doesn’t fall.)
3 Cylinder stands.
4 Ball moves.

5 1* Man, stick, and ball on stage. same as A
2 Man looks at stick and ball.
3 Ball moves.
4. Ball hits stick.
5. Stick falls over.
6. Man looks (during action).

7a. 1* Gorilla and camel on stage. Camel and elephant on stage.
2. Gorilla chases camel. (cut) Camel chases elephant. (cut)

7b. 1* Gorilla and stag on stage. Deer and elephant on stage.
2. Gorilla chases stag. (cut) Deer chases elephant. (cut)

7c. 1* Gorilla and elephant on stage. Gorilla and elephant on stage.

8a. 1* Apple and stool on stage. Apple and stool on stage.
2. Apple chases stool. (cut) Apple chases stool. (cut)

8b. 1* Apple and dog on stage. Stool and dog on stage.

8c. 1* Cow and zebra on stage. Cow and zebra on stage.
2. Cow chases zebra. (cut) Cow chases zebra. (cut)

8d. 1* Zebra and seal on stage. Cow and seal on stage.

8e. 1* Camel and giraffe on stage. Camel and giraffe on stage.
2. Camel chases giraffe. (cut) Camel chases giraffe. (cut)

8f. 1* Bottle and camel on stage. Bottle and giraffe on stage.

8g. 1* Chair and table on stage. Chair and table on stage.
2. Chair chases table. (cut) Chair chases table. (cut)

8h. 1* Hippo and chair on stage. Hippo and table on stage.

8i. 1* Cow and apple on stage. Cow and apple on stage.

9a. 1 (Big) orange on plate. (Big) orange on plate.
2. (Big) orange moves. (cut) (Big) orange moves. (cut)

9b. 1 (Big) oranges on stage. Big and small orange on stage.
2. One (big) orange moves. (Big) orange moves.
3. (Big) orange hits (big) orange. (Big) orange hits (small) orange.
4. (Big) orange moves. (Small) orange moves.
5. (Big) orange stays. (cut) (Big) orange stays. (cut)

9c. 1 (Big) orange on plate. (Big) orange on plate.

10a. 1 Man and woman face each other. Man and woman face each other.
2. Man has ball. Woman has ball.
4. Man hands woman a ball. Man grabs ball from woman.

10b1. 1 Woman sits down. (cut) Woman sits down. (cut)
10b2. 1 Woman and man stand facing. (cut) Woman sitting; man facing. (cut)
2. Woman moves to man (cut) Man moves to woman. (cut)

10b3. 1 Camera behind woman. Camera behind man.
2. Woman grabs ball. Man gives ball to woman.

10c1. 1 Woman skips. (cut) scene omitted
10c2. 1 Woman and man stand facing. scene omitted
2 Woman moves to man.  scene omitted
10c3 1 Camera behind woman.  scene omitted
2 Woman gives ball to man.  scene omitted
10d1 1 Woman reads a book. (cut)  Woman reads a book. (cut)
10d2 1 Woman stands alone.  Woman stands alone (cut).
2 Camera zooms in.  omitted
3 Woman walks, raises hand. (cut)  omitted
10d3 1 Man and woman standing close.  same as A
2 Man drops orange in her hand.  same as A
11a 1 Man 1 and man 2 standing.  same as A
2 Man 1 paints himself.  Man 2 bounces a ball.
3 Man 2 bounces a ball.  Man 1 paints himself.
4 Man 1 climbs a ladder.  same as A
11b 1 Woman 1 and 2 are standing.  same as A
2 Woman 1 turns around.  Woman 2 bounces a ball.
3 Woman 2 bounces a ball.  Woman 1 turns around.
4 Woman 1 sits down.  same as A
12a 1 Tree pulls active walrus.  Tree chases panther.
12b 1 Tree circles alligator.  Tree pulls camel. (cut)
12c 1 Tree pushes active penguin.  Tree circles alligator.
12d 1 Tree chases panther.  Tree pushes sheep.
12e 1 Tree falls on bear.  Tree pulls active walrus. (cut)
12f 1 Tree pushes neutral sheep.  Tree falls on bear.
12g 1 Tree hits gorilla.  Tree pushes active penguin.
12h 1 Tree pulls neutral camel.  Tree hits gorilla.
12i 1 Buffalo pushes active tree.  Buffalo pushes passive tree.
12j 1 Buffalo pushes neutral tree.  Buffalo pushes active tree.
13a 1* (Lemon, apple, orange on stage.)  same as A
2 Lemon moves.
3 Lemon hits apple.
4 Apple moves.
5 Apple hits orange.
6 Orange moves.
13b 1* (Lemon, apple, orange on stage.)  same as A
2 Lemon moves.
3 Apple and orange on stage.
4 Lemon moves.
5 Lemon hits apple.
6 Apple moves.
7 Apple hits orange.
8 Orange moves.
13c 1* Apple and orange on stage.  same as A
2 Lemon moves.
3 Lemon hits apple.
4 Apple moves.
5 Apple hits orange.
6  Orange moves.
13d 1*  Apple and orange on stage.  same as A
  2  Lemon moves.
  3  Orange, lemon, apple on stage.
  4  Lemon moves.
  5  Lemon hits apple.
  6  Apple moves.
  7  Apple hits orange.
  8  Orange moves.
13e 1*  Lemon, apple, orange on stage.  same as A
  2  Lemon moves.
  3  Lemon hits apple.
  4  Apple moves.
  5  Apple hits orange.
  6  Orange moves.

Film A order was 8a–d, 11a, 8e–f, 13c, 8g–h, 10a, 8i, 4a-e, 10c, 12, 11b, 2, 13e, 7, 13b, 1, 10d, 9, 13d, 5, 10b, 3, 13a.

Film B order was 8a–d, 13a, 8e–f, 11b, 8g–h, 4b–e, 8i, 10a, 3, 13e, 10b, 5, 13c, 10d, 9, 1, 13d, 7, 13b, 2, 11a, 12.
7.2 Reasons for Stimuli

1. The treatment of the orange as an animate should be strongest in Film A where focus is achieved cinematographically. Focus on the kangaroo should induce more backgrounding of the hitting by the orange.
2. It is interesting to compare “bottle” with “orange” in (1) and to compare across versions for effects of the previous priming scene. Perhaps “dog” will be passivized?
3. Here we might expect conjunctions instead of relatives and fewer passives than in (1) and (2). Do children have a hard time “letting go” of the fish when it is already the perspective?
4. These should elicit forms such as “will hit,” “will miss,” “hit,” and “missed.”
5. Backgrounding of “man,” cause inferred in (3), multiple setting in ASL.
6. Omitted.
7. This should elicit devices that mark anaphoric givenness and singularity of newness.
8. These should be marked by coordination in compound presentation, givenness of distant items, with possible referent confusion.
9. Set operation, confusability.
10A. Segment a should elicit “gives”, whereas segment b should elicit less “gives” than 10B. Segments a, c, and d should elicit “gets.”
10B. Segment a should elicit “takes” and “grabs.” In segment b “gives” should be stronger than in 10A. Segments a and d should elicit “gets.”
11. Relative clauses, and so forth.
13. Stage setting, ASL marking, shifts.
7.3 Devices Used Across Segments

1. Scene contrast: Contrast with a whole previous scene. Examples: *in this scene, however; and now in this scene*. Code on each segment of the entire scene.

2. Scene parallelism: Parallel to a previous scene. Examples: *exact same thing as before, but with a goat; the same black ball hit the orange stick again*. Code on each segment of the entire scene.

3. Movie scene specification: May apply to either a segment or a whole scene. Code on segment on which occurs. Examples: *in this episode, scene, shot, in the picture, in this one, here in the second one.*

4. Global description: Example: *fruit rolling around.*
7.4 Devices Used in Particular Segments

1. Segment ellipsis: No material in the segment at all.
2. Interactionals. Examples: OK, allright, well, um, and so forth. Code on segment if occurs anywhere within the segment.
3. Evidentials: References to the perceiver. Examples: we see that, it appears, we can see, you see a, and so forth.
4. Location specification: Physical location within the “room” or frame depicted. Examples: in the room, on the table, it left the screen, off stage, came on stage, and so forth. Do not code plain off, in, or away as in then the cow went in/off/away. Although the latter may imply location, they are not very specific and are already captured as the main function of the directional adverb category. Also, do not code the presentative (e.g., there is a dog), because it may imply existence rather than location; however, do code utterances such as the dog was there, because these are locative and not captured by any other category. Note that this category should not be confused with pansegmental #3, which refers to the contents of a segment or scene as a whole. For instance, in this scene, a dog chased a cow would be coded #3 pansegmental, because elements of the scene are not assigned locations.
5. Segment parallelism: Like #2 pansegmental, but referring only to segment rather than scene similarity. The referents involved will determine which coding is applicable. Examples: in both of these, the same as, the exact same thing, in the same sequence. More specifically, an example would be the hippo hit the ladder and then did the same to a lock — the referents being parallel within multiple segments of a single scene. Code parallelism on the latter segment (e.g., hippo hit lock).
6. Segment contrast: Analogous to #1 of pansegmental, but referring to segment rather than scene contrast. The referents involved will determine which coding is applicable. Examples: here, however, in this case, in turn, and now. (The very common phrase and then has been double-coded as #7 and #8. Although such a combination might conceivably be considered a method of segment contrast, it was not so regarded here. Instead it was considered to be primarily a mode of coordination and ordering.)
7. Segment order: Examples: then, first, and so forth. May be double-coded #3 pansegmental, but only when order also suggests a movie scene, e.g., in the first one/in the second scene. In these cases a noun or pronoun must indicate reference to the scene.
8. Coordination: Example: and. Code when occurs within a segment (e.g., a cow and an apple were there. An exception would be for adjectival coordinations because they are script-external (e.g., a brown and white cow hit the ladder. When coordination occurs between segments code on the following segment. In the case of acow ran and it hit an apple, the cow hit apple segment would be coded as #8).
10. Restrictive relativization relevant to script: Relativizations that must discriminate among two similar scripted objects. Example: *the man who painted his face* when one man did and one man did not.

11. Nonrestrictive relativization relevant to script: Example: *the apple hit the lock which went off stage.* Code under the segment *lock moves.* Causatives such as *the apple hit the lock which made the lock go off stage* were not coded here because the relativizer “which” refers to the contents of the entire previous phrase, rather than a single noun. These causatives are captured by #9 of Verbal Specification.

12. Segment negation: Examples: *it didn’t hit the cylinder; you don’t see if it falls or not.* Do not code nominal negation here.

13. Backward conflater: Transitive action conflating previous intransitive motion. Example: *the apple rolled into the orange* implies previous movement of the apple. Code the segment *apple hits orange* as a backward conflater even if the previous segment is not ellipsed. Other likely backward conflaters are *runs into* and *bumps (into).*

14. Backward conflatee: Segment ellipsed as indicated in #13. Also code as #1 (ellipsed). Do not code as a conflatee unless actually ellipsed.

15. Forward conflater: A transitive that conflates a following intransitive motion. Example: *the goat knocked over the bottle* on the segment *goat hits bottle* implies movement in the following *bottle moves* segment. Other likely forward conflaters include *pushed over, push,* and *nudge.* Code as a conflater even if the following segment is not ellipsed. For some speakers, bump seems to have the potential to be considered as both a backward and forward conflater. It was coded only as a backward conflater because it necessarily implies some movement on the part of the one who bumps. It may or may not imply a substantial degree of movement (although it may tend to imply recoil or surprise) on the part of the bumpee. Other somewhat ambiguous cases include *the goat knocked into the bottle, the apple pushed into the orange.* The unusual pronouns may tend to move these in the direction of backward conflation: they were nonetheless treated as forward conflaters, that is, as if they were more typical usages of these verbs.

16. Forward conflatee: Segment actually ellipsed by #15. Code only if ellipsed and double code for ellipsis (#1).

17. Double forward conflation: Examples: *the giraffe pushes the chair into the lock.* Three segments are involved: “giraffe hits chair”, ”chair moves”, “chair hits lock”. Code #17 on “chair hits lock.” Code #15 and #16/#1 on “giraffe hits chair” and “chair moves,” respectively.

18. Error: The purpose of this code is to identify those segments that seem to be difficult to perceive. Examples: Role errors (example, *the dog pushes the cow* when the cow has pushed the dog) and errors that make it impossible to tell what roles are assigned. Also apply this codeto major nominal errors that are not role errors (example, *the dog pushes the cow* when there is no dog in the scene, and a giraffe has pushed the cow). Do not use this code for minor nominal errors with no impact (e.g., calling a horse a donkey). Also use this code for errors requiring prompting (indicated on transcripts as empty parentheses).

19. Self-correction: Retraced false starts (not hesitations, interactionals, or simple repeats). Use the corrected production for all other coding.
20. Out of script order: Whole segments are out of order. This is indicated in the transcripts by segment numbers that are out of sequence.
7.5 **Nominal Specification Devices**

1. **Ellipsis:** Any reference to the nominal (via noun, pronoun, and so forth) is missing. This is also used when the nominal is incorrect and additional coding would be misleading.
2. **Definite article:** *the.*
3. **Indefinite article:** *a, an.*
4. **Definite pronoun:** *he, she, it, they, we, this, that, those* (Note that the use of relative pronouns is indicated by #13, #14, or #15.)
5. **Indefinite pronoun:** *one, some, something, another.*
6. **Deictic adjective:** *that dog, this cow.*
7. **Indefinite adjective:** *another animal, some fruit, one cow, the other cow, all the objects.*
8. **Deverbal adjective:** the *moving goat, the fallen cylinder.* Such adjectives lead to the ellipsis of a scripted segment. For example, *the moving goat hit the bottle* ellipses the segment “the goat moves.”
9. **Negated nominal:** nobody, none, nothing, none of the, not the ball etc.
10. **Possessive adjective:** *his, its.*
11. **Other Adjective:** *big, fat.*
12. **Prepositional phrase used to modify scripted noun.** Example: *with the stripe.* Sometimes, within the prepositional phrase, possessives or other adjectives appear that refer indirectly to the scripted noun or its parts (e.g., the hippopotamus *with his big mouth*). These are given the codings appropriate to their respective types of adjective (#10 for the possessive and #11 for other adjectives) in addition to the preposition code, because they refer indirectly or in part to the scripted noun. Prepositional phrases used as adverbs are not coded here. For example, in *pushed with his nose* the prepositional phrase is treated as an adverb and thus does not receive either a #10 coding for *his* or a #12 coding.
13. **Nominal specified by restrictive relative clause.**
14. **Nominal specified by nonrestrictive relative clause:** in, *the apple was hit by a lock which went off stage,* the lock is nonrestrictively specified by the relative *which went off stage.* The relative is coded on the segment “lock moves.” Note: Do not code “lock” as ellipsed because it is substituted for by the relative pronoun in this segment.
15. **Nominal specified by script-external relative clause or script-external adverbial:** *the goat that had the red collar hit the table; there is an orange sitting on the plate.* (In the latter example, where the script has “the orange” simply “on stage,” then the specification of it as “sitting” is taken as a script-external adverbial that is auxiliary to the scripted presentative. Where the script actually has the stative scripted, as in “cylinder lying down,” the coding is #23 rather than #15. Unfortunately, there is some inconsistency in the scripting. In the cylinder sequences, the distinction between “on stage” and “lying down” or “standing up” does not seem to be entirely principled. That is, it does not seem to relate to whether the object is to undergo a change of state or to any other discernable principle. Unfortunately, such differences in the script must make other
differences in the coding, as noted below. Since the decision had been made not to alter the script, the only reasonable response was to acknowledge it in a consistent manner, which has been done.) When #15 is coded for unscripted relative clauses, also indicate via #13 or #14 whether the relative is restrictive or nonrestrictive. Do not code at the level of segmental device.

16. Stress: (Note that this device is not yet marked on the transcripts and thus cannot be coded.)

17. Noun phrase coordination: the lemon and the orange. Mark each noun coordinated, for example, code both the lemon and the orange with a #17.

18. Left dislocation: Initialization in clause with a pronoun copy in standard position, as in the table, it hit the lock.

19. Preverbal positioning: Immediately preceding the scripted noun: not separated by another noun. In “the dog chased the stool,” the dog is obviously the preverbal nominal element. In we saw the dog chasing the stool, the dog is still considered to be preverbal because “chase” is the scripted verb. There are a few cases in which a noun is coded as immediately preverbal although it may not be in the strictest sense. For instance, in the spool of thread bumped the lock, “spool” is the scripted noun. Although spool does not immediately precede the verb, spool of thread functionally refers to the same entity. Indeed, thread alone may often be used to make the same reference. In the few cases like this, the presence of the prepositional modifier did not seem to justify disqualifying its noun as preverbal. A final note: again, there is some confusion in the sequences scripted “on stage.” The coding of there is a cylinder standing up therefore depends on whether the script reads “cylinder on stage” or “cylinder standing up.” In the former, the scripted verb is taken to be the presentative, which cylinder follows postverbally. In the latter the scripted verb is taken to be the stative, which cylinder precedes. (Responses vary from a cylinder is there to there is a cylinder to a cylinder is standing there to there is a cylinder standing there. Perhaps one should rethink the position of the “on stage” vs. its stative alternatives in the script, although this was not originally seen as viable.)

20. Postverbal positioning: Immediately following the scripted verb; not separated by another noun. In the dog chased the stool, the stool is the postverbal nominal element. In simple, scripted presentatives (e.g., there was/we had an orange, a lemon, and an apple) in the scripted “lemon, apple, orange on stage”), code the noun that follows the presentative verb as immediately postverbal. (Here that would be the orange.) In the case of a presentative that is not scripted and an adverbial that is, (e.g., there is a dog chasing a cow for the scripted “dog chases cow,” dog is considered to be immediately preverbal vis-a-vis the scripted verb and cow is considered to be immediately postverbal.

21. Placement in a by-clause: Indexes the use of the passive. Mark only the noun that actually occurs in the clause, (e.g., in the dog was chased by the stool mark only the stool.)

22. Presentative nominal: Any presentative, whether scripted or not. Examples: There’s a dog, We had a dog, They showed an apple chasing an orange.

23. Nominal specified by a scripted adverbial: There’s a dog chasing a cow: We see a kangaroo turn.
7.6 Verb Specification Devices

1. Ellipsis: Include all types of gapping as well as complete ellipsis of scripted verb. Examples: NP + NP VP (the dog and ball ran); VP NP + NP (there was an orange and an apple); presentative without any form of the scripted verb.

2. Presentative: Any form that introduces a referent. Examples: there is, they show, we have. Most frequently these will appear when there is a new “on stage” element. When a presentative is not scripted in this way but used to introduce scripted action, code on the scripted action verb, in addition to other applicable codings (e.g., in we see an apple chasing a stool, code a presentative on chasing.

3. Deleted presentative: A sentence in which the main verb is missing. Example: an apple which gets hit by a dog (where there is should occur). Only coded on the first segment of each coordinated sequence, such as apple moves for an apple running and hitting a stool.) Where sequences are not coordinated with and, each is coded for a deleted presentative. Example: Code #3 on both apple hits stool and apple hits cow for a noncoordinated response such as Apple hitting stool. Apple hitting cow.

4. Stative: Examples: an orange is sitting on a plate, standing, resting. Code even when statives are not specifically scripted.

5. Directional adverb: after, around, on, into.

6. Exit directional: A subset of directionals which indicate something has left. Even though a subset, do not double-code a #5. In the interests of objectivity and uniformity, code all instances of the following: off, away, off stage. The “exit” component of these may be vague, because off may mean “off the plate” or “off stage” and away may simply mean “away from another animal or object.” All “exists”, including those simply relative to another object or position should be coded.

7. Manner adverb: Put any nondirectional adverbs in this category. Examples: slowly, in a circle, twisting, sort of, just etc. However, adverbs of time or order such as then were not included, because they already make up pansegmental category #7.

8. Causative replacing action: Examples: the lemon made the apple roll into the orange.

9. Causative coded on resultant: Examples: the orange hit the lemon, making it roll into the apple; the orange hit the kangaroo, which made it hit the pencil. Code causative on the segment with “roll into” or “hit,” respectively, because the causative does not replace the action as in #8.

10. First member(s) of coordinate verb phrase: All verbs in script (in different segments). To be assigned the #10/11 coding an explicit and must occur at the end of the sequence. Examples: “moved” in moved and hit; both “turned” and “moved” in the hippo turned, moved over, and hit the ladder.

11. Last member of coordinate verb phrase: All verbs must be in the script (in different segments); an explicit and must occur at the end of the sequence. Examples: “hit” in moved and hit.
12. Double transitive: Two transitive verbs in one segment where script has only one verb. Example: *hit it and knocked it over.*
13. Double coding of motion: Two intransitive verbs where script has only one. Example: *the ball rolls and then it goes and hits.* This is coded on “hits”.
15. Reciprocal: Example: *touched each other.*
16. Reflexive: Examples: *himself, herself, themselves.* Code only for scripted verbs. Some scripted verbs cannot be coded according to this scheme, (e.g., if the scripted verb occurs as an adverbial, in the infinitive).
17. Past imperfect: Examples: was chasing, was being chased by, there was.
18. Past perfect: Examples: *chased, was chased by, we saw.*
19. Present imperfect: Examples: is chasing, is being chased by, there is, we have.
20. Present perfect: Examples: *chases, is chased by.* In segments like “apple moves,” a description such as *goes rolling* was classified as present perfect (for *goes* as an indicator of movement) plus an adverb indicating the manner of movement (rolling).
21. Retrospective: Example: *has chased.*
22. Inchoative: Example: *is about to.*
23. Inceptive: Examples: *starts to, started to.* (Also code for tense, when possible.) Examples: *started spinning* would be coded as the imperfect past as well as inceptive. There is no tense coding for the infinitive plus inceptive, as in *started to spin.*
24. Generalized verb: Verb that does not specify who does what. Example: *the orange and the apple chased around.*
25. Adverbial: Scripted verb is in adverbial form. Example: “chasing” in *there was an apple chasing a stool.*
7.7 *Segmental*

1. Is there a preceding segment? **PSEG**
2. Is there a following segment? **FSEG**
3. Is the preceding segment dynamic (i.e. movement)? **PDYN**
4. Is the following segment dynamic? **FDYN**
5. Is the preceding segment transitive? **PTRAN**
6. Is the following segment transitive? **FTRAN**
7. Is there a preceding cut? **PCUT**
8. Is there a following cut? **FCUT**
9. Is there an observer present? **OBS**
10. Is there an inferred action (i.e. not actually seen)? **INF**
11. Does the previous scene have parallel structure? **PAR**
7.8 Nominal

1. Is the word a noun or a pronoun? \( \text{N} \)
2. Is it exophorically given (achieved by leaving projector on)? \( \text{EXO} \)
3. Is it anaphorically given (achieved by presence in current scene)? \( \text{ANA} \)
4. Was it nominated as topic (in constrained production task)? \( \text{NOM} \)
5. Is it the only element that is given or only that is new? \( \text{SANA} \)
6. Is the referent confusable? \( \text{CONF} \)
7. Is the element cinematographically salient? \( \text{CSAL} \)
8. Was it previously salient (in current scene)? \( \text{PSAL} \)
9. Is it potent? \( \text{POT} \)
10. Is it animate? \( \text{ANI} \)
11. Is it human? \( \text{HUM} \)
12. Singularity of animacy? \( \text{SANI} \)
13. Is it currently the perspective (based on verb markedness)? \( \text{CPER} \)
14. Was it previously a perspective? \( \text{PPER} \)
15. Is it moving? \( \text{MOV} \)
16. Singularity of motion? \( \text{SMOV} \)
17. Was it the first mover? \( \text{FMOV} \)
18. Was it used as reference location? \( \text{LOC} \)
19. Was there an earlier segment within this scene with same verb? \( \text{PARS} \)
20. Did the item appear in the same role in an earlier segment? \( \text{SAME} \)
21. Different animacy in the parallel segment in the previous scene? \( \text{ADIF} \)
### 7.9 Verbal

1. Action?  
   - ACT
2. Transitive?  
   - TRAN
3. Static?  
   - STAT
4. Collision?  
   - COL
5. Percussive?  
   - PERC
6. Result of previous action?  
   - RES
7. Negation of expectation?  
   - NEG
7.10 *Participant-Generated*

1. Was the first segment ellipsed? FELL
2. Was the last segment ellipsed? LELL
3. Was the previous segment ellipsed? PELL
4. Was the following segment ellipsed? FSELL
5. Was the item previously mentioned? PMEN
6. Was there a preceding presentative? PPRE
7.11 Script of Cartoon

1. Woodpecker is pecking in a tree.
2. Dog runs in from off stage and runs up to the tree.
3. Bird stops pecking and dog barks at bird.
4. Bird looks at dog.
5. Bird pecks at tree again (pecks until has almost pecked through tree branch).
6. Bird stops pecking and tree top falls to the ground.
7. Dog and bird run off stage.
8. Tree top follows (chases) dog and bird.
9. Bird and dog and tree top come back on stage and tree top chases dog which chases bird.
10. Bear walks into cave.
11. Bird and dog go into cave.
12. Bird and dog come out of cave then bear follows (chases) them.
13. Bird and dog run up an evergreen tree together.
14. Tree runs away with bird and dog in it.
15. Bear picks up a stick.
16. Bear walks along and encounters a banana tree and hits the tree with the stick.
17. Banana falls on bear.
18. Bear picks up banana and eats it.
19. Bear walks along and encounters an apple tree, picks up another (duplicate) stick and hits the tree.
20. An apple falls on the bear.
21. Bear drops stick and then picks up the apple and eats it.
22. Bear picks up the same stick and walks to a banana tree and hits the tree with the stick. (Note: the banana tree now has a monkey in it.)
23. Monkey falls on bear.
24. Monkey gets up and growls at bear.
25. Monkey grabs stick from bear.
26. Bear points at banana tree.
27. Monkey hits banana tree with stick.
28. Bananas fall on bear and monkey.
29. Bear and monkey shake hands and then they each pick up a banana and eat it.
8. VionColas

Vion, Monique  
Laboratoire Langage et Parole  
Université de Provence  
Les vergers C, 11 ave. Marius Joveau  
Aix-en-Provence  
F-13090, France  
mvion@up.univ-mrs.fr  

Colas, Annie  
CREPCO-CNRS  
Université de Provence  
29, avenue R. Schuman  
Aix-en-Provence  
F-13621, France  
acolas@newsup.univ-mrs.fr

This corpus was designed to study the cognitive constraints (memory based and/or inferential) that affect the establishment and management of links between events. The general hypothesis was that the linguistic expressions that structure discourse are the manifestation of conceptual constraints imposed by the information management process. By varying the conditions of information availability, inference making, and thematic continuity in pictorial narratives (silent comic strips), we provided verbalizing conditions that were more or less favorable to establishing conceptual relationship.

Each comic strip contained eight frames (8 x 8 cm). The first frame showed two characters. All subsequent frames showed only one of the two characters carrying out various activities. A minimal link between the frames was achieved by the continuous presence of one of the characters from the first frame.

Four different comic strip versions were constructed using a factorial combination of two variables: thematic continuity and layout. The first variable concerned “thematic continuity”. In the maintained topic condition, the materials were designed in such a way that a topic would be induced after the first frame by the repeated presence of the same character in every frame, up to and including the last one. In the changed topic condition, the materials were designed in such a way that a thematic break was generated by the reintroduction in the last picture of the other character from the first frame (in other words, frame 1 had both characters, frames 2 through 7 showed only one of the two characters, and frame 8 showed only the other). The second variable was a secondary one used to control the layout of the characters in the frames. To avoid any bias in referent marking brought about by the greater salience of one of the two characters due to its location in the picture, the layout (left, right) of the characters in the first frame was counterbalanced.

The comic strips differed as to whether event sequence was arbitrary or ordered. In the arbitrary sequences, the events although presented as a sequence, could have occurred in any order (e.g., in A1, the daily activities depicted are relatively independent of each
other, and thus required inference making: the woman getting dressed - or undressed - could have been placed after the woman putting on - or taking off - her makeup, or anywhere else in the sequence, for that matter). In this case, the speaker's had to infer the links between the pictures from the proposed sequence in order to build an overall representation of one story. In the ordered sequences, the order of the events could not be changed (e.g., in O12, before potentially catching a fish, the man had to put on his fishing gear, go to the water's edge, and cast the line). The ordered sequences still did not have a script structure because the normal sequence of events was modified by the sudden appearance of an obstacle. The obstacle was always an event over which the main character had little or no control. In some of the comic strips, the obstacle interrupted the causal chain of events (e.g., in O15, the car hit a hedgehog crossing the road). In others, the obstacle did not interrupt the causal chain but created a surprise effect that sometimes substantially changed the expected course of events (e.g., in O16, the air bubble the fish entered so it could fly burst) and sometimes did not (e.g., in 013 the cereal bowl fell and made a hat for the cat hanging on the tablecloth). For each type of sequence, the materials consisted of 32 test comic strips (8 pairs of characters x 4 versions).

The last variable manipulated was the frame display mode. In the simultaneous display mode, all pictures were on one page. The speaker was asked to look at the comic strip and to prepare to tell the story immediately afterwards. In the consecutive display mode, the comic strip was presented in booklet format, with one picture per page. Subjects were instructed to turn the pages one by one and to say what was happening on each page. As such, the events had to be verbalized on-line, as they were discovered.

Participants
One hundred and ninety-one native French-speaking subjects (98 males and 93 females) participated in the study. There were 63 seven-year-old children (attending first grade), 64 nine-year-old children (attending third grade), and 64 eleven-year-old children (attending fifth grade).

Data collection design
Each speaker was tested in only one frame display mode and on one type of sequence. During testing, a given participant saw eight test comic strips (each presented in one of the four versions).

Procedure
Testing was individual and lasted approximately 20 minutes. In the room where the experiment took place, there were three persons, the speaker, the experimenter, and the addressee of the narration. The addressee was a same-age peer from the speaker's grade in school. He/she only acted as listener once during the experiment.

Recording
1528 narratives were audiotaped in several public elementary schools in Aix-en-Provence (Château-Double, Henri Wallon, Les Granettes) and Luynes (public elementary school and St François d'Assise) France. We would like to thank Delphine Baigue and Aïcha Idriss-Abdalla (graduate students at the time) for their help in preparing the materials and collecting the data.
Pictures

The description of the picture stimuli is as follows. First, the arbitrary sequences:

A1: Un homme et une femme  (A man and a woman)
A2: Un adolescent et un garçonnet  (An adolescent and a little boy)
A3: Un homme et un adolescent  (A man and an adolescent)
A4: Une femme et une fillette  (A woman and a little girl)
A5: Une tortue et un crocodile  (A tortoise and a crocodile)
A6: Un singe et un lion  (A monkey and a lion)
A7: Une poule et des poussins  (A hen and chicks)
A8: Un chat et un âne  (A cat and a donkey)

And then the ordered sequences:

O9: Garçon et grand-père  (Boy and grand-father)
O10: Homme et femme  (Man and woman)
O11: Garçon et fille à la plage  (Boy and girl at the beach)
O12: Fils et père à la pêche  (Son and father fishing)
O13: Chien et chat  (Dog and cat)
O14: Ver et escargot  (Worm and snail)
O15: Hérissón et lapin  (Hedgedog and rabbit)
O16: Poisson et grenouille  (Fish and frog)

File names are constructed using the first three number for the participant ID and the fourth and fifth for the age (07, 09 and 11). Then come three letters. The first two are either im (arbitrary) or ex (ordered) and the last is either g (simultaneous) or s (consecutive).

The gem codes in the files begin with V for vignette. Then there is a number for a picture number of the letter “u” for the picture series 2 through 7 and then the letter “m” for maintaining topic or “c” for changing topic. If the child went straight from the “u” sequence on to sequence 8 without a break, then the CHAT will include +… followed by +^ as in this example:

VU:  aprè il joue du piano +...
(V8m) *CHI:  +^ et il va se coucher à l'ombre.

Researchers using these data should cite one of these sources: