Children's sentence planning: Syntactic correlates of fluency variations*

DANA MCDANIEL

University of Southern Maine

CECILE MCKEE AND MERRILL F. GARRETT

University of Arizona

ABSTRACT

This paper argues for broader consideration of children's language production systems and, in that context, describes research on children's planning of syntactic structures. The research presented here measures non-fluency patterns in elicited utterances of varied syntactic type. We describe and interpret several regularities in these patterns for two groups of children ('young': three–five-year-olds; and 'older': six–eight-year-olds) and an adult comparison group. The evidence indicates a strong correspondence of adult and child responses to structural complexity, both in terms of global fluency measures and in terms of more detailed indicators of planning load. In addition, we report some specific contrasts in the patterning for children and adults that suggest disparities in processing resources and/or in local planning strategies.

Children's utterances are typically used to evaluate their knowledge of words and structures. Their utterances are less often studied for what they reveal about language processing per se or to address the effects of interactions between a performance system and the linguistic competence that

---

it integrates. Instead, most research on language development has children's linguistic knowledge as its core concern: Are children's grammars adult-like in various respects? And, if not, how are they different, and how do they become adult-like? Our objective here is to target the complementary processing questions: To what extent are children's language production systems adult-like, and in what respects do they differ?

The extensive literature on children's linguistic competence dwarfs study of children's comprehension and production mechanisms. The contrast is especially striking in the case of production, where a vast body of spontaneous speech has been collected and analyzed. Though obviously relevant, it is not used to explore the production system itself. Wijnen (1990: 651), echoing Marshall (1979), made this point almost twenty years ago: 'Although most child language research is based on spontaneously produced speech, the predominant approach is competence-based.' The irony here is that production data are arguably poorly suited for studying grammar (for all the reasons that lead to its near-total neglect by syntacticians studying adults). The data represent competence through a production filter, so an appropriate analysis must provide principled ways to subtract effects of that filter; that is, the analysis must presume some theory of real-time sentence generation. Further, to the extent that grammar is influenced by innate components, it could be that much of the developmental trajectory is determined by the ontogeny of the processing systems that are responsible for production and comprehension of sentences, since those control real-time performance. The principles on which these systems operate may be adult-like even from the beginning, but the way those principles are deployed cannot initially be adult-like because at the very least their effective application requires practice. Such processing includes both the specific mechanisms for integrating linguistic form and the more general mechanisms of memory and attention. These two aspects of processing interact to generate the performance profile for both child and adult.

On both linguistic and non-linguistic comparisons, children differ from adults. For example, the working memory that supports various aspects of cognition is thought to increase during development. The lexicon clearly starts out smaller than it ends up, and it probably undergoes some restructuring during development. These and related observations invite the expectation that the language processing systems may differ in children and adults. We focus here on that expectation, analyzing the distribution of different types of non-fluency in elicited utterances of varied syntactic complexity. Available evidence does not provide a basis for detailed claims about similarities and differences between child and adult production systems. But a reasonable working hypothesis is that children's more limited processing resources could lead to smaller and possibly structurally different planning domains. To get more detail on development per se, we

compare three age ranges: younger and older children, and adults. A natural distribution of differences across those groups would show our older children sharing features with both younger children and adults.

The study we report here assumes a framework characteristic of several existing models for adult production (Bock, 1987; Dell, 1986; Garrett, 1980; Levelt, 1989). These have in common a three-level staging that begins with non-linguistic planning for content (message level processing), followed by successive stages of syntactic integration (sentence formulation) and sound structure planning (phonological and prosodic processing). Lexical processes are linked to these stages, with an initial retrieval of semantic and syntactic lexical representations (lemmas) followed by retrieval of lexical representations that provide morphophonological information (lexemes). As we detail below, these adult modeling claims can be applied to an account of developing production capabilities with a focus on syntactic planning domains.

Most developmental research on production has emphasized phonological and lexical errors in children's spontaneous speech (e.g. _bolar pears_ for _polar bears_; _easy_ for _hard_; Stemberger, 1989; see also Jaeger, 1992; Jaeger, 2004, provides an excellent overview and extension of such research). This approach has strong parallels in research on adults. But, while spontaneous speech errors are quite revealing, they cannot carry the empirical burden alone. Bock (1991) points out that the linguistic and extralinguistic contexts of natural errors may vary freely in ways that bear on both the occurrence of error and the validity of different explanations, making some form of experimental control essential. Experimental paradigms are as necessary for studying sentence production as they are for studying sentence comprehension. In both domains, many crucial questions can only be settled in an experimentally controlled environment.

In the study of adult production, a combination of naturalistic and experimental data has proven powerful. Observations from early speech error and fluency analyses motivated general claims about the architecture, or fundamental design, of the production system. Subsequent experimentation extended and modified those claims. Early research often focused on the controlled elicitation of the phenomena measured in spontaneous speech both as a way of testing the generalizations derived from the observational data and as a way of developing experimental methods. Contemporary adult production research now employs an array of techniques for inducing error and for measuring the time course of targeted features of language generation (see, e.g., Bock, 1996, for a systematic review). We have pursued a similar course for the study of emerging production processes in children. That project includes both the reinterpretation of some past observational data on child production and the exploration of experimental modes of evaluating child language production.

61

In prior work, we have used an elicitation technique with children aged two to twelve (described in the next section). That work provides information on the accuracy of performance with different structures and information on the time course of producing those structures. We explore here fluency patterns in utterances elicited with the same experimental techniques. Our aim is to assess locations where planning load varies. Several investigations of language production have examined the detailed structure of non-fluencies (e.g. pauses, both unfilled and filled) based on the assumption that the distribution of such indicators of production difficulty signals at least in part the ebb and flow of processing demands. Most previous work has focused on adult language, with a smaller but significant body of developmental study. The non-fluencies in these studies are broken down in ways that differ slightly from each other and from our approach. For later clarity, we exemplify our terms in (1).

(1)   Non-fluency types with examples (italicized):
     Unfilled pause: A moment of silence.
        The one that [*960 ms*] Big Bird thought the princess was
        kissing. (participant 7;1)
     Filled pause: A 'filler' word, such as *um* or *uh*.
        The one that Dorothy said *um* was tapping the horse.
        (participant 5;3)
     Restart: A repetition or repair, where the speaker returns to an
              earlier part of the utterance and continues again from that
              point.
        Pick up the one – pick up the bear that the evil king is pushing.
                        (participant 5;1)
     Part-word repetition: Part of a word precedes the successful
                           utterance of the word.
        The baby that's *pu- pulling* the rooster. (participant 5;3)
     Prolongation: A word is unnaturally lengthened.
        The one that *he's* ([*hi::z*]) touching. (participant 4;9)

Our working assumption throughout is that the nature of the material following a non-fluency is frequently and systematically a cause of the non-fluency. We recognize that pauses and other non-fluencies may arise from sources other than the structure of impending speech (e.g. changes of intention or distraction and attentional shifts). We rely on the systematic character of the language-generated effects to help us filter out the relevant class of observations. There is substantial intuitive and experimental support for this assumption. Early work by Maclay & Osgood (1959) used filled and unfilled pause distributions to argue for planning units larger than the word, though the specific triggering structures were not detailed. Pioneering work by Goldman-Eisler and colleagues established systematic

relations between non-fluency phenomena and complexity of speech planning tasks (see Goldman-Eisler, 1972, for a review). Of particular interest for our work are the numerous studies reporting pauses clustering at the onset of clauses (Beattie, 1980; Boomer, 1965; Butterworth, 1980; Ford, 1978). We are interested in the indication in that literature that processing load increases near the beginnings of planning units (see Bock, 1996, and Garrett, 1982, for relevant comment). The adult studies of non-fluency are complemented by several related observations in children. These have convincingly tied fluency phenomena to the relative facility with which children control and integrate language structure (e.g. Rispoli, 2003; Wijnen, 1990, 1992). These and similar studies link linguistic development to fluency changes and thus buttress the assumption that non-fluency patterns provide a plausible tool for investigating specific features of the developing production system.

Our previous research (McDaniel, McKee & Bernstein, 1998; McKee & McDaniel, 2001), as well as Hawkins (1971) and research investigating part- and whole-word repetition in children who are not stutterers (Bernstein, 1981; Bloodstein, 1974; Wall, Starkweather & Cairns, 1981), suggest that clause boundaries may be critical planning junctures for children as they are for adults. If this is true, children's indicators of production difficulty should reflect this. For example, non-fluencies should occur at the onset of relative clauses more than inside them (e.g. *The rabbit that um … the girl tickled* rather than *The rabbit that the girl um … tickled*). Our experiment pursues this and related predictions in detail.

Two major sources of difficulty are typically discussed in linguistic accounts of fluency variation: word finding and syntactic planning. Our research is designed to focus on the latter. A key question is what specifically stresses the sentence planning system and is likely to surface at planning boundaries. To explore such questions, we examine relative clauses. Our past research investigated children's knowledge of these in both production and judgment paradigms. The relative clause construction is revealing because of its syntactic properties: It is a multi-clause structure, with a clause embedded in a noun phrase. The embedded (relative) clause includes a 'filler–gap' construction, exemplified in (2).

(2) a. I found the boy *who* the cat scratched [*gap*].
    b. I found the boy that the cat scratched [*gap*].

The noun *boy* in these examples, referred to as the 'head', is modified by a relative clause. The sentences convey the message that the cat scratched the boy, even though *the boy* does not follow *scratched*. In this structure, *boy* (or *who*, which relates back to *boy*) is 'split' between the position where it occurs (after *found*) and the position after *scratched*. The element hypoth-esized to occur after *scratched* is a gap (also referred to as a 'trace'), and it

relates to the filler *boy/who*. In English relative clauses, it is common for *that* to occur rather than a *wh*-word, as in (2b). This element is analyzed as a complementizer (the same element that introduces complement clauses, as in *You said that the cat scratched the boy*). Since the relative clause structure with *that* has the same basic properties as the one with a *wh*-word, it has been analyzed similarly, as a filler–gap structure with the gap in the same position (and, in some analyses, with a null *wh*-phrase occurring in the same position as *who*). For purposes of exposition, we refer to both *who* and *that* as 'relativizers' when they occur in the initial position of a relative clause. We use the term 'complementizer' only to refer to *that* when it introduces a complement clause.

There is very little research on the production of filler-gap structures.[1] In a production study that included relative clauses, Ford & Holmes (1978) measured adult response times for a tone detection task performed during spontaneous speech. They found that reaction times were shorter before, and longer just after, the onset of a relative. They attributed this to an increased processing load, which in current terms could be related to the filler–gap structure. Our previous research suggests a strong preference in both children and adults for having the gap close to the filler. For example, they used the generally dispreferred passive to turn structures that would otherwise be object relatives into subject relatives (e.g. *The sheep that the doctor is rubbing* became *The sheep that is being rubbed by the doctor*; McDaniel *et al.*, 1998).

It might seem that producing filler–gap constructions would be less challenging than comprehending them. The listener needs to figure out the filler–gap relation, but the speaker – being the creator of the structure – already knows this relation. However, filler–gap structures that involve more than one clause do pose an interesting problem for the production system. If sentence planning proceeds by clauses, the clause with the filler could be planned separately from the clause with the gap. This raises questions about organizational levels and memory. Consider, for example, the relative clause in (3).

(3)   I found the boy *who* you said the cat scratched [*gap*].

If sentence planning proceeds in units delineated by clause boundaries, *I found the boy* would be planned first, then *who you said*, and finally *the cat scratched*. The filler *who* would be planned at an earlier point in the process than the gap it relates to. If this is true, then the speaker must to some

---

[1] In contrast to their near-neglect in production research, filler–gap structures have been explored in numerous comprehension studies on adults, where the broader question is how the listener relates the filler to its gap. Roberts, Marinis, Felser & Clahsen (2007) summarizes this research and also reports an investigation of children's parsing of filler–gap structures.

64

extent have the gap clause in mind when saying the filler clause. This also means that the relativizer (*who* in (3)) should be an especially revealing part of the structure, since at this point both the preceding relative clause head (*boy* in (3)) and the upcoming gap must be represented.

The experiment presented here elicited four types of relative clause structures. We have elicited similar structures in other studies (e.g. McDaniel & Lech, 2003; McDaniel *et al.*, 1998; McKee & McDaniel, 2001; McKee, McDaniel & Snedeker, 1998). These earlier studies of children's utterances, often in conjunction with grammaticality judgments, convinced us that young children's grammars are adult-like with respect to relative clauses. We coded utterances from the present experiment for structural properties and fluency phenomena. We emphasize here the distribution of unfilled and filled pauses and restarts. Exploring their distribution in complex structures produced by children of different ages and by adults is one way of addressing continuity questions. If hesitation phenomena distribute differently in child and adult speech, it would suggest that the production system undergoes significant changes during development. On the other hand, if the distribution is the same for adults and children, with or without a difference in magnitude, it suggests that similar mechanisms underlie the child and adult sentence planning process.

METHOD

We compared fluency indicators across structural variations (subject vs. object relatives; depth of clause embedding) and across three age groups. The comparison across structures evaluates planning units and what might stress sentence planning, such as distance between a filler and gap. The comparison across age groups evaluates developmental changes in the system that handles such stresses.

Participants

Two groups of children and one group of adults participated in our experiment. We aimed for approximately 25 to 30 participants in each group, but the actual numbers were affected by some variation in exclusion rates. Our Young group included 23 three–five-year-olds (range: 3;5–5;9, mean: 5;0; 14 girls, 9 boys). Our Older group included 24 six–eight-year-olds (range: 6;1–8;10, mean: 7;6; 13 girls, 11 boys). Our Adult group included 30 participants (26 women, 4 men). All participants were monolingual, native speakers of English with no known speech, language or hearing disorders. The children attended schools and daycare centers in Portland, Maine, and the adults were students in introductory-level linguistics courses at the University of Southern Maine. Data from 16 additional child participants

65

(10 in the Young group and 6 in the Older group) were not analyzed because they did not complete the task.

Task

Each participant was seen in one session that lasted approximately 15 to 30 minutes. Our elicited production task involved two experimenters working with each participant. One experimenter told stories using toys as props. The stories were always about two identical toys, which were distinguished by an event. At the end of each story, the storyteller asked the other experimenter to cover her eyes. Then the storyteller pointed to one of the toys, and the participant told the other experimenter to pick that toy up. Since the experimenter so directed could not see and since the relevant toys were identical, the most effective way for a participant to describe the designated toy was with a relative clause. Once the participant described the toy, the experimenter uncovered her eyes and picked it up. A sample protocol is given in (4).

(4) **Storyteller**: This is a story about two sheep, and they look exactly the same. There's also a doctor in this story [places doctor behind sheep]. In this story, the doctor's going to rub one of the sheep. [doctor says:] 'I feel like rubbing a sheep today … hmm … not this one … I think I'll rub <u>this</u> sheep!' Rub, rub, rub. [to second experimenter] Now cover your eyes. [to participant] I'm going to point to one of the sheep, and you tell [second experimenter] to pick it up. [points to sheep that the doctor is rubbing]
Participant [targeted response]: Pick up the sheep that the doctor is rubbing.

Elicited production has an advantage over spontaneous production in that the message level (in models like Garrett, 1980, and Levelt, 1989) is somewhat controlled for. That is, the task specifies exactly what needs to be communicated. Once people understand the task (which is easy, even for three-year-olds), they understand that they need to tell an experimenter which toy to pick up, and that the best way to do so refers to an event. Therefore, using the above example, the message for most participants would be the same: *Pick up the sheep that the doctor is rubbing*. Also, the storyteller's narration emphasizes the critical lexical items, which were pilot-tested earlier. In the above example, *doctor, sheep* and *rub* are all mentioned before the participant plans an utterance with these words. This is important because we want to minimize pauses reflecting lexical retrieval and focus on those reflecting sentence planning. In this way, the message and lexical items are controlled for to some extent. But the syntactic structure is up to the speaker.

66

Materials

We targeted four types of relative clauses: subject extraction from one clause (Short-Subject), object extraction from one clause (Short-Object), subject extraction from a lower clause (Long-Subject) and object extraction from a lower clause (Long-Object). There were three tokens of each type. Examples are given in (5). Since the stories preceding each item served as relatively lengthy distractions, we did not include filler items. We kept the order of the items the same for all participants to facilitate smooth delivery of the stories. (See the Appendix for the full set of 12 targeted utterances.) An example of a protocol for a Short item was given in (4); an example of a Long item is given in (6).

(5) a. Short-Subject   Pick up the robber that __ is touching the dog.
    b. Short-Object   Pick up the sheep that the doctor is rubbing __.
    c. Long-Subject   Pick up the queen that Grover dreamed __ was washing the pig.
    d. Long-Object   Pick up the duck that Big Bird thinks the princess was kissing __.

(6) Long-Object
     Storyteller: This is a story about Big Bird, and Big Bird is going to do some thinking in this story. Then there are these two ducks, and they look exactly the same. There's also a princess in this story. So Big Bird's standing back here, and the princess is whispering to the ducks. [Princess zigzags between the ducks.] Psh, psh, psh, psh, psh, psh. Big Bird can't see very well from back here, and he thinks that the princess was KISSING one of the ducks! [Big Bird to princess:] 'I think you were just whispering to this duck, but I think you were KISSING this duck!' Now, WE could see that she was really just whispering to both of the ducks, but Big Bird THINKS the princess was kissing THIS duck! [to second experimenter] Now cover your eyes. [to participant] I'm going to point to one of the ducks, and you tell [second experimenter] to pick it up. [points to duck that Big Bird thinks the princess was kissing]
     Participant [target]: Pick up the duck that Big Bird thinks the princess was kissing.

In Short items, the action always continued after the story, until the point where the second experimenter picked up the toy. Long items were more complicated. In these, the upper verb was cognitive (e.g. *think* in the above example) and therefore could not be acted out as clearly. Further, the event in the character's mind did not correspond with reality. In (6), for example, Big Bird was mistaken in thinking that the princess was kissing a duck. We designed the stories this way to discourage participants from responding with utterances like *the duck that the princess was kissing*. Due to these

67

complications, the storyteller always summarized the story with a sentence at the end of the event. The summary sentence included the targeted lexical items. The upper verbs were *think, dream* or *guess*; they were emphasized in the summary sentence. In the *think* items, the thinking was still occurring after the end of the story, so the verb in the summary sentence was in the simple present tense. In the *guess* and *dream* items, the verbs were in the past tense in the summary sentences. The lower verb in the summary sentence was always in the past progressive. We kept the summary sentences of each item lexically and structurally consistent across presentations to minimize variation in the prompt and across participants' responses. Importantly, although we modeled the lexical items and morphology in the summary sentence, we did not model the target response. In fact, we completely avoided relative clauses during all experimental sessions.

Coding

We will cover the following points here: sectioning the utterances, degree of nearness to the target, structural and lexical features, and production-specific phenomena like pauses. Our emphasis is on pauses and restarts, so we will go into considerable detail on the coding of those.

Sectioning the utterances. In order to code the utterances, we broke the targeted structures into six sections. (i) The Pick-up section was the part with the directive verb, regardless of whether the verb was actually *pick up* or something similar like *choose, point to* or *pick*. (ii) The Head section was the head of the relative clause, which was most often either *the* plus the noun, as in the above examples, or *the one*. (iii) The Relativizer section was the word used to introduce the relative clause, which was usually *that* but sometimes *who, what* or NULL (no relativizer). (iv) The Upper Clause section only occurred in Long items. It was the top clause within the relative clause. This clause usually consisted of the character's name or a pronoun and the verb of cognition (not always the targeted verb; e.g. *thought* instead of *guessed*). (v) The Complementizer section, also only occurring in Long items, was the complementizer introducing the lower clause. It was usually NULL (as in the above example) or *that*. (vi) The Lower Clause section was the complement clause for Long items and the single relative clause for Short items.

Nearness to the target. Only target utterances were coded. An utterance was considered target if its general structure (e.g. Short-Object) was what the item was designed to elicit and if it communicated the information in the story. A target utterance was not necessarily grammatical. Participants did vary in how they instantiated the targeted message and sentence structure, as examples (7) and (8) show. Table 1 uses these utterances to illustrate the coding sections described above.

68

TABLE 1. *Example utterances assigned to coding sections of targeted structures*

Section	Example (7)	Example (8)
Pick-up	Beth, pick up	
Head	the one	the one
Relativizer	that	that
Upper Clause	Big Bird thinks	Big Bird thought
Complementizer	that	NULL
Lower Clause	the princess kissed	the princess was kissing the duck

(7) Beth, pick up the one that Big Bird thinks that the princess kissed. (participant 8;9)

(8) The one that Big Bird thought *PAUSE* um *PAUSE* the princess was kissing the duck. (participant 6;0)

Both (7) and (8) are Long-Object structures, since the lower clause object is relativized. But they vary in the Pick-up section (included or omitted), the complementizer (overt or NULL), the choice of verb tense and aspect, and the type of 'gap' (NULL vs. resumptive phrase). Other kinds of variation across utterances included wording of the Pick-up section (choice of verb, inclusion of experimenter's name, inclusion of *please*), choice of relativizer (*that*, *who*, NULL), nouns vs. pronouns, and choice of lexical items (e.g. *pet* vs. *rub*).

Linguistic and fluency features. Each section was coded for a variety of structural and lexical characteristics. For example: Was the section grammatical? Did the nouns and verbs match the target? Which function elements were used? Were there extra elements? We also coded for the following fluency phenomena in each section: number and length of unfilled pauses, number of filled pauses, number of part-word repetitions and number of prolongations.

Coding such non-fluencies required more than the structurally defined sections in Table 1. In particular, we measured phenomena in the spaces between sections. We focus first on time to onset of the target utterance – what we refer to here as the Initial Pause. In most instances, this pause occurred in the interval between the experimenter's last word and the participant's first word. Usually, the storyteller's last utterance was the directive to the participant (e.g. *I'm going to point to one of the ducks, and you tell* [second experimenter] *to pick it up*) and the participant's first utterance was the target. Exceptions to this occurred when other discussion preceded the participant's coded utterance. When the participant said something non-communicative to the blindfolded experimenter, she was asked for clarification. For example, if she said *Pick up the duck*, the blindfolded experimenter replied *Which duck? There are two of them*. In this case, an

69

Initial Pause would be measured between the end of the word *them* and the first word of the participant's next utterance (if it was the target). Occasionally, a participant started with something other than the coded utterance. For example, she might say *This is an easy one* and then the target utterance. In that case, the Initial Pause would be between the end of the word *one* and the beginning of the target utterance. Space 0 was used for cases where the utterance began with a filled pause (e.g. *um, pick up* …). In such cases, the Initial Pause was between the end of the preceding sentence and *um*. Space 0 was the locus of the pause filled with *um* and also any other filled or unfilled pauses occurring between *um* and the Pick-up section.

The remaining spaces are defined by the sections illustrated in Table 1. We called the space between the Pick-up section and the Head section Space 1, for example. A complication arose when an utterance contained a null relativizer or complementizer, and we could not determine whether non-fluencies should be characterized as occurring before or after that section. Consider (8) again, repeated in (9) with our sections indicated below the utterance and our spaces above it. (Spaces irrelevant to the coding of this utterance are in parentheses.)

(9)   (0) (1)   2   3   4·5

     the one that Big Bird thought *PAUSE* um *PAUSE*

       Head   Rel   Upper Clause   Comp
     the princess was kissing the duck

             Lower Clause

The Space between the Upper Clause section and the Complementizer section is Space 4; the space between the Complementizer section and the Lower Clause section is Space 5. Without *that* between *thought* and *the princess*, we cannot designate the pauses and *um* as occurring before the complementizer (Space 4) or after it (Space 5). We thus labeled this part of the null complementizer utterances as Space 4·5, and, similarly used Space 2·5 for utterances with a null relativizer. Example (10), similarly notated, further illustrates our scheme. The full set of our Sections and Spaces is given in Table 2, with (9) and (10) represented in the rightmost columns to show its application to actual utterances. Note that Spaces 2 and 3 are complementary to Space 2·5; Spaces 2 and 3 are used when a relativizer is overt, and Space 2·5 is used only for a null relativizer. The same applies for Spaces 4, 5, and 4·5 with respect to a null vs. overt complementizer.

(10)   (0)   1   2   3   4   5

     pick up the duck that um Big Bird thought *PAUSE* that um *PAUSE*

      Pick-up   Head   Rel   Upper Clause   Comp
     the princess was kissing (participant 5; 11)

            Lower Clause

TABLE 2. *Assignment of example utterances to sections and spaces*

Sections & spaces	Example (9)	Example (10)
Space 0	NA	NA
Pick-up	NULL	pick up
Space 1	NA	
Head	the one	the duck
Space 2		
Relativizer	that	that
Space 2·5		
Space 3		um
Upper Clause	Big Bird thought	Big Bird thought
Space 4		PAUSE
Complementizer	NULL	that
Space 4·5	PAUSE um PAUSE	
Space 5		um PAUSE
Lower Clause	the princess was kissing the duck	the princess was kissing

An important question regarding unfilled pauses was what to count as a pause. Except in a few cases where a coded utterance interrupted a preceding utterance, there was always an Initial Pause of some length and so we always measured this entire pause. But the other pauses required a cut-off length. For adults, we used a 200 ms cut-off, as in most other research on adult hesitations (Bock, 1996). We chose 500 ms as the cut-off for children. Two factors influenced our decision. First, evidence that speech rate is slower for children (see, e.g., Sturm & Seery, 2007, and references therein) suggested that their pauses might also be longer. Second, we wanted a stable measure to apply uniformly across children. The noisy conditions in which many of the children were recorded (adults were recorded in a university laboratory, whereas most children were recorded in daycare centers and schools) made it difficult to pick out all 200 ms pauses. The conservative criterion helped ensure that we were identifying a planning pause. This may, of course, lead to under-representing the number of pauses in children's utterances, but our primary focus is the DISTRIBUTION of pauses, and hence this seemed the best trade-off for our current research.

The restarts presented another coding challenge. Restarts are any material that appears to be 'overwritten' by a later part of the utterance, including both repetitions and repairs. Our coding system, which we designed for comparison across utterances, would not work if we did not give restarts special treatment. For example, an utterance beginning *Pick the – pick up the one* … would include two Pick-up sections with part of a Head section between them. Such a situation would make it impossible to compare the properties of Pick-up sections across different utterances. We therefore cleaned up each utterance prior to coding it by 'covering' restarted material.

71

Starting from the part of the utterance after the part to be overwritten, we hid material leftward word by word until we reached a part of the sentence that made the continuation grammatical. In the example in (11a), the under-lined occurrence of *who* would in some sense 'cover' [*who Bird who um PAUSE*] to its left. Specifically, starting from the third *who* and moving leftward, we 'hid' the above bracketed string from our coding. Similarly, the sequence *the um princess PAUSE um kissed* forms a grammatical phrase that covers [*the PAUSE queen ts*] to its left. The resulting sentence (i.e. without the covered parts) is the codable (11b). In this sense, it is a cleaner version of (11a), which includes the actual messiness of planning the utterance and executing the plan.

(11) a. The one [who Bird who um *PAUSE*] who um *PAUSE* Ernie thinks *PAUSE* [the *PAUSE* queen ts] the um princess *PAUSE* um kissed. (participant 3;8)

    b. The one who um *PAUSE* Ernie thinks the um princess *PAUSE* um kissed.

Although (11b) was the coded utterance, we did not want to lose track of the important information regarding sentence planning that such coding hides. So we kept track of the covered material in each utterance, specifying whether there were any indicators of planning difficulty in the covered parts (i.e. unfilled pauses, filled pauses, part-word repetitions, prolongation or restarts). This information contributed to one of our global measures of production difficulty.

In order to investigate the restarts themselves, we did additional coding that involved comparing the original utterances to their cleaned-up counter-parts. We indicated the section of the utterance where the restart began and the section that was returned to (where the utterance started up again). Consider (12). In this case, the restart begins in the Lower Clause section, after the verb *pet*. The section that is returned to is the Relativizer section, *that*.

(12) Pick up the robber that pet PAUSE that PAUSE touched the dog. (participant 8;6)

In coding the restarts, we conflated the Pick-up and Head sections of the utterances. We did this in order to have a larger comparison set, and also to avoid difficult decisions about which section the restart was meant to cover.

Finally, our coding system also included two composite measures of processing load, one that incorporates non-fluencies across the full utterance (the DIFFICULTY score, abbreviated DIFF) and one that focused on non-fluency at sentence onset (the DELAY score). We calculated DIFF from the sum of the number of the non-covered filled pauses, part-word repetitions and prolongations across the utterance, as well as one point for

every 200 or 500 ms of unfilled pause (so a 1000 ms pause would get two points for a child participant), including the Initial Pause, one point if the sentence included covered material, and one point if the covered material included any non-fluencies. A high DIFF score therefore indicates a relative lack of fluency. The raw DIFF scores ranged from 0 to 13; these were transformed to ratios to compensate for differences in the number of coding sections in the Short (Short-Subject/Short-Object: six coding sections) and Long (Long-Subject/Long-Object: eight coding sections) sentences. (Initial Pause was recorded as 0 in cases where the utterance interrupted the preceding speech. This situation would make the DIFF score 0 if the utterance also contained no filled pauses, unfilled pauses, part-word repetitions, prolongations or restarts.) The DELAY score summed values for Initial Pause and any filled or unfilled pause following the Initial Pause but preceding the first word of the utterance. Filled pauses were counted in the DELAY measure with a time value (300 ms for each filled pause).

We recorded the utterances digitally and analyzed the unfilled pauses using Amadeus II (an acoustic analysis program by HairerSoft, 2008). The procedure we used to code each participant's data involved several steps, including reliability checks. First, one researcher transcribed (orthographically) the utterances to be coded and listened for pauses. This transcriber then measured those pauses and, if they were at least 500 ms long (or 200 for adults), she included them and their length in the transcription. A second experimenter listened to each sentence again to check for transcription errors, including where she perceived the pauses to be. This person also entered the sentences into an Excel file and coded their sections and spaces for structural and lexical features and fluency phenomena. A third researcher was consulted regarding discrepancies between the transcriber and checker/coder. This person also checked the coding and resolved coding questions with the checker/coder (and sometimes with the whole research team).

SPECIFIC RESEARCH QUESTIONS

A wide variety of issues could be investigated using the production data we collected and coded in this study. We emphasize here an exploration of fluency-related phenomena involving pauses and restarts. Specifically, we ask the following five questions that are simultaneously aimed at validation of the various fluency measures and at specific issues related to syntactic planning.

(i) Do the production phenomena correspond to the overall complexity of the structures? The DIFF and DELAY scores address this question. Specifically, we predicted that Short structures would be easier to produce than Long structures, due to the additional embedding in the latter. It also seems reasonable to

predict that Object structures should be more difficult than Subject structures, due to the greater distance between the filler and gap. DIFF and DELAY scores should therefore relate to sentence type and produce a detailed complexity ordering: $LO > LS > SO > SS$. The corresponding developmental question is whether the complexity patterns prove to be similar across age groups. In this regard, we may note that the DIFF score includes fluency features that distribute across the entire sentence and thus reflects both global and local processing challenges. The DELAY score, by contrast, may more strongly reflect early stage planning processes, and could reveal differences between age groups in the depth of planning.

(ii) How do the UNFILLED PAUSES distribute? Studies of adults have revealed tendencies to pause at certain points associated with increased planning loads. As noted earlier, the most common position for a pause is near a clause boundary. If our adult data replicate these and related findings, the question is whether children's pauses distribute similarly.

(iii) How do the FILLED PAUSES distribute? Are the loci of filled and unfilled pauses the same? In adult production, filled pauses have been argued to play multiple roles, reflecting both processing problems and conversational signaling (Maclay & Osgood, 1959; Levelt, 1983; Smith & Clark, 1993). Our experimental situation focuses primarily on the former. Again, the developmental question is whether the children's pattern is the same as the adults' pattern.

(iv) How do the RESTARTS distribute? In general, repetitions and repairs reflect syntactic and phonological structures in complex ways (e.g. Levelt, 1989; Postma & Kolk, 1993). These non-fluent stretches are also often accompanied by filled and unfilled pauses. The developmental question is again similarity in the general character of the child and adult patterns.

(v) What factors influence overt relativizers or complementizers, and how do non-fluencies distribute around these elements? Several possibilities can be developed. Our past research indicates that both children and adults prefer an overt relativizer (almost always *that*) where there is a choice (e.g. object relatives like *the book (that) Grover read*). Ferreira & Dell (2000), using a recall/imitation task with adults, found a similar preference for the overt element in non-movement structures (e.g. *The boy thinks (that) the dog chased the cat*). Specifically, complementizer use depended on availability of material in the embedded clause: complementizers were uttered more when the embedded material was not previously mentioned. They argued that the complementizer may aid speakers in maintaining fluency while planning the next part of a sentence. The general idea is

74

that reduced availability of upcoming structure may trigger expression of optional elements. If this is correct, then the use of *that* and associated fluency measures may indicate the burden that the filler–gap relation imposes and how planning of the gap proceeds in children and adults. A related consideration is the role of the *that*-trace constraint in sentence planning. The complementizer is possible in object extraction structures, where the trace (gap) is after the verb, but not in subject extraction structures, where the trace directly follows the complementizer *that* (e.g. *the duck that Big Bird thinks* (*that*) *the princess was kissing* [*trace*] vs. *the queen that Grover dreamed* (**that*) [*trace*] *was washing the pig*). Given Ferreira & Dell's (2000) analysis, we might expect to find inappropriate use of the complementizer, which would reflect processing demands overriding a grammatical principle. The adult/child comparison is of interest here too: Does this override occur in adults as well as in children? Are there differences in the two child groups?[2]

These five areas of analysis have dual functions. They establish adult patterns for these specific structures as reference points for replication and extension of prior processing claims, and they provide the points of effective contrast between adult and child production performance.

RESULTS

Of the total possible 924 responses, 87 were non-target, yielding a primary database of 837 utterances. The rate of target response in the three groups was 88% for Young, 87% for Older and 96% for Adult. Table 3 summarizes the number of responses for each of the four types: Short-Subject (SS), Short-Object (SO), Long-Subject (LS), Long-Object (LO) in each age group. As the table indicates, sentence difficulty affected target success rate: longer sentences had fewer target responses in each group, but all cells had substantial numbers of observations.[3] Because of unequal numbers of participants across groups, a hierarchical general linear factors analysis (PROC-GLM in SAS) was applied to mimic ANOVA measures. This was applied for DIFF and DELAY scores in a 3 (age) by 4 (structural type)

---

[2] A complication comes from evidence that young children's grammars may be non-adult-like with respect to the *that*-trace constraint (McDaniel, Chiu & Maxfield, 1995; Thornton, 1990). We return to this issue in the 'Discussion' section.

[3] Length affected success rate as expected (0·21 of the total number of 87 non-target responses were Short; 0·79 were Long), but Subject vs. Object sentence types were not uniform. Though SO had more non-target responses compared to SS, LS and LO non-target rates were reversed, with LS having somewhat more (0·49 of the total non-target responses) than LO (0·29 of the non-targets). We have no explanation for the reversal at present.

TABLE 3. *Numbers of utterances analyzed by age group and sentence type*

	Total possible	SS	SO	LS	LO	Total analyzed
Young ($n=23$)	$23 \times 12 = 276$	67	63	51	61	242
Older ($n=24$)	$24 \times 12 = 288$	69	67	55	58	249
Adults ($n=30$)	$30 \times 12 = 360$	90	89	81	86	346
Totals	924	226	219	187	205	837

TABLE 4. *Mean DIFF score ratios by age group and sentence type*
(standard error in parentheses)

	SS	SO	LS	LO
Young	0·522 (0·304)	0·579 (0·398)	0·657 (0·341)	0·756 (0·310)
Older	0·406 (0·230)	0·497 (0·296)	0·555 (0·283)	0·636 (0·348)
Adults	0·406 (0·240)	0·328 (0·160)	0·289 (0·174)	0·342 (0·201)

design. All contrasts of interest within the model are reported as F tests. Effect scores are expressed as partial R^2 (the proportion of the variance that can be attributed to a specific effect).

We look first at the DIFF score, comprised of indices across the full sentence as it sums across the various non-fluencies distributed through the utterance. These scores ranged from 0 to 13, with 13 indicating greater non-fluency; for analysis, scores were expressed as ratios of coding slots to compensate for length differences between Short and Long sentences. Table 4 shows the mean DIFF score ratios in each group for the four sentence types. Analysis showed significant differences among groups ($F(2, 74) = 28 \cdot 41$, $p < 0 \cdot 01$, $R^2 = 15 \cdot 39\%$), as well as significant differences among the structural types ($F(3, 222) = 7 \cdot 42$, $p < 0 \cdot 01$, $R^2 = 1 \cdot 92\%$) and a significant interaction for the structural types across groups ($F(6, 222) = 6 \cdot 80$, $p < 0 \cdot 01$, $R^2 = 3 \cdot 53\%$).[4] Inspection of the patterns of the DIFF score ratios from Table 4 illustrates both the complexity correspondences across age groups and the greater difficulty of the longer (more deeply embedded) constructions for children.

As expected, DIFF score ratios were higher in children than in adults ($F(1, 74) = 50 \cdot 08$, $p < 0 \cdot 01$, $R^2 = 13 \cdot 58\%$), and higher in Young than in Older children ($F(1, 74) = 6 \cdot 75$, $p < 0 \cdot 05$, $R^2 = 1 \cdot 83\%$). With the important exception of the Adult group's relatively high score for the Short-Subject type,

---

[4] We tested for item/order effects in the GLM model to determine whether fluency measures differed, e.g. for early occurring items (trials 1–4) vs. late (trials 9–12). No trial effects or interaction of trials with structures or with age groups approached significance.

the scores showed the expected complexity ordering. Subject structures (SS, LS) were generally easier to produce than Object structures (SO, LO) ($F(1, 74) = 10\cdot14$, $p < 0\cdot01$, $R^2 = 1\cdot56\%$) and Short structures (SS, SO) were generally easier to produce than Long structures (LS, LO) ($F(1, 74) = 10\cdot10$, $p < 0\cdot01$, $R^2 = 1\cdot13\%$). The elevation of Adult scores for SS was reflected in an interaction of adult/child differences with both length ($F(1, 74) = 22\cdot87$, $p < 0\cdot01$, $R^2 = 2\cdot56\%$) and structure ($F(1, 74) = 12\cdot57$, $p < 0\cdot01$, $R^2 = 0\cdot70\%$).

We look next at the DELAY score, which was the interval immediately prior to production of the target and hence more specific to the scope of advance planning activity. Recall that we excluded from the DELAY calculation the time taken by occasional interchanges between participant and experimenter following a scenario presentation. Adults had 22 such instances in 346 trials with a target utterance, Older children 35 of 249 trials, and Young children 57 of 242 trials. Incorporating the time taken by these would have skewed the DELAY scores for those trials and introduced substantial variability given their heterogeneous character. For DELAY analysis, we therefore focused on the point at which we were confident that the speaker had undertaken to generate the response that we scored for its fluency characteristics (viz. the target utterance). The brief exchange interludes included asides, clarification requests and some aborted attempts at compliance. It is unlikely that these are intervals in which the speaker could be directly deliberating about details of the target utterance. However, such trials potentially reflect relative difficulty with formulating an effective response to the test scenarios, and in some cases may have contributed to the likelihood of achieving a correct target response. Their distribution has suggestive information that complements the DELAY score. Three observations are of interest. First, the age groups differed in frequency of such trials: Adult 6%, Older 14% and Young 24% of successful trials. So, children more often engaged in such extra interactions. Second, these were not affected by length/complexity variation: interruption rates were similar overall for Long and Short sentence types. Third, there was, however, an apparent trial order effect that could have inflated incidence for SS types in children. The first experimental trial was an SS item. A total of 37 of the 225 SS trials had an experimenter/participant exchange prior to target utterance, and 27 of these occurred on trial 1; the child groups accounted for 22 of these and adults only 5. No other structural type showed any bias toward its first occurrence as the occasion for a conversational exchange. Thus, to some extent, the impetus to comment or the need for clarification may have been influenced, for children in particular, by learning how to respond appropriately to the test situation.

Table 5 shows the mean DELAY across age groups and sentence types. Analysis showed significant differences among age groups ($F(2, 74) = 10\cdot05$, $p < 0\cdot01$, $R^2 = 6\cdot92\%$), as well as significant differences among the structural

77

TABLE 5. *Mean length in milliseconds for the DELAY score by age group and sentence type* (*standard error in parentheses*)

	SS	SO	LS	LO
Young	1783 (945)	1782 (1133)	1951 (1059)	2017 (1213)
Older	1360 (665)	1763 (1252)	1666 (924)	2096 (877)
Adults	1506 (939)	1172 (565)	1150 (560)	1348 (642)

types ($F(3, 222) = 3\cdot22$, $p < 0\cdot05$, $R^2 = 0\cdot87\%$) and a significant interaction for the structural types and age groups ($F(6, 222) = 4\cdot27$, $p < 0\cdot01$, $R^2 = 2\cdot31\%$). The DELAY score patterns are broadly compatible with the complexity ordering. But the impact of structural detail is not as evident as for the DIFF score ratios in the young children, and the elevation of SS score for the adults also diminishes the overall effects of structure.

The children's DELAY scores were substantially longer than those for adults (with the exception of SS as noted) ($F(1, 74) = 18\cdot86$, $p < 0\cdot01$, $R^2 = 6\cdot49\%$); Young and Older children did not differ ($F(1, 74) = 1\cdot25$, $p = 0\cdot27$, $R^2 = 0\cdot43\%$). In general, DELAY scores tended to be higher for Long structures and Object structures, but neither reached significance: Subject vs. Object ($F(1, 74)$, $F = 3\cdot01$, $p = 0\cdot09$, $R^2 = 0\cdot27\%$); Long vs. Short ($F(1, 74) = 2\cdot17$, $p = 0\cdot15$, $R^2 = 0\cdot22\%$). The high score for the Short-Subject items for Adults seen in the DIFF score ratios was expressed here as a relatively long DELAY; this elevation of adult scores for SS was reflected in an interaction of adult/child differences with both length ($F(1, 75) = 6\cdot33$, $p < 0\cdot05$, $R^2 = 0\cdot65\%$) and structure ($F(1, 74) = 7\cdot13$, $p < 0\cdot01$, $R^2 = 0\cdot64\%$). Indeed, the DIFF score effects for adults at SS can mostly be accounted for by the contribution of Initial Pause values to that measure. With the SS exception, DELAY scores generally fit the expected ordering where differences appeared. The clearest pattern reflecting complexity ordering appears for the Older children. The similarities and differences in these two measures, DIFF and DELAY, we will later argue, can be understood in terms of differences in the degree of advance planning available to the different age groups.

We now turn to the distribution of non-fluencies within the target utterances. These include unfilled and filled pauses, restarts and some specific features of the two clause types associated with relativizers and complementizers. We begin with a report on the distribution of unfilled pauses. We considered the locations in (13) for this analysis.

(13) (Space 0) Pick-up Space 1 Head Space 2/3 Upper Cls
     Space 4/5 Lower Cls

We limited this analysis to sentence-internal pauses; Initial Pauses and pauses in Space 0 are not included in this comparison. Further, some of the

78

TABLE 6. *Relative frequency of unfilled pauses in target utterances for each age group*

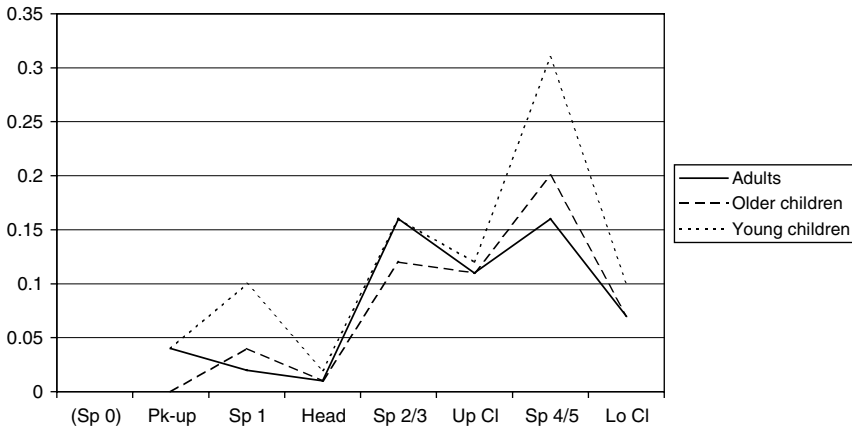|	Pk-up	Sp 1	Head	Sp 2/3	Up Cl	Sp 4/5	Lo Cl
Young	0·04	0·01	0·02	0·16	0·12	0·31	0·10
Rank	5	7	6	2	3	1	4
Older	0·00	0·04	0·01	0·12	0·11	0·20	0·07
Rank	7	5	6	2	3	1	4
Adult	0·04	0·02	0·01	0·16	0·11	0·16	0·07
Rank	5	6	7	1·5	3	1·5	4

Fig. 1. Unfilled pause patterns in child and adult groups.

locations we coded are omitted or combined for the unfilled pause analysis. Spaces 2, 2·5 and 3 are combined (labeled 2/3 above), as are Spaces 4, 4·5 and 5 (labeled 4/5 above) in order to collect all the pauses occurring at each of the two clause onsets. Note also that all the positions reported did not appear in every utterance. The Upper Clause and Space 4/5 positions could occur only in Long items, and the Pick-up and Space 1 positions occurred only in utterances that included a Pick-up section. The numbers of analyzable utterances were all high enough for meaningful analyses, ranging from 102 to 346. We calculated the frequency of the pauses in different positions by considering the actual number of pauses in that position in relation to the total number of occurrences of the section or space in the different groups. Table 6 presents the incidence rates for unfilled pauses across positions. The pattern exhibited by unfilled pauses is graphed in Figure 1. (In some utterances, more than one pause occurred in a single section. For this analysis, we considered only the first pause in such cases.)

79

TABLE 7. *Relative frequency of filled pauses in target utterances for each age group*

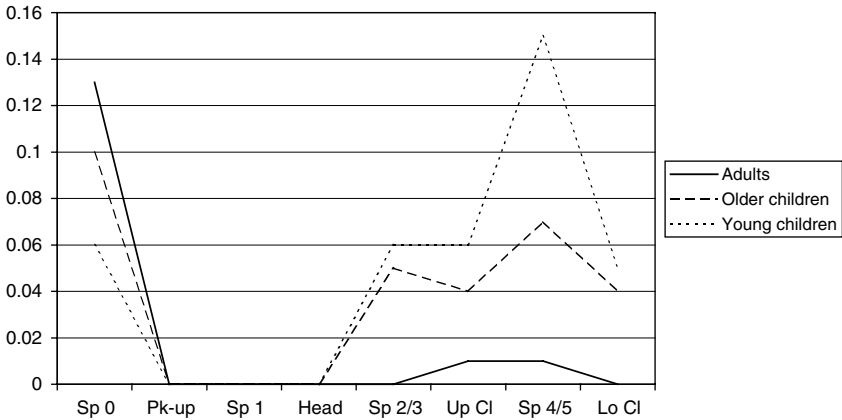|	Sp 0	Pk-up	Sp 1	Head	Sp 2/3	Up Cl	Sp 4/5	Lo Cl
Young	0·06	0·00	0·00	0·00	0·05	0·06	0·15	0·05
Older	0·10	0·00	0·00	0·00	0·06	0·04	0·07	0·04
Adults	0·13	0·00	0·00	0·00	0·00	0·01	0·01	0·00

Fig. 2. Filled pause patterns in adult and child groups.

To illustrate the calculation, the Adult group produced a total of 346 target utterances containing the relativizer spaces (Sp 2/3 in Table 6); 54, or 16%, of those 346 utterances contained one or more pauses. Similarly, the Young group produced 242 target utterances; 38 contained one or more pauses at the relativizer spaces, which also yields a pause rate of 16%. Note that values do not sum to 100 across positions because each position value is computed relative to its own occurrence rate. We used percentages calculated in this way to compare the relative incidence of pauses in the different utterance positions.

The unfilled pause distributions reflect the clausal effects that have been reported in past studies: relativizer and complementizer positions are the most prominent pause loci. For our purposes, however, the salient issue is the relation between adult and child performance. To demonstrate this, the rank value of the frequency at the seven pause locations for each age group is included in Table 6. This indicates, with respect to clausal structure and other features of the distribution, that pausing is highly similar across the three age groups. The match across age groups for ordering of pause

80

TABLE 8: *Number of utterances with filled pauses at three selected regions*

	Sp 0, Pk-up, Sp1, Head	Sp 2, 2·5, 3, Up Cl	Sp 4, 4·5, 5, Lo Cl
Young	16	20	28
Older	26	18	17
Adults	45	2	2

locations is powerful: Spearman rank correlations for child and adult groups across the positions in Table 6 are significant: Adult/Older ($r = 0.88$) and Adult/Young ($r = 0.96$), ($p < 0.05$, critical value 0.714 for $n = 7$).

We will return to further details of pause patterns in connection with the analysis of relativizer and complementizer distributions below. These additional findings also bear on the similarity of processing across the age groups.

Next, we consider the distribution of filled pauses. Filled and unfilled pauses are both prima facie indicators of a processing delay of some kind, though as noted earlier, some claims for different functions of the two pause types have been put forward. We approached the filled pause analysis as we did the unfilled pauses with one difference. Recall that for unfilled pauses, we did not plot those prior to the target utterance (i.e. in the delay region). For filled pauses, however, we did include those that occurred in Space 0 since that location was defined by the occurrence of a filled pause. Table 7 presents the incidence rates for filled pauses across serial positions. Figure 2 graphs the distribution. The low incidence in several categories and the many tied ranks in each group make a ranking comparison such as we applied to the unfilled pause of little use. Even so, it is clear that there is a difference in distribution across the age groups. Adults used filled pauses almost exclusively at the beginning of the utterance: 90% of their filled pauses were in Space 0, compared to 46% for Older children and 18% for Young children. Older children used filled pauses approximately equally in the beginning of the utterance and at the four preferred positions for unfilled pauses; the Young children distributed the non-initial filled pauses into the same positions that were preferred for unfilled pauses, with elevated incidence in the complementizer section. To evaluate these patterns, we condensed the categories in Table 7 in order to collect all filled pauses in loci prior to the relative clause (Space 0, Pick-up, Space 1, Head), those for relativizer and upper clause group (Spaces 2, 2·5, 3; Upper Clause), and those for complementizer and the lower clause group (Spaces 4, 4·5, 5; Lower Clause). Unlike Tables 6 and 7, entries in Table 8 are actual numbers of filled pauses at those combined loci, not corrected for differences in base rates of trials. Comparing Adult and Older groups showed a significant contrast ($p < 0.001$, Fisher's exact test, two-sided), as did Adult and Young groups ($p < 0.001$, Fisher's exact test, two-sided). Older and Young groups

81

TABLE 9. *Number of utterances with filled pauses uniquely at utterance onset or at other locations*

	beginning only	other places only	both
Young	13	39	5
Older	24	29	3
Adults	46	2	3

did not differ ($p = 0.06$, Fisher's exact test, two-sided), though there was the suggestion of greater dispersion in the Young group, and this shows up more strongly in the related comparison that is reported immediately below.

The contrasts so far noted in adult and child patterns with respect to filled pauses suggest a difference in their planning. But another striking feature of these distributions indicates an underlying similarity in the activity that triggers filled pauses. In all age groups, the tendency was to use a filled pause at the beginning of the utterance OR in some other position in the utterance, but not in BOTH the beginning and other positions. In other words, for any given utterance, there tended to be one filled pause location and this was true for all participants. The filled pause surfaced at the beginning for adults, but arose in later locations for the children. The distribution of filled pauses is given in Table 9; again, these are actual numbers of filled pauses, not corrected for different base rates of trials across groups or conditions. To compare age groups, we set aside the 'both' category in Table 9. The interesting contrast lies in the change in distribution across age groups. The filled pause patterns of Adult and Older groups differ ($p < 0.001$, Fisher's exact test, two-sided). A similar comparison shows that Older and Young groups also differ ($p = 0.04$, Fisher's exact test, two-sided). Inspection shows that adults and children differ and that the tendency to distribute filled pauses across sentence-internal locations is stronger in the Young group.

We now report on the restarts. (These include both repetitions and repairs.) Children's utterances contained many more restarts than did adults': 37% of Young children's utterances contained at least one restart, 27% of Older children's did, and only 11% of Adults' did. The typically greater fluency of the adults is on display here, and the contrast becomes sharper when the incidence of multiple restarts within an utterance is identified: 10% of utterances in the Young group had multiple restarts, 5% in the Older, and less than 1% in the Adult. Table 10 shows the proportion of utterances in each structural type that contained restarts; proportions are relative to the number of utterances of each type within an age group. Complexity effects are not evident in the restart measure. Inspection shows little difference between Subject and Object structures across age groups, and analyses comparing restart incidence in Long and Short structures

TABLE 10. *Proportion of restarts by age and sentence type*

	SS	SO	LS	LO
Young	0·25	0·33	0·49	0·54
Older	0·27	0·23	0·31	0·35
Adults	0·11	0·07	0·16	0·15

TABLE 11. *Number of restarts that began in each section*

	Pk-up/Head	Relativizer	Up Cl	Comp	Lo Cl	Totals
Young	26	28	9	7	47	114
Older	28	10	11	1	33	82
Adult	13	3	2	0	21	38

when corrected for length (as in the DIFF analysis) did not differ significantly.

A particularly interesting property of the restarts is how they cluster. Table 11 shows the number of restarts that began in each section. (The numbers slightly under-represent the Upper Clause and Complementizer sections, since the Short structures did not include these sections.) The table shows different patterns for the adults and children. Two things are evident. There is both a similarity and a difference between adult and child patterns. The beginning and final sections of the utterances are common restart loci for all three age groups. Indeed, the relative ordering of restart loci was identical for adult and combined child data. The difference between adults and children was more a matter of degree: Whereas adults preferred the Lower Clause section, and rarely restarted in the middle sections, children's restarts distributed more generally over the utterance; this was more pronounced for the Young group.

A more striking difference in adult and child performance emerged when the Lower Clause restarts were evaluated to determine the domain over which any given restart ranged: What section did the restart return to? To assess this, we focused on Lower Clause restarts to see how many stayed within the Lower Clause as compared to those that ranged outside the clause. The sentences in (14) exemplify these cases.

(14) a. Dana please pick up *PAUSE* the woman that is patting the sheep *PAUSE* the goat. (participant 8;8)
   b. The girl who was pushing *PAUSE* the woman who was pushing the goat. (participant 8;0)

Both of these restarts begin in the Lower Clause section. In (14a), the participant replaces a DP (*the sheep*) with another DP (*the goat*), staying

83

TABLE 12. *Proportion of utterances with a non-fluency at the relativizer locus*

	SS	SO	LS	LO
Young	0·08	0·14	0·33	0·41
Older	0·06	0·13	0·22	0·33
Adult	0·10	0·04	0·21	0·38

within the Lower Clause. In (14b), the participant returns to the Head section.

In Adult utterances, the Lower Clause restarts stayed within the Lower Clause most of the time (81%), whereas Young participants more often returned to an earlier section of the sentence; only 39% remained in the Lower Clause. The Older group's Lower Clause restarts stayed within the Lower Clause around half of the time (55%): for the Adult/Older comparison ($p = 0·082$, Fisher's exact test, two-sided); for the Adult/Young comparison ($p = 0·003$, Fisher's exact test, two-sided); for the Older/Young comparison ($p = 0·179$, Fisher's exact test, two-sided). These observations suggest interesting differences in the scope of planning, to which we will return.

Our last topic concerns findings on the relativizer and complementizer. There are several noteworthy features of these distributions. We will first look at features of the relativizer distributions for indications of processing effects. Recall that the overt relativizer was strongly preferred over the null relativizer by all age groups and for all sentence types. There were only 61 null relativizer structures (about 7% of the total target utterances) and they were about evenly divided across the age groups. (The percent of targets with a null relativizer out of the total number of targets for each age group was: Young: 6%, Older: 4%, Adults: 10%.) The 61 null relativizer structures broke down by sentence type as follows: SS: 19; SO: 25; LS: 7; LO: 10. Though the numbers are small, a complexity effect was observable. The preference for the overt relativizer was significantly greater in Long items than in Short items: binomial test for Long vs. Short ($p < 0·001$). (The Short Subject utterances in this category were reduced relatives, e.g. *the woman pushing the goat*.)

We earlier suggested that overt expression of the relativizer may reflect processing demands. The above pattern of null relativizers fits this expectation: the easier structures used relativizers less often. But, the high rate of relativizer use across the board suggests that its overt expression is only partly driven by processing demands. Another indication of processing demand at the locus of the relativizer is the incidence of non-fluency in that section of the utterance. Table 12 gives the incidence of filled/unfilled pauses and restarts at the relativizer locus (for this analysis, we counted all non-fluencies: before, after or for null cases, at Spaces 2, 2·5 and 3). The

TABLE 13. *Proportion of non-fluencies before and after relativizer*

	before relativizer	after relativizer	both
Young	0·09	0·86	0·05
Older	0·17	0·67	0·16
Adult	0·39	0·51	0·10

structural factors emerged in the expected complexity ordering, and the age groups showed a noteworthy similarity in overall incidence of non-fluencies associated with launching the relative clause structures. Here we applied a hierarchical general linear factors analysis (PROC-GLM in SAS), as previously noted for DIFF and DELAY scores in a (3) age by (4) structural type design. Contrasts of interest within the model are reported as F tests. The structural types differed substantially in incidence of non-fluency ($F(3, 217) = 28·09$, $p < 0·01$, $R^2 = 9·39\%$). The complexity ordering seen in other measures appears here as well, with a clear increase in non-fluency for all age groups – again with the exception of the Adult SS scores. No significant difference among age groups was observed ($F(2, 74) = 1·07$, $p = 0·35$, $R^2 = 0·32\%$); the interaction of group and structural type was not significant ($F(6, 65) = 0·84$, $p = 0·54$, $R^2 = 0·56\%$).

Even with the similar complexity profile, there is a striking difference in the distribution of the non-fluencies when one looks in detail at the spaces in that locus: children are much more prone to non-fluency FOLLOWING the relativizer (Space 3) than are adults. This can readily be seen in Table 13, which breaks down the serial position of the non-fluencies collapsed in Table 12; entries are proportions of all the non-fluencies at the relativizer locus within each age group. To analyze the distribution of pause incidence around the relativizer, we again applied the GLM procedure, with serial position added to the model. Results were as follows: the effects of age group were not significant ($F(2, 74) = 0·48$, $p = 0·62$, $R^2 = 0·09$); however, the serial position effect was significant ($F(1, 74) = 29·43$, $p < 0·01$, $R^2 = 1·97$), as was the interaction of age group and serial position ($F(3, 74) = 5·32$, $p < 0·01$, $R^2 = 0·71$). (The structural effects reported with Table 12 appeared again and in the same way and to comparable degree for this analysis as well; for simplicity, we omit those here.)

Finally, we turn to the other major correlate of clausal structure: the complementizers. The use of the complementizer differed sharply across the age groups. None of the groups manifested the subject/object asymmetry corresponding to a *that*-trace effect, but for seemingly different reasons. The child groups used the complementizer in both the Subject and Object items, whereas the adults avoided it in both. This is shown in Table 14. The figures in the LS column represent *that*-trace violations. There are

85

TABLE 14. *Proportions of utterances with overt complementizers in each age group*

	LS	LO
Young	0·53	0·39
Older	0·25	0·26
Adult	0·04	0·08

too few observations to support a claim about adult performance, though they did trend in the direction of observing the *that*-trace constraint. But, for both Older and Young children, inspection shows that likelihood of complementizer use was not affected by LS and LO environments. The *that*-trace violations did decline with age, but so did the use of *that* in the grammatical cases (LO) for the Older group (the incidence of complementizer was significantly greater for the Young than for the Older group across structural environments ($p = 0.002$, Fisher's exact test, two-sided)).

We summarize all of the findings reported here in Table 15.

DISCUSSION

We have described patterns of fluency in children and adults for seven indicators of sentence production. We looked for clustering of pauses, restarts and so on, in four structures with a prima facie claim to complexity variation. Two major themes emerge from the examination of the non-fluency distributions for the three age groups. The first is the remarkable similarity of the complexity ordering across the age groups. The response to processing demands imposed by the different sentence types is repeatedly demonstrated to correspond for children and adults. The second theme is the evidence for specific differences in the amount of advance planning undertaken by adults and children. The differences in the way children and adults deal with particular processing challenges suggest that children's processing for sentence detail is more locally determined. We comment on some particular features of these two themes.

We began with two broad measures of processing difficulty: the DIFF and DELAY scores. The measures differ in their degree of incorporation of early vs. later processing demands, but both yielded similar general conclusions regarding the complexity of the structures in the experiment. And, children and adults responded similarly to the complexity variation – though with some important exceptions that we discuss below. The same complexity ordering also surfaced in aspects of the restarts and in the relative incidence of the overt markers of clause structure, again with consistent effects across age groups. In addition to these measures, detailed

86

TABLE 15. *Summary of results (RC = relative clause, NFs = non-fluencies, rel = relativizer, comp = complementizer)*

	RC type DIFF scores	Unfilled pauses	Filled pauses	Restarts	Relativizers	Complementizers
Young	SS < SO < LS < LO	mainly at clause boundaries	position varied; usually one place per utterance	all over utterance; tended to range over sections	usually overt; rate of NFs varied with RC type; NFs tended to follow rel	overt about half the time; no *that*-trace effect
Older	SS < SO < LS < LO	mainly at clause boundaries	about half at beginning, half elsewhere; usually one place per utterance	all over utterance; ranged over sections about half the time, local half the time	usually overt; rate of NFs varied with RC type; NFs tended to follow rel	overt about a fourth of the time; no *that*-trace effect
Adults	SO < LS < SS < LO	mainly at clause boundaries	almost exclusively at beginning; usually one place per utterance	tended to be early in utterances or Lower Clause; tended to be local	usually overt; rate of NFs varied with RC type; NFs about equally before and after rel	almost always non-overt; weak indication of *that*-trace effect

87

sentence-internal measures based on unfilled pauses and restarts also yielded strong evidence of correspondence between adult and child processing systems, and this cut across the structural types: dominant positions for increased processing load were similar in children and adults.

These several indicators suggest that from early on, children use adult-like planning strategies. The positions where we found pauses correspond to planning points described in previous studies. This is particularly clear in the case of pauses at clause boundaries (Spaces 3 and 4/5). The case of the Upper Clause pause, which was between the subject and the verb, is less clear with regard to earlier findings. Since the subject was also the first word of a clause (e.g. *Dorothy*), this may correspond to Boomer's (1965) report of a pause locus after the first word of a 'phonemic clause' (a prosodically defined unit that probably corresponds to a simple phrase) but our evidence is inconclusive on this point. Further study is necessary to better evaluate links to Boomer's finding. Another possibility is that this increase in pauses in our data is due to difficulty with cognition verbs, since the upper clause verb was *think*, *guess* or *dream*. Again, this possibility is interesting, but our construction types do not allow us to distinguish serial position and verb type. We did not find any differences across these verbs. Both children and adults often replaced *dream*, and especially *guess*, by *think*, but pauses were no more likely to occur in one situation than in another.

The several indications of similarity across child and adult participants are balanced by non-fluency measures that indicate differences in processing capacity and/or strategy. These indicate that some development of sentence planning does occur in the age range we investigated. Briefly, we think these differences signal that adults undertake planning over longer spans than children, and that children plan more frequently and more locally than adults. That is, children need to stop to plan at each (major) phrase, whereas adults can engage in planning phrase $X+1$ while uttering phrase X. A useful way to think of this contrast is to describe it in terms of the ability to sustain concurrency of processing at different levels of planning. Concurrent processing for planning at message, syntactic and morphophonological levels is a core feature of the production models earlier cited. Planning domains at higher levels span larger ranges and control activity at the smaller scope lower levels. Children may have greater difficulty in coordinating activity across levels and require more frequent access to higher levels to ensure the adequacy of their (lower level) immediate or impending output. This may reflect in part memory resources (e.g. planning chunks held accurately for shorter time spans) and in part less realistic knowledge of the limits of their own capacity (children rush in where adults may pause to reflect).

Several patterns support this view. We comment on four: (i) differences in child and adult filled pause distribution; (ii) differences in child and adult

88

restart distribution; (iii) differences in child and adult pauses at relativizers and complementizers; and (iv) violation of the complexity ordering in the form of elevated processing on some measures for SS structures for adults (and in one instance for Older children).

(i) *Filled pause measures*. The observed patterns suggest that filled and unfilled pauses may in some instances reflect different aspects of the planning process. Recall that filled pauses for adults showed a strong elevation at the onset of the utterance and virtual absence at later points. Children, on the other hand, distributed filled pauses more across the utterance, with clusters at later clause onsets. Unlike adults, children's filled pauses matched their unfilled pause profile. One possibility is that for adults, unfilled pauses arise more often at locations that normally represent predictable disruption to the flow of planning (i.e. places where the processor has typically completed the integration of current structure and detailed planning for the next section is necessary), and that load may vary in terms of specific local conditions. By contrast, filled pauses may more often be points where the speaker is working on multiple levels. The difference between adults and children may then have to do with working memory capacity. Adults do the heavy lifting at the beginning, where they plan the sentence corresponding to the message and begin to project the initial portions of utterance. Children may have to do more planning for both message level and linguistic form at the predesignated stopping points.

(ii) *Restart measures*. The pattern of restarts comports with this account of adult/child planning differences, indicating that children do less preutterance planning than adults do. Adult restarts tended to occur within the Lower Clause, suggesting that they had planned out the major structural units of the sentence – with the occasional exception of the details of the Lower Clause. Children, on the other hand, restart throughout the utterance, and their Lower Clause restarts often return to earlier parts of the utterance. This suggests that they begin the utterance without having worked things out at the message level well enough to formulate a clausal representation that can control detailed, lexically interpreted phrasal structures, and/or that when they lose the thread, they must go back to features of the message in order to recover information that will support detailed local planning.

(iii) *Relativizers and complementizers*. We can also bring to bear on our working hypothesis some properties of the frequency and fluency with which relativizers and complementizers were used. Our findings on the relativizer and complementizer are in some respects

89

compatible with Ferreira & Dell's (2000) account in terms of a stalling device.

All age groups used the relativizer more in Long items (which were of greater complexity) than in Short items. And, the frequency of non-fluencies at this locus matched that complexity profile closely. These observations must be balanced against the brute fact that more than 80% of all utterances had the relativizer, so to think of the overt use of the relativizer as a 'stalling device' may be misleading. It may better be thought of as an 'opportune spot' for planning; it is typically present because there is typically a significant processing load at that point. Variations around that central fact may lead to the occasional elimination of the relativizer (in the simpler processing environments of SS and SO) and to occasional non-fluency (in the more complex environments of LS and LO). The striking asymmetry in pause distribution around the relativizer for the children fits well with this picture. Adults paused with roughly equal (and lower) frequency on either side of the relativizer. Both child groups, however, showed an asymmetry, with the great majority of the non-fluencies occurring after the relativizer: 83% of the non-fluencies followed the relativizer in the Older group and 91% in the Young. Children tended to produce the relativizer and then pause. They 'committed to the relative' but then had to work out the details.

The patterns with respect to the complementizer are also relevant to our claim that children and adults differ in preplanning activity. Children used the overt complementizer significantly more often than adults, and did so even in the LS cases, which yielded *that*-trace violations. Adults for the most part avoided overt complementizers (and therefore produced few *that*-trace violations). The fact that children used the overt complementizer in the LS structures as frequently as in the LO structures is compatible with the possibility that the complementizer is a stalling device for them and that the overuse of this device is responsible for the *that*-trace violations. That possibility fits with the conclusion that children are more 'local planners' with regard to structural detail. The absence of detailed advance planning in this case would mean lack of information distinguishing the LS from the LO constraints on complementizer use.

Adults, on the other hand, clearly did not use the complementizer as a stalling device. Given the option to use the complementizer and its acceptability in the (most complex) LO environment, why did adults avoid it there as well as in the LS environment, especially if the complementizer can aid planning? For them, a

concern about grammaticality might have overridden marginal benefits of the complementizer as a stalling device. Specifically, adults may avoid *that* in LO structures (where it is grammatical) to preclude the possibility of error in LS structures. This makes sense for a system that plans sentences by clausal units, since the filler and gap are in different clauses. The adults' strategy would thus be a global one to reduce the possibility of ungrammatical utterances. This strategy would be substantially more salient in this experiment than in normal language exercise: 25% of the trials involved the *that*-trace constraint. We return to this idea below in discussion of the adult SS processing, where we think heightened sensitivity to the *that*-trace configuration could be a factor.

We note here that an alternative interpretation of our data is that children are grammatically insensitive to the *that*-trace constraint (see footnote 2). If this is the case, it is still plausible that children's frequent choice of the overt complementizer is due to its use as a stalling device for this complex planning environment. Although our data do not distinguish between a planning account and a grammatical account of children's *that*-trace violations, this research does provide evidence against a grammatical account like Thornton's (1990) that claims that the overt complementizer is OBLIGATORY for children in LS structures; 47% of the Young children's, and 75% of the Older children's, LS utterances were complementizer-less, in spite of the complementizer's potential use as a stalling device.

(iv) *The SS complexity reversal*. Our data were orderly with respect to expected effects of age and structural type. The principal exception was the surprising elevation in adult non-fluency for SS structures, which represent the simplest of the planning challenges in our study. We have no conclusive explanation for this, but we suggest that in this case as well, differences in advance planning in adults and children may play a role. It has long been argued that non-fluencies arise at different levels of processing (e.g. Goldman-Eisler, 1972; Butterworth, 1980; Bock, 1996): some non-fluencies arise at the message-to-sentence mapping stage, while others arise during subsequent mappings from a functional structure to the explicit syntactic and lexicalized form of the sentence. We assume that the overall complexity orderings that we found for several measures in the data are the product of these several processing factors. The relative weight of early and late stage processing may differ for adults and children on the assumption that adults 'see farther' into the upcoming planning space than children do. On these grounds, adults grapple with options at the explicit sentence planning level

91

for the early portions of the SS sentences (e.g. gap filling, overt vs. reduced relatives, lexical retrieval and phonological interpretation) that the children have not yet elaborated. And, we suggest that the processing load at exactly this point may be additionally exacerbated by the adult ability to foresee occurrence in SS of the *that*-trace configuration. Though with the relativizer, the configuration is licit, the sequence of elements is the same as in the illicit *that*-trace sequence, as shown in (15).

(15) *that*-trace sequences in SS and LS structures:
SS: pick up the robber *that* [*trace*] is touching the dog
LS: *pick up the robber that Grover thinks *that* [*trace*] is touching the dog

Recall that we discussed the concern of *that*-trace violations as an explanation of adult avoidance of explicit complementizers in LO sentences. The salience of the (potential) *that*-trace configuration in the experiment could lead to additional checking to ensure that utterance of the *that*-trace sequence is acceptable, or to a switch to the reduced relative structure. These issues do not arise, or do not do so immediately, for the other three structural types in which such problems of detailed planning for sentence form arise later in the utterance. On this analysis, it is extra processing capacity on the part of adults compared to children that leads paradoxically to the elevated onset times for the SS structures.

The overall picture that emerges from this investigation is that the architecture of the child and adult formulators are very much the same. Sentences are planned in similar ways and the planning points are the same. The difference between children and adults lies in the amount of advance planning they undertake, and possibly in the levels of simultaneous planning they can sustain during sentence formulation. A good analogy is the process of crossing a creek by jumping from stone to stone. Children and adults land on stones that are positioned in the same places. But adults figure out a path before starting across, whereas children do some of the figuring on the way.

REFERENCES

Beattie, G. (1980). The role of language production processes in the organization of behaviour in face to face interaction. In B. Butterworth (ed.), *Language production: vol. 1, speech and talk*, 69–107. London: Academic.
Bernstein, N. E. (1981). Are there constraints on childhood disfluency? *Journal of Fluency Disorders* **6**, 341–50.
Bloodstein, O. (1974). The rules of early stuttering. *Journal of Speech and Hearing Disorders* **39**, 379–94.

Bock, J. K. (1987). Coordinating words and syntax in speech plans. In A. Ellis (ed.), *Progress in the psychology of language, vol. 3*, 337–90. London: Erlbaum.

Bock, J. K. (1991). A sketchbook of production problems. *Journal of Psycholinguistic Research* **20**, 141–60.

Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review* **3**, 395–421.

Boomer, D. (1965). Hesitation and grammatical encoding. *Language and Speech* **8**, 145–58.

Butterworth, B. (1980). Some constraints on models of language production. In B. Butterworth (ed.), *Language production: vol. 1, speech and talk*, 423–59. London: Academic.

Dell, G. (1986). A spreading activation theory retrieval in language production. *Psychological Review* **93**, 283–321.

Ferreira, V. & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* **40**, 296–340.

Ford, M. (1978). Planning units and syntax in sentence production. Doctoral dissertation, University of Melbourne.

Ford, M. & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition* **6**, 35–53.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (ed.), *Language production, vol. 1: speech and talk*, 177–220. London: Academic.

Garrett, M. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (ed.), *Normality and pathology in cognitive functions*, 19–76. London: Academic.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech* **15**, 103–113.

Hairer, Martin (2008). Amadeus Pro (Version 1.3.3) [Computer software]. New York: HairerSoft.

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech* **14**, 277–88.

Jaeger, J. J. (1992). Not by the hair of my hinny hin hin: Some general properties of slips of the tongue in young children. *Journal of Child Language* **19**, 335–66.

Jaeger, J. J. (2004). *Kids' slips: What young children's slips of the tongue reveal about language development*. Mahwah, NJ: Erlbaum.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition* **14**, 41–104.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT.

Maclay, H. & Osgood, C. (1959). Hesitation phenomena in English speech. *Word* **15**, 19–44.

Marshall, J. C. (1979). Language acquisition in a biological frame of reference. In P. Fletcher & M. Garman (eds), *Language acquisition*, 437–53. Cambridge: Cambridge University Press.

McDaniel, D., Chiu, B. & Maxfield, T. L. (1995). Parameters for wh-movement types: Evidence from child English. *Natural Language and Linguistic Theory* **13**, 709–753.

McDaniel, D. & Lech, D. (2003). The production system's formulation of relative clause structures: Evidence from Polish. *Language Acquisition* **11**, 63–97.

McDaniel, D., McKee, C. & Bernstein, J. B. (1998). How children's relatives solve a problem for Minimalism. *Language* **74**, 308–334.

McKee, C. & McDaniel, D. (2001). Resumptive pronouns in English relative clauses. *Language Acquisition* **9**, 113–56.

McKee, C., McDaniel, D. & Snedeker, J. (1998). Relatives children say. *Journal of Psycholinguistic Research* **27**, 573–96.

Postma, A. & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research* **36**, 472–87.

Rispoli, M. (2003). Changes in the nature of sentence production during the period of grammatical development. *Journal of Speech, Language, and Hearing Research* **46**, 818–30.

Roberts, L., Marinis, T., Felser, C. & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research* **36**, 175–88.

93

Smith, V. L. & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language* **32**, 25–38.

Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language* **28**, 164–88.

Sturm, J. A. & Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Language and Speech and Hearing Services in Schools* **38**, 47–59.

Thornton, R. (1990). Adventures in long-distance moving: The acquisition of complex wh-questions. Doctoral dissertation, University of Connecticut.

Wall, M. J., Starkweather, C. W. & Cairns, H. S. (1981). Syntactic influences on stuttering in young child stutterers. *Journal of Fluency Disorders* **6**, 283–98.

Wijnen, F. (1990). The development of sentence planning. *Journal of Child Language* **17**, 651–76.

Wijnen, F. (1992). Niet vloeiende spraak bij kinderen: planningsproblemen op verschillende niveaus [Disfluent speech in children: Planning difficulties at different levels]. *Stem, Spraak, en Taalpathologie* **1**, 23–41.

APPENDIX: LIST OF EXPERIMENTAL MATERIALS

These targeted utterances are in the order the items were presented. Each begins with *Pick up*.

(1) SS the baby that is pulling the hen
(2) SO the fish that the man is licking
(3) LS the girl that Stitch thinks was kicking the cow
(4) LO the cat that Belle guessed the boy was patting
(5) SS the robber that is touching the dog
(6) SO the sheep that the doctor is rubbing
(7) LS the queen that Grover dreamed was washing the pig
(8) LO the duck that Big Bird thinks the princess was kissing
(9) SS the woman that is pushing the goat
(10) SO the bear that the king is hitting
(11) LS the pirate that Dorothy guessed was tapping the horse
(12) LO the wolf that Lucy dreamed the clown was hugging