Journal of Memory and Language **46,** 57–84 (2002)
doi:10.1006/jmla.2001.2797, available online at http://www.academicpress.com on IDEAL®

How Incremental Is Language Production? Evidence from the Production of Utterances Requiring the Computation of Arithmetic Sums

Fernanda Ferreira and Benjamin Swets

Michigan State University

The incremental approach to language production assumes that the production system interleaves planning and articulation processes. Two experiments examined this assumption. In the first, participants stated the sums of two two-digit numbers in one of three different kinds of utterances, the sum by itself, the sum followed by the sequence "is the answer," or the frame "The answer is" followed by the sum. Problem difficulty was manipulated as well, so that in some conditions, speakers could (in principle) state the tens component of the sum while planning the ones. Latencies to begin to speak were the same for all three utterance types and were affected by the difficulty of the problem as a whole. Utterance durations were unaffected by problem difficulty. In the second experiment, participants were induced to speak incrementally through the use of a deadline procedure. Both latencies and utterance durations were influenced by the difficulty of the problem. This latter finding supports a basic premise of the incremental approach: Speakers sometimes speak and plan simultaneously. Nevertheless, the language production system appears not to be architecturally incremental; instead, the extent to which people speak incrementally is under strategic control.

Key Words: language production; arithmetic; incrementality; syntax; phonology.

Almost since the modern era of psycholinguistics began in the 1960s, researchers have argued for a hierarchical model of language production (Fromkin, 1971; Garrett, 1975, 1976, 1980) in which processing proceeds from semantic intention to articulation. According to a recent version of this model (Bock & Levelt, 1994) activity begins in the Message Component when a speaker forms an idea of what he wishes to say. The Grammatical Component operates next, and is itself subdivided into two stages of processing. The "Functional Level" selects the meaning and syntactic features of words—so-called "lemmas" (Levelt, 1989)—and assigns them a grammatical role such as subject or object. Then, positional processing takes place: The word-forms corresponding to lemmas are retrieved, and the serial order of individual phrases is put together. The next component engages in phonological processing (but see F. Ferreira (1993) for evidence that the computation of the intonational and metrical features of an utterance *precede* retrieval of word-forms), and from there the system sends orders to the articulatory organs to produce speech sounds appropriately.

Any adequate model of language production must incorporate mechanisms to explain three important features of normal speech: First, utterances generally conform to the speaker's communicative goals. Second, speakers choose from among various ways they might express an idea. For instance, a speaker may say *Ten plus seven make seventeen*, or *Seventeen is the sum of ten plus seven*, and so on. The final choice reflects both the communicative needs of the speaker and the processing states of the various components of the language production system. Finally, although pauses, repetitions, and repairs certainly do occur in normal speech (Clark & Wasow, 1998), speakers are usually reasonably

This study was supported in part by grant SBR-9319272 and BCS-9976584 from the National Science Foundation, and an All University Research Initiation Grant from Michigan State University. We thank Marc Brysbaert for his advice and ideas that helped motivate this study, and Wim Fias for providing the materials. We also thank Gary Shrock, Richard Falk, Mel Johnson, Robin Revette, Leah English, Carrick Williams, and Janis Stacey for their assistance. Finally, we are grateful to Karl Bailey, Herb Clark, Doug Davidson, Vic Ferreira, Zenzi Griffin, John Henderson, and Mike Tanenhaus for helpful discussions about the studies, and the Stanford Language Users Group (SLUGs) for their comments on the work.

Address correspondence and reprint requests to Fernanda Ferreira, Department of Psychology, Psychology Research Building, Michigan State University, East Lansing, MI 48824-1117. E-mail: Fernanda@eyelab.msu.edu.

0749-596X/01 $35.00

fluent (Levelt, 1989). At the extreme, no competent speaker needs to pause before retrieving each successive word of his utterance.

In many recent models of how people talk, these features of speech are explained with the single assumption that language production is incremental (V. Ferreira, 1996; Levelt, 1989; Roelofs, 1998; Wheeldon & Lahiri, 1997). Once a piece of information at a level of processing becomes available, it triggers activity at the next level down in the production system (Levelt, 1989, p. 23: "Each processing component will be triggered into activity by a minimal amount of its characteristic input"), and this sequence may continue all the way to articulation. This "vertical" aspect of the architecture is coordinated with what we might term its "horizontal" activities: At the same time that a piece of information works its way from idea to articulation, other pieces are constructed and make their way through the system as well. Thus, incremental models assume that various levels can operate in parallel. At the same time, this vertical parallelism implies a certain amount of horizontal seriality. That is, because the system is incremental, information at a *particular* level of processing is not necessarily handled in parallel. For example, a word at the end of an utterance would not be syntactically available in parallel with one at the beginning, because the earlier word was shunted off for phonological processing after it was syntactically encoded.

To illustrate how an incremental system operates, consider a speaker who wishes to express the fact that 17 is the sum of 10 and 7. The processing events might take place as follows. First, assume that the concept TEN becomes activated first. That activation will trigger retrieval of the corresponding lemma, which will cause the lemma to take its place at the left edge of a syntactic frame. (Simultaneously, other concepts become activated, but their presence is not relevant at the moment.) These events will cause the word-form for TEN to be accessed, so the word *ten* may be articulated. A consequence of this architecture, then, is that the speaker could say *ten* before knowing exactly how the utterance will end. As Wheeldon and Lahiri (1997) state, "What a processor is doing with a particular fragment of an utterance should not be dependent on information available in later fragments of the utterance. For example, constructing the initial prosodic units of a sentence should not be dependent on how the sentence will end" (p. 361). Having now said *ten*, the speaker is committed to some sort of a global form in which the addends occur earlier in the utterance and the sum occurs later. Therefore, a likely utterance given these events is something like *Ten and seven make seventeen*.

The incrementality hypothesis is appealing because it accounts for all three features of human speech mentioned above. A concept corresponding to what a speaker intends to say becomes activated based on information contained in the message-level representation. The early activation of the concept leads to early placement in the utterance. Incrementality, then, explains at least in part how speakers choose from among different sentence forms for expressing their ideas. Syntactic variations are taken advantage of by the production system to accommodate the states of activation of information in the production system. And speakers are fluent because they do not need to pause for long periods of time in order to plan their utterances. Instead, in a highly parallel system, a concept may make its way vertically through the architecture while, simultaneously, other concepts are becoming activated and encoded. Therefore, speakers do not need to plan whether to say *ten and seven make seventeen* or *seventeen is the sum of ten and seven*. The decision is made by the availability of concepts in long-term memory.

Furthermore, the incrementality hypothesis has received a fair bit of empirical support in recent experimental studies of language production. The consequence that has been most thoroughly examined is the notion that syntactic forms are chosen to reflect states of activation of the production system. Bock (1986) demonstrated that a semantically primed word tends to occur early in a sentence. Thus, if an experimental participant is shown the word *worship* and then sees a picture of lightning striking a church, she is likely to produce a passive utterance such as *the church is being struck by lightning*. The idea is that the semantic prime caused the concept CHURCH to become activated. As a result (i.e., because the system is incremental), the word *church* grabbed the subject position of the sentence, which then committed the production system to an utterance in which the other nominal concept occurred later in the sentence, as in a passive, for example. Furthermore, if the system operates in this fashion, one would expect that fluent production would be easier when there is syntactic choice rather than when there is only one grammatical way to express a particular idea. V. Ferreira (1996) found evidence to support this implication of incrementality. Participants in his experiments saw the initial frame I GAVE on a computer monitor, and then either the word TOYS followed by CHILDREN or the word CHILDREN followed by TOYS. In another condition, instead of I GAVE, participants saw I DONATED. The idea is that if a verb permits either order of the postverbal arguments, production will be faster because the system can handle either sequencing of TOY and CHILDREN. In contrast, with a rigid verb such as *donate*, it is more difficult to produce a proper utterance when the concept CHILDREN becomes available sooner than TOYS. The results supported these predictions. Because the system is incremental, it likes to have syntactic choices.

Wheeldon and Lahiri (1997) conducted a more direct test of incrementality. Participants received a noun phrase such as *het water* (“the water” in Dutch) followed by the question *Wat zoek je* (“What do you seek?”), and their task was to answer the question in a full sentence using the noun phrase they were given (i.e., *ik zoek het water*). Participants were encouraged to begin to speak as quickly as possible. Wheeldon and Lahiri recorded utterance initiation times and found that latencies were shorter when the first phonological word of the utterance was less complex. The characteristics of the other phonological words had no effect. They concluded that the production system is radically incremental: What determines when a speaker begins to talk is how long it takes her to get the first phonological word ready. This conclusion is reinforced by results from their other three experiments, which demonstrated that latencies are affected by the complexity of the entire utterance when participants were asked to prepare carefully before beginning their utterances. Thus, if speakers are allowed to speak naturally, the system behaves incrementally. As they argue, “when it is possible to do so, speakers preferentially initiate articulation following the phonological encoding of the initial phonological word of an utterance” (p. 377).

On the other hand, some findings in the literature are inconsistent with incrementality. For example, Lindsley (1975) asked participants to respond as quickly as possible to a simple picture showing a transitive action (e.g., one person touching another). He found that speakers began to phonologically encode their utterances before they had syntactically encoded the object of a transitive action but not before they knew the verb. In another study, Meyer (1996) used a word distractor paradigm to examine how much information about the words of an utterance is accessed prior to articulation. Speakers (of Dutch) produced simple utterances and were presented with a spoken distractor that was either semantically or phonologically related to either the subject or object noun of the sentence. Meyer found that a semantic distractor for either noun increased initiation times, while a phonological distractor impaired performance only when it was related to the subject noun. Meyer concluded that sentence production requires the retrieval of the semantic/syntactic information associated with most of the utterance, but it only requires the retrieval of phonological information for the first word or phrase. Thus, Meyer’s experiments provide evidence that grammatical encoding requires simultaneous knowledge about both preverbal *and* postverbal material, a finding that is inconsistent with incrementality.

Further evidence that is inconsistent with incremental production comes from a study by Stallings, MacDonald, and O’Seaghdha (1998), who examined the tendency for sentences with main verbs such as *transferred* and *introduced* to occur in either a canonical or a shifted form (e.g., *Mary introduced the new neighbor to Bill* vs *Mary introduced to Bill the new neighbor*, respectively). They found that the more often a verb tended to occur in structures separating it from its complement, the more likely participants were to produce shifted structures. More importantly, latencies to begin to speak were longer when participants produced sentences that were incompatible with the main verb's shifting history, indicating that the structures competed to be the ultimate syntactic frame for the sentence. This conclusion in turn implies that language production is not always incremental: Not only were structures planned over fairly large domains, but even more than one structure became activated simultaneously.

Finally, just how fluent is real language production? Studies of naturally occurring speech have shown that speakers do tend to pause before major syntactic constituents. For example, Henderson, Goldman-Eisler, and Skarbek (1966) found that periods of fluency alternated with phases during which the speaker engaged in a great deal of pausing. Also, more demanding speaking tasks (e.g., interpreting a cartoon versus merely describing it) result in less fluent speech (Goldman-Eisler, 1968). Similarly, Ford (1982) examined spontaneous speech for pauses longer than 200 ms and found that they preceded about 20% of deep clauses. Holmes (1988) asked speakers to talk spontaneously on various topics and then asked another group of participants to read the utterances the former group produced. She found that pauses and hesitations occurred before complement and relative clauses in spontaneous but not read speech. These pauses suggest that speakers disrupt fluency in order to plan their speech (see also F. Ferreira, 1991, 1993), and the fact that they plan in units that are approximately clausal in size (see also Bock & Cutting, 1992) indicates that the system has nonincremental aspects. Finally, Clark and Wasow (1998) have shown that disfluencies in normal speech are quite common and, furthermore, that they are more likely to occur the more complex the upcoming constituent. Thus, although an "ideal delivery" (Clark & Clark, 1977) might be the speaker's goal, under most normally demanding speaking circumstances, pauses, hesitations, and other disfluencies are fairly frequent.

Where are we so far? On the one hand, experiments in which speakers are asked to begin to speak as soon as they can tend to show that people plan more than one word at a time, as do studies that examine pauses and other disfluencies in spontaneous speech. On the other hand, studies of syntactic choice in production seem to favor incrementality, as does the Wheeldon and Lahiri finding that only the characteristics of the first phonological word of an upcoming utterance influenced latencies to initiate speech. Evidence for quite radical incrementality comes from what might appear on the face of it to be a rather unlikely source, a study by Brysbaert, Noel, and Fias (1998) on numerical cognition. The Brysbaert et al. results motivated to some extent the experiments described in this article, so their study will be described in detail.

The participants in Brysbaert et al. (1998) stated the sum of a one-digit number and a two-digit number as quickly as possible. For example, someone might see "21 + 4" and respond "25". The problems were presented either with the larger number preceding the smaller or in the other order. In addition, Brysbaert et al. tested two groups of participants, native French speakers and native Dutch speakers. The reason for looking at these two groups is that the languages differ in whether the ones or tens are pronounced first for a number such as "twenty-five". In French, one says the equivalent of "twenty-five", but in Dutch one says the equivalent of "five and twenty". Thus, if participants are faster to initiate articulation of the sum when the format of the problem allows them to prepare the first phonological word ("five" or "twenty") of their utterance more easily, then we will have support for radical incrementality. This was indeed the result Brysbaert et al. reported: For no-carry items such as 21 + 4, French participants showed a 56-ms advantage for the 21 + 4 order, and Dutch participants showed a 51-ms advantage for the 4 + 21 order. Brysbaert et al. interpreted the results as follows: "the Dutch speakers try to get access to the unit of the response first, because they can start programming the pronunciation of the answer as soon as the value of the unit is known. In contrast, the French speakers have to capitalise on the value of the ten, which they must know before response execution can be started" (p. 67). This, then, is incrementality of the sort proposed by Wheeldon and Lahiri (1997): Speakers prefer the addends to be ordered so that the first phonological word of the sum can be computed first, allowing the calculations required to determine the other part of the sum to be done during articulation. These results were obtained even though carry and no-carry items were randomly intermixed in the experiment. Furthermore, when both French and Dutch participants were required to type their responses, the language difference disappeared: Both groups preferred to have the two-digit number followed by the one-digit number. This result is also consistent with incrementality, because in both languages numbers appear in print with the tens before the ones.

Overall, research in language production does not provide clear evidence one way or the other for the incrementality hypothesis. What does seem apparent is that under some circumstances speakers are able to begin to speak as soon as they have formulated the smallest bit of linguistic structure (a phonological word), but that in other situations speakers are more cautious and do not speak until a reasonably large chunk of an upcoming utterance has been planned. This observation reveals that the language production system is strategically incremental: An important component of speech planning is to determine whether the situation calls for "blurting" or for more careful planning. This perspective suggests an important empirical question for researchers in language production: Even when speakers find that it is in their interests to articulate as quickly as possible in order to hold the floor, do they still plan more than just the first phonological word of the utterance anyway? A finding that they do would suggest that the language production system mandates some planning regardless of the speaker's goals and strategies.

To examine these issues, we conducted two experiments that utilized an arithmetic task much like the one used in the Brysbaert et al. (1998) study. In our first experiment, participants responded to addition problems by producing three different types of sentence forms, just the answer itself, the answer at the beginning of a sentence, and the answer at the end of a sentence. This experiment yielded no evidence for incrementality: Participants planned the entire utterance before speaking in all three conditions (and latencies were systematically influenced by the difficulty of the arithmetic problem). The second experiment was designed to strongly encourage participants to adopt an incremental strategy. Speakers had to begin to speak before a variable deadline. This manipulation drastically reduced utterance initiation times and yielded some evidence for incrementality; yet, clear evidence also emerged that speakers planned the sum even though it was sentence-final. In the General Discussion, we argue that these results support only moderate incrementality. Moreover, they suggest that even when speakers are trying to speak incrementally, the system still engages in a certain amount of utterance planning.

EXPERIMENT 1

The goal of this experiment was to examine incrementality in language production by seeing whether we could take the Brysbaert et al. (1998) effect one step further. Participants added together two two-digit numbers and stated the sum in the form of an utterance. We varied the difficulty of computing the tens column and of computing the ones column independently. In addition, participants were told to state their answers in one of three ways: The utterance could consist of just the sum, a sentence in which the sum was the grammatical subject, or a sentence in which the sum was a grammatical object. Recall that what Brysbaert et al. reported was that speakers preferred to have the addends arranged so that the leftmost addend allowed them to compute and articulate just the first phonological word of the sum. An extension of this result would be a finding that, in the sum-alone and sum-first conditions, initiation times are influenced only by the difficulty of calculating the tens column. Moreover, even if this quite radical incrementality is not observed in the experiment, the design allows us to test whether speakers are at least somewhat incremental. If they are, then problem difficulty (of the entire problem, not just the tens column) should influence initiation times but only in the conditions in which the sum occurred utterance-initially.

Method

Participants. All participants tested for the experiment were undergraduate students at Michigan State University, and all participated in exchange for partial credit in their introductory psychology courses. In the sum-only condition, 40 participants were tested in order to end up with 32 participants whose data could be used. To be included in the analyses, a participant had to produce correct answers to more than 75% of the problems. This criterion eliminated 5 participants; another 3 people were excluded because they did not speak loudly enough to trigger the voice-key. For the sum-frame condition ("SUM is the answer"), 3 people were not included in the analyses because they made too many errors, and 2 because of voice-key problems. For the frame-sum condition ("The answer is *SUM*"), 4 participants were excluded because of high error rates, and 1 because of voice-key errors. Thus, in total, data from 96 participants were analyzed, 32 in each of the three between-participants conditions.

Materials. Two different stimulus lists were created,[1] and a given participant saw only one of those lists. Each list contained 168 problems: 56 of those did not require a carrying operation (no-carry problems), 56 did (carry problems), and 56 were filler problems. The no-carry problems occurred in four different conditions: for 14 of them, computation of the tens part of the sum and the ones part of the sum was easy; for another 14, computation of the tens was easy and of the ones was hard; for another 14, computation of the tens was hard and of the ones was easy; and for the final 14, computation of both tens and ones was difficult. A column was considered easy if the sum was smaller than 5 and hard if the sum was greater than 5 (Geary, Bow-Thomas, & Yao, 1992). (Problems for which either column summed to 5 were excluded from the no-carry and carry sets.) An example of each problem type is given in Table 1. The carry problems also participated in four conditions, "easy", "hard", "harder", and "hardest". This continuum is based on the size of the resulting sum (Ashcraft, 1992). There were three types of filler problems, those that included an addend smaller than 20 but did not require a carrying operation, ones that included an addend less than 20 and that did necessitate carrying, and ones in which an addend was divisible by 10. The no-carry, carry, and filler problems were put into a random order, constrained so that two problems of the same type (same condition or filler type) did not occur in a row. The lists were identical, except that the order of addends was swapped from one list to the other.

Procedure. Participants were tested individually. The session began with the experimenter reading the instructions. Participants were told that they would see an addition problem on the computer monitor, and their task was to state the sum as quickly as possible into the microphone. They were reminded that they should respond as soon as possible, but that they should make sure they had the correct answer. A trial began with a fixation cross located in the center of the screen. One second later, the problem was presented with both addends on the same line separated by a plus sign, with spaces around the plus. The participant responded and then pushed a button to proceed to the next trial. He or she was allowed to correct the response initially given. The experimenter pushed one key on the key-

[1]All the materials were generously provided by Wim Fias of the Catholic University of Leuven, and Marc Brysbaert, now of Ghent University.

TABLE 1

Examples of the Problems Used in Experiment 1

No-carry problems		Carry problems		Filler problems	
Condition	Example	Condition	Example	Type	Example
Easy 10s, easy ones	21 + 22 = 43	Easy	23 + 28 = 51	Addend < 20, no-carry	16 + 11 = 27
Easy 10s, hard ones	21 + 26 = 47	Hard	23 + 38 = 61	Addend < 20, carry	17 + 36 = 53
Hard 10s, easy ones	21 + 62 = 83	Harder	23 + 58 = 81	Addend divisible by 10	12 + 70 = 82
Hard 10s, hard ones	21 + 66 = 87	Hardest	23 + 68 = 91		

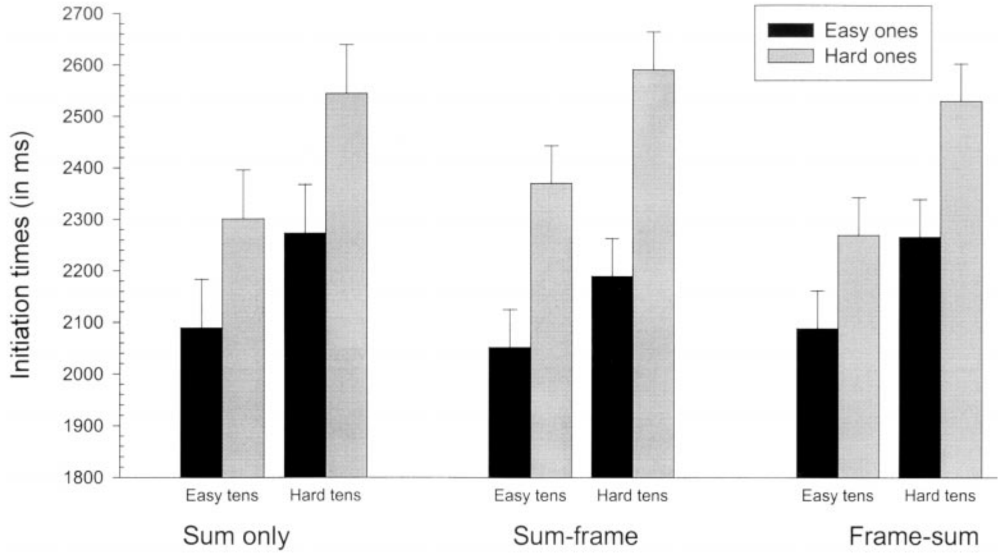


FIG. 1. Initiation times (in ms) for no-carry problems, Experiment 1.

board if the answer given was correct, and another if it was incorrect. The sessions were tape-recorded to allow those experimenter decisions to be checked off-line, and to allow the durational properties of the utterances to be analyzed at a later point.

For the sum-only condition, participants were asked to state the sum by itself. For the sum-frame condition, they were asked to give their answers "in the format 'X is the answer', where X is the sum." They were then given an example ("if given 25 + 25, you would say 'Fifty is the answer'"). For the frame-sum condition, they were asked to state their answers "in the format 'The answer is X', where X is the sum." They were then given an example ("if given 25 + 25, you would say 'The answer is fifty'").

An experimental session lasted about 45 min, and participants were free to take a break between trials whenever they wished. The session began with a practice session consisting of 30 problems (with the same characteristics as the fillers) and ended with a debriefing in which they were told the purposes of the study.

Results and Discussion

Latencies and accuracy data, no-carry problems. A trial was excluded if the response time was less than 300 ms or greater than 20 s (2% of the data were excluded) and if the response was incorrect. Latencies were analyzed as a 3 × 2 × 2 mixed factorial. The first variable is between-participants and represents the three utterance types (sum-only, sum-frame, frame-sum). The other two variables are within-participants and concern the difficulty of the problem: The tens column was either easy or harder to calculate, and similarly for the ones. (The two different lists produced the same results, so all analyses collapse over this variable.) All reported effects are significant at $p < .05$.

The latencies for the no-carry items are shown in Fig. 1. People took longer to begin saying the answer when it was difficult for them to calculate the tens and when it was difficult to calculate the ones. The effect of the two variables was additive, and it was identical for all three between-participant conditions. The statistical analyses are as follows: Participants took longer to initiate production of the sums when the tens were more difficult (2397 vs 2194 ms), $F(1,93) = 24.22$, MSe = 164606, and longer when the ones were more difficult (2433 vs 2159 ms), $F(1,93) = 101.86$, MSe = 70999. The two variables did not interact, $F(1,93) = 2.44$, MSe = 52436. There was no main effect of utterance type, $F < 1$, and no significant interactions between this variable and the others.

The average latency was 2299 ms in the sum-only condition, 2307 ms in the sum-frame condition, and 2281 in the frame-sum condition, $F < 1$.

The latency data place constraints on the extent to which the language production system is incremental, because the difficulty of both the tens *and* the ones contributed to initiation times. In other words, participants began their utterances only once they knew both parts of the sum, not just the part corresponding to the first produced phonological word. (Recently, Brysbaert and Fias have conducted a study identical to this one but with Dutch-speaking participants from the Catholic University of Leuven. They have reported the same results: both the difficulty of the tens and the difficulty of the ones influence time to initiate production of the sums.) Perhaps more surprisingly, the type of utterance in which the participant had to state the sum did not matter at all. Indeed, not only was the pattern the same across conditions, the absolute reaction times were remarkably similar as well. It does not appear, then, that participants produced their utterances in an incremental fashion. If they had, then latencies in the frame-sum condition would have been the same for the different problem types, because latencies would have been controlled by the time to produce the same phonological word across all conditions ("The answer" or "the answer is"). In addition, the requirement to produce not just the sum but also to remember to place it in a particular sort of frame does not seem to have affected latencies either (and we will see next that it does not affect accuracy).

The accuracy data are given in Table 2. Accuracy was unaffected by the between-participants manipulation: People were 95, 94, and 94% accurate in the sum-only, sum-frame, and frame-sum conditions, $F < 1$. There was no main effect of the difficulty of the tens or the difficulty of the ones, but the two variables did interact significantly, $F(1,93) = 7.04$. When the tens were easy, difficulty of the ones did not matter; when the tens were difficult, accuracy was worse when the ones were hard as well.

The accuracy results make clear that the requirement to produce the sum in a carrier frame did not impose the sort of burden on participants that would result in lower accuracy. Furthermore, the need to hold the sum in memory while saying "the answer is" does not seem to have caused any particular burden either, as indicated by the equivalent accuracy in the sum-frame and frame-sum conditions.

Latencies and accuracy data, carry problems. Three percent of trials were eliminated from the analyses because of response times less than 300 ms or greater than 20 s, or because of an incorrect response. Latencies were analyzed as a 3×4 mixed factorial: The first variable is between-participants and represents the three utterance-type conditions. The other variable is within-participants and concerns the difficulty of the problem as reflected in the progressively increasing size of the sum (easy, hard, harder, hardest).

The latencies for the carry problems are shown in Fig. 2. Latencies were unaffected by utterance type, $F < 1$. In general, and equivalently across all three between-participant conditions, latencies were longer the larger the sum, $F(3,93) = 45.97$, MSe $= 231628$: latencies were 3145, 3580, 3839, and 3862 in the easy, hard, harder, and hardest conditions. Accuracy too was unaffected by utterance type ($F < 1$), but it did vary according to problem difficulty, $F(3,93) = 7.78$, MSe $= 0.0099$. People were 87% accurate in the easy and hard conditions,

TABLE 2

Percentage Correct for No-Carry Problems, Experiment 1

	Utterance type					
	Sum-only		Sum-frame		Frame-sum	
	Easy ones	Hard ones	Easy ones	Hard ones	Easy ones	Hard ones
Easy tens	94	94	94	94	93	94
Hard tens	96	94	97	92	96	93

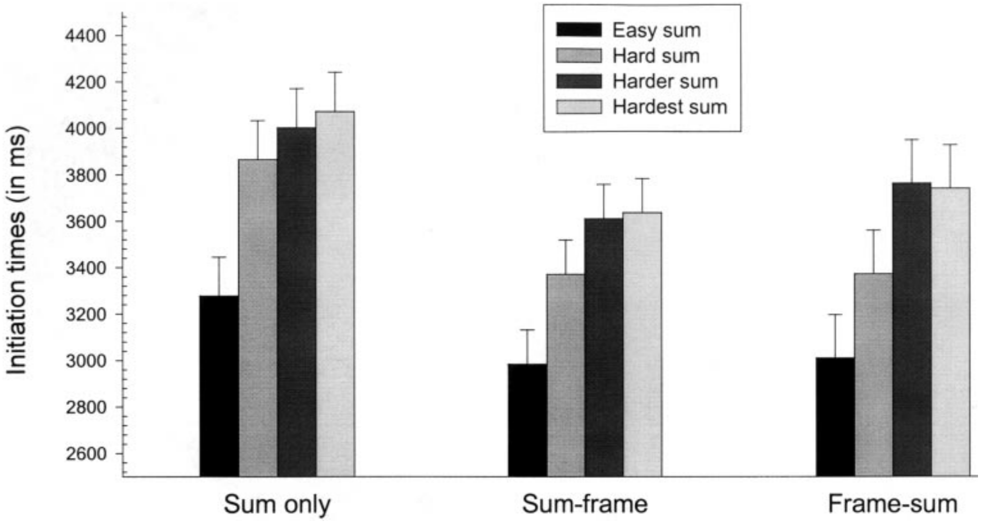


FIG. 2. Initiation times (in ms) for carry problems, Experiment 1.

and 81 and 83% accurate in the harder and hardest conditions.

The data from latencies and accuracy for both the no-carry and carry problems provide little support for the notion that the language production system is highly incremental. To explore the incremental hypothesis further, the durations of the utterances were measured and analyzed in order to assess whether participants computed the sum online during utterance production (i.e., as they were saying "the answer is") as well as prior to its initiation. Disfluencies were analyzed as well.

Utterance durations. Data from the frame-sum condition were analyzed in order to assess whether speakers might have engaged in some planning, or checking of their answers, as they produced the frame component of their utterances. Unfortunately, not all the recordings made of the experimental sessions were of uniform quality, and therefore only 26 of the 32 participants whose data were included in the analyses for the frame-sum condition could be included in these analyses of duration and disfluencies. (Analyses of the latencies and accuracy levels of just these 26 participants determined that this subgroup showed the same pattern as the larger one.) The utterances were digitized at a rate of 20 kHz using Computerized Speech Laboratory (CSL, from Kay Elemetrics), and durations were measured using CSL's waveform editor. The utterance was divided into parts as follows, the sequence *the answer is*, the sum, the first digit of the sum, and the second digit of the sum. Any pause time before the first digit was included as part of its duration, and any pause time before the second digit was included as part of the second digit's duration. One concern is that the digits themselves are different in the different experimental conditions (see Appendix A), and it is possible that some of the differences in sum durations are due to the digits themselves and not the experimental manipulations. We will present the data anyway because we wish to explore any measure that might potentially reveal incrementality, and because the results are not discrepant from those for *the answer is*. Results are shown in Table 3.

For the no-carry utterances, the duration of *the answer is* was longer when the tens were *easy* to compute than when they were more difficult (475 vs 445 ms), $F(1,25) = 16.84$, MSe = 1431. Also, the duration was longer when the ones were difficult (454 vs 467 ms), $F(1,25) = 3.88$, MSe = 1489, $p < .10$. These two variables did not interact, $F < 1$. Clearly, the pattern here

TABLE 3

Durations of Utterances in the Frame-Sum Condition (in ms), Experiment 1

	Duration of the answer is	Duration of sum	Duration of digit 1	Duration of digit 2
No-carry problems				
Easy 10s, easy ones	466	654	333	321
Easy 10s, hard ones	485	647	337	310
Hard 10s, easy ones	442	685	346	339
Hard 10s, hard ones	448	688	348	340
Carry problems				
Sum easy	504	713	390	323
Sum hard	483	754	420	334
Sum harder	501	751	413	338
Sum hardest	485	758	422	336

is not straightforward. It appears that speakers took longer to get through the frame when they noted that the first part of the sum would be easy to compute, and at the same time, there was some tendency for the speakers also to take longer when the ones were difficult.

The duration of the sum was longer when calculation of the tens was difficult vs easy (687 vs 650 ms), $F(1,25) = 8.33$, MSe = 2858. The duration of the first digit of the sum (the tens) was longer when the tens were difficult (347 vs 335 ms), $F(1,25) = 2.87$, MSe = 1222, $p < .10$, and similarly for the duration of the second digit—the second digit was longer when the tens had been difficult (340 vs 315 ms), $F(1,25) = 6.84$, MSe = 2273. It appears that when the tens were more difficult speakers stretched out the time they spent saying the sum.

For the carry utterances, the durations of *the answer is*, the first digit, and the second digit were unaffected by problem difficulty (see Table 3), all Fs < 1.

Analyses of disfluencies for frame-sum utterances, both problem types. A laboratory assistant unfamiliar with issues in psycholinguistics and cognitive science listened to the utterances and noted any disfluencies. Disfluencies were categorized into three major types, based on the type of editing term the speaker used (Clark & Wasow, 1998): “uh”, “um”, and other (e.g., words such as “geez” and profanities). There were five potential locations for a disfluency, before the utterance, after “the”, after “answer”, after “is”, and after the first digit of the sum. (No disfluency term occurred inside a word.) Overall, editing terms of any kind were quite rare. Of the 3854 utterances that were digitized for the duration and disfluency analyses, only 69 contained a disfluency (and only 2 of those involved the term “um”), and half of those occurred after the word “is”. A total of 3530 utterances were associated with correct responses; 63 (1.78%) included some disfluency. A total of 324 utterances were associated with incorrect responses; only 6 (1.85%) included a disfluency.

In sum, this experiment yielded little evidence for incrementality. A skeptic could argue, however, that the results are due to the nature of our task. The experimental situation calls for the speaker to produce error-free utterances, and the speaker has no need to begin to speak quickly. The problems themselves were quite difficult, particularly the carry problems. Therefore, calculation might have required all the speakers’ attention and not permitted the participants to “dual-task” such that they were speaking and calculating at the same time. The second experiment, then, was changed so as to encourage participants to speak incrementally as much as possible.

Before turning to the second experiment, we will address one final issue concerning these results. Recall that our goal was to try to extend Brysbaert et al.’s (1998) result that seemed to support radical incrementality. If French-speaking participants (for instance) preferred the $xx + x$ order because it allowed them to articulate the tens while calculating the ones column, then perhaps with two two-digit addends our speakers of English would plan only the tens component of the sum before beginning to speak. One possible challenge to this logic that can almost certainly be dismissed centers around the language difference across the two studies. This lead is almost certainly unpromising, because in all relevant respects English and French are alike. A more likely possibility focuses around the difference associated with adding two two-digit addends versus adding a one-digit and a two-digit addend. The former calculation is more difficult, and perhaps is too demanding to allow the dual-tasking (articulation and calculation) that incrementality in this task demands. The next experiment provides us with an opportunity to evaluate this suggestion, because we included the simpler mixed addend problems that Brysbaert et al. used and varied the order in which the addends occurred.

EXPERIMENT 2

Because of the surprisingly strong nonincremental effects obtained in Experiment 1, we made changes to the experimental task in order to maximize the chances that incremental behavior could emerge. First, problems were made easier. Second, by introducing a timing bar to the task, we gave participants an incentive to begin to speak quickly rather than to wait until they felt entirely ready. Finally, the stimulus problems were left on the screen when participants began to speak, unlike in Experiment 1, in which participants' voices caused the stimuli to disappear. All utterances were produced in the form "The answer is SUM" (see the section on pilot work below).

Method

Participants. Fifty participants were tested, all of whom were undergraduates at Michigan State University receiving partial credit in their introductory psychology courses. To be included in the analysis, a participant had to produce correct answers to more than 75% of both the practice problems and the experimental problems. All participants met these criteria. An additional 4 participants were tested prior to the 50 whose data we are reporting. These 4 individuals were tested in pilot work designed to assess what would happen if speakers could use any utterance form they wished (see "Pilot instructions," below, for more details).

Materials. Two different stimulus lists were created, and a given participant saw only one of them. Each list consisted of 112 problems, 56 of the no-carry experimental items from Experiment 1 (all of which consisted of two two-digit addends), and 56 new no-carry problems consisting of a one-digit addend and a two-digit addend. Computation of the ones column could be easy or hard, and similarly for the tens column. A column was considered easy if the sum was between 1 and 5, and hard if the sum was between 6 and 9. Examples illustrating the conditions are shown in Table 4.

Problems were presented in a random order. The lists were identical, except that the order of addends was swapped from one list to the other. In any given list, half of the mixed addend problems had the one-digit problem to the left of the two-digit problem, and half had the addends in

TABLE 4

Examples of the Problems Used in Experiment 2

Mixed addends problems		Same addends problems	
Condition	Example	Condition	Example
Easy tens, easy ones	2 + 41 = 43	Easy tens, easy ones	21 + 22 = 43
	41 + 2 = 43		22 + 21 = 43
Easy tens, hard ones	5 + 24 = 29	Easy tens, hard ones	21 + 26 = 47
	24 + 5 = 29		26 + 21 = 47
Hard tens, easy ones	3 + 71 = 74	Hard tens, easy ones	21 + 62 = 83
	71 + 3 = 74		62 + 21 = 83
Hard tens, hard ones	4 + 75 = 79	Hard tens, hard ones	21 + 66 = 87
	75 + 4 = 79		66 + 21 = 87

the other order; for the same addend problems, half were arranged so that the smaller addend was to the left of the larger addend, and the remaining half of the problems were arranged in the other order.

Pilot instructions. The goal of this experiment was to allow participants to speak as incrementally as possible. As we were developing this experiment, we reasoned that speakers might even choose to use different utterance types to facilitate this goal. For example, if they knew the sum quickly and wanted to get it out of the way, they might choose to say "SUM is the answer"; if they needed extra time, they might prefer the other arrangement, and they might even want to include extra "filler" words to stretch out the amount of planning time they would have before having to articulate the sum (e.g., "I think that the answer to that one is SUM"). If this strategy were observed, we would have more evidence for incrementality; as Levelt argued (1989, p. 245), speakers might choose the form and even the content of their utterance to accommodate the way that an utterance is evolving as it is encoded left to right.

Thus, the first four participants we tested were told to use any utterance form they liked, except that we required them to place the sum into a complete sentence. No examples were provided, to prevent our biasing their responses. After four participants, it became clear that speakers did not wish to take advantage of the latitude we had given them; on all 448 trials (112 trials $\times$ 4 participants), the speakers chose the form "The answer is SUM" (the frame-sum condition of Experiment 1). The instructions for the actual experiment, then, which excluded these four participants (because this latter group received different instructions from the subsequent participants), required participants to use the frame-sum format for their responses.

Procedure. Participants were tested individually. The session began with the participants reading a set of written instructions. They were told that they would see arithmetic problems on the computer monitor, and that their task was to give the answer in the format "The answer is SUM". They were then given an example, e.g., "if you saw 25 + 25, you would say 'The answer is fifty'".

Participants were also warned that the problem on the screen would be accompanied by a timing bar, and that they should answer the question before the timing bar counted all the way down and produced a loud "beep" sound. The timing bar was a horizontal rectangle that gradually shrunk in size until it disappeared, whereupon a beep sound was emitted by the computer. The duration of the timing bar varied randomly between 2 and 4 s. Participants were told that they should begin to speak quickly to avoid being beeped. They were also assured that they could correct their response if they believed it to be incorrect.

The 2- to 4-s value range for the timing bar that we ultimately used for this experiment was the first range that we attempted, and because it worked so well we saw no need to change it. Our goal at the start was to pick a value for the lower bound of the deadline that was long enough that it could be beaten on almost all trials, because we did not want to eliminate a large proportion of our data based on the participants' having been beeped. The lower bound of 2 s was about the average time that participants required in the first experiment to begin to respond to the no-carry problems in the frame-sum condition. The 4-s upper bound was chosen simply to allow there to be trials on which participants would easily beat the deadline. After analyzing the data from the pilot participants it became clear that the deadline values we had chosen worked optimally: Response latencies were reduced by more than two-thirds, and accuracy was not compromised (see Results and Discussion).

Finally, participants were told that the problem would remain on the computer monitor throughout the trial. In Experiment 1, the problems disappeared as soon as participants began to speak, which may have made it difficult for people to time-share the tasks of articulating the frame of the utterance while computing the sum of the problem from memory.

A trial began with a fixation cross located in the center of the screen. Participants hit a button to bring up the addition problem. They were allowed to correct their responses if they wished, but only the first response was considered in categorizing a trial as correct or incorrect. The experimenter then typed in their response and hit "enter" to proceed to the next trial. The sessions were tape-recorded to allow utterance durations to be analyzed at a later point.

An experimental session lasted about 35 min. The session began with a practice session consisting of 32 problems (with the same characteristics and conditions as the experimental stimuli) and ended with a debriefing in which participants were told the purposes of the study. It is also relevant to evaluating the effectiveness of the timing bar to note that at the beginning of the practice trials, participants sometimes got beeped—that is, they allowed the timing bar to elapse. However, by the end of the practice trials, participants were never beeped, nor did they allow the timing bar to elapse during the experimental trials. Clearly, for whatever reason, participants found the beep aversive and performed on every experimental trial so as to avoid hearing it.

Results and Discussion

We will first describe the latency and accuracy data. We will begin with the problems consisting of two two-digit addends (same addend problems), because these data are the most comparable to the data collected in Experiment 1. Then we will proceed to the problems consisting of one- and two-digit addends (mixed addend problems). Next, we describe utterance durations for both types of problems. The final section summarizes the data on disfluencies. All effects are significant at $p < .05$ unless otherwise stated.

Latencies and Accuracy Data

For latencies, a given trial was excluded if the response time was less than 300 ms or greater than 5 s. Trials associated with incorrect responses were also excluded (trials were considered incorrect if the *first* response given as part of the utterance was the wrong sum). Latencies were analyzed as a $2 \times 2 \times 2$ factorial. All three variables were within-participants. The first variable was order of addends. The other two variables concerned the difficulty of the problem: The tens column was either easy or harder to calculate, and the same was true for the ones column.

Same addend problems. No differences based on order of addends were observed, so all analyses collapse over this variable. Latencies are shown in Fig. 3. First, latencies to begin articulation were much faster in Experiment 2 than they were in Experiment 1. Whereas the average latency for frame-sum, no-carry problems in Experiment 1 was 2281 ms (the exact same problems analyzed here), the average latency for the two two-digit addend problems in this experiment was 634 ms. This marked difference suggests that the changes implemented in the task were extremely effective in getting subjects to respond more quickly—latencies were reduced by over 70%. Even more surprisingly, this increase in speed did not compromise accuracy: People were correct on 96% of trials. As comparison of Table 2 and Table 5 reveals, accuracy was just as high in this experiment as in the comparable conditions of the first.

Contrary to the incrementality hypothesis, latencies were longer when the ones column was more difficult to calculate, $F(1,49) = 11.84$, $MSe = 6052$. For easy ones, latencies were 621

TABLE 5

Percentage Correct, All Problems, Experiment 2

	Same addends problems: Smaller addend first, Easy 1s	Same addends problems: Smaller addend first, Hard 1s	Same addends problems: Larger addend first, Easy 1s	Same addends problems: Larger addend first, Hard 1s	Mixed addends problems: Single-digit addend first, Easy 1s	Mixed addends problems: Single-digit addend first, Hard 1s	Mixed addends problems: Double-digit addend first, Easy 1s	Mixed addends problems: Double-digit addend first, Hard 1s
Easy tens	97	96	95	97	99	96	98	99
Hard tens	99	97	97	93	99	96	99	98

(Problem type)

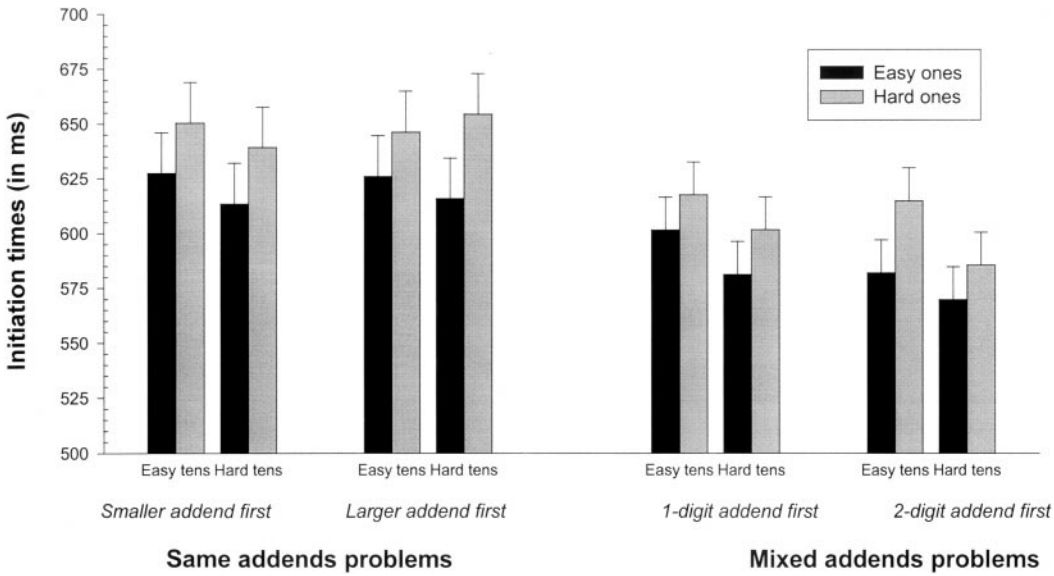


FIG. 3. Initiation times (in ms) for all problems, Experiment 2.

ms; for hard ones, 647 ms. The difficulty of the tens had no effect: latencies were 637 ms in the easy tens condition and 631 in the hard tens condition, $F < 1$. There was no interaction between difficulty of the tens and the ones, $F(1,49) = 1.13$, MSe = 2387, $p > .25$.

The pattern of results here is somewhat different from the pattern observed in Experiment 1: Only the difficulty of calculating the ones influenced latencies, rather than that of both the tens and the ones. This difference indicates that participants planned to a lesser extent in Experiment 2 than they did in Experiment 1, which is consistent with their being more incremental. On the other hand, the effect of ones difficulty makes clear that, to some extent, speakers planned their utterances all the way to the end before they began to speak. This would certainly seem to represent evidence against the argument that language production is radically incremental, even in a situation that encourages incrementality and discourages long-term planning.

The accuracy data are given in Table 5. Overall, accuracy was very high, and performance varied little over conditions (from 93 to 99% correct). Significant effects of the manipulated variables were observed, nonetheless. First, there was a main effect of addend order (97.1% for the order in which the small addend was to the left of the larger addend, and 95.6% for the opposite order), $F(1,49) = 5.00$, MSe = 0.0041. Second, there was a significant interaction between the difficulty of the tens and the ones, $F(1,49) = 5.39$, MSe = 0.0046. Third, order and difficulty of tens interacted, $F(1,49) = 5.02$, MSe = 0.0033. Finally, even the three-way interaction was significant, $F(1,49) = 5.58$, MSe = 0.0037. The one effect that is straightforward to describe is the interaction between the difficulty of the tens and the difficulty of the ones. If the tens were easy to calculate, then the difficulty of the ones did not matter; if the tens were more difficult, then accuracy was 98% in the easy ones condition and 95% in the harder ones condition. The rest of the accuracy differences are difficult to interpret. Fortunately, the differences are small, and our main concerns regarding accuracy are that error rates be relatively low (which they were) and that there be no evidence of speed–accuracy trade-offs (which there was not).

Mixed addend problems. The latencies for the problems consisting of a one-digit and a two-digit addend are shown in Fig. 3. Overall, participants were faster to begin speaking when presented with these problems than with those consisting of two two-digit addends (594 vs 634 ms), $F(1,49) = 20.61$, MSe = 15425. In addition, main effects were found for difficulty of both the tens and the ones columns. People were faster to begin to speak when the ones were easy to calculate than when they were hard to calculate (583 vs 605 ms), $F(1,49) = 8.54$, MSe = 5270. The effect for the tens column went in an unexpected direction: Participants were faster to begin their utterances when the tens were *difficult* rather than easy (603 vs 584 ms), $F(1,49) = 16.88$, MSe = 2223.

There is a suggestion in the data that people responded faster overall when given the two-digit addend before the one-digit addend (588 vs 600 ms for the opposite order), $F(1,49) = 3.56$, MSe = 4386, $p < .07$. Further evidence for this possibility comes both from the accuracy data and from comments made by participants to the experimenter during the experimental session. First, participants were more accurate when the two-digit number was the leftmost addend (98.4 vs 97.3% when the one-digit addend was on the left), $F(1,49) = 5.61$, MSe = 0.0023 (although a significant interaction between addend order and ones difficulty qualifies this main effect; see below). Second, after the experiment was over, participants commented that when a one-digit number was the first addend, its position to the left of the two-digit number made it difficult for them to conceptualize the one-digit number as belonging to the ones column. They reported that they sometimes incorrectly considered adding that number to the tens place of the second addend.

This result concerning order of the addends replicates the Brysbaert et al. (1998) effect: People prefer the $xx + x$ order over its opposite. They have a tendency with the other order to try to add the single addend to the tens column of the second addend because they are trying to deal with the columns in the order in which they will be spoken. Of course, the fact that this finding emerges in the latency data—at a point well before the speaker actually has to articulate the sum—suggests that this preference is in the planning and not in the articulation. In other words, when speakers *plan* their utterances, they prefer that material unfolds in the ultimate, to-be-articulated order. As will be described below, the preference for the $xx + x$ order is evident in the duration data as well, and these results are more compatible with actual incremental production.

Overall, the main results for the mixed-addend problem types clearly show that even for these easier problems, participants do not begin to speak until they have taken account of at least some aspects of the sum. Even with a great deal of incentive to speak incrementally, participants showed signs of planning all the way to the final word of the utterance.

The accuracy data are given in Table 5. Participants were more accurate when the ones were easier to calculate (98.7 vs 97.0%), $F(1,49) = 9.04$, MSe = 0.0033, and there was no effect of tens difficulty. A significant interaction was found between the order of addends and the difficulty of the ones, $F(1,49) = 5.57$, MSe = 0.0023. Participants were equally accurate for both orders of addends when the ones were easy (98.7%), but if the ones were hard, participants preferred to have the two-digit addend to the left of the one-digit addend. Accuracy was 98.1% for the former order and 95.9% for the latter.

To summarize, latency data from both types of problems revealed that, despite all of the changes made to the procedure to induce participants to speak incrementally, people nonetheless persisted in showing nonincremental, planning effects from the difficulty of the problems. First, people took longer to begin to speak for the more difficult two two-digit addend problems than for the mixed addend problems. Second, difficulty of the ones affected initiation times for both types of problems. While these data are difficult to reconcile with a radically incremental view of language production, they may be consistent with a more moderate version. In order to test whether participants planned and calculated during articulation, thereby demonstrating some incremental tendencies, we analyzed utterance durations.

Utterance Durations

Utterances from the first 24 participants were digitized at a rate of 20 kHz, and durations were measured using a waveform editor. (We analyzed only 24 of the 50 participants' data because waveform measurements are extremely time-consuming to obtain, and based on the first experiment we estimated that 24 participants would suffice to provide stable estimates of these duration means.) Each utterance was divided into three parts, the sequence *the answer is*, the first digit of the sum, and the second digit of the sum. (As with Experiment 1, it is probably prudent to put more weight on the duration data for the frame than for the sum, because the digits of the sum differ across conditions, and therefore intrinsic durations are not controlled.) Incorrect trials and trials on which corrections to previously given incorrect partial and/or full answers were made were excluded from the data. Results are given in Table 6.

Overall, the duration of *the answer is* was longer for the more difficult two two-digit addend problems than for the mixed addend problems (750 vs 610 ms), $F(1,23) = 58.33$, MSe = 32334. The same pattern held for the duration of the first digit of the sum (479 vs 417 ms), $F(1,23) = 18.18$, MSe = 20534, and the duration of the second digit of the sum (435 vs 393 ms), $F(1,23) = 21.93$, MSe = 7851. These differences are the first bit of evidence that people are planning as they speak, and that calculation and articulation may go on in parallel.

Same addend problems. The duration of *the answer is* was longer when the tens were difficult rather than easy to calculate (765 vs 735 ms), $F(1,23) = 5.17$, MSe = 8211, and when the ones were difficult to calculate (771 vs 729 ms), $F(1,23) = 16.20$, MSe = 5211. This result suggests that while uttering the frame, participants were planning or computing both digits of the sum. No other significant effects were observed for the duration of *the answer is*. The data were similar for the duration of the sum's first digit. Its duration was longer when the tens were more difficult (506 vs 453 ms), $F(1,23) = 5.99$, MSe = 22044—that is, its own duration was affected by how difficult it was for the participant to come up with its value. The duration of the tens was longer also when the ones were more difficult (524 vs 435 ms), $F(1,23) = 15.57$, MSe = 24063. These two variables did not interact. For the duration of the sum's second digit, only one effect was significant: the interaction between difficulty of the tens and difficulty of the ones columns, $F(1,23) = 9.03$, MSe = 6313. When the tens were easy to calculate, the duration of the second digit of the sum was longer in the easy ones condition than in the hard ones condition (459 vs 400 ms). When the tens were harder to calculate, the opposite was true (435 vs 446 ms). Again, because different digits are spoken in the experimental conditions, the duration data for the problems must be interpreted with some caution.

Mixed addend problems. Participants took longer to say *The answer is* when the 1-digit ad-

TABLE 6

Durations of Utterances (in ms), All Problems, Experiment 2

	Same addends problems		
	Duration of frame	Duration of digit 1	Duration of digit 2
Easy tens, easy ones	703	414	459
Easy tens, hard ones	767	492	400
Hard tens, easy ones	755	457	435
Hard tens, hard ones	775	555	445

	Mixed addends problems					
	Single-digit addend first			Double-digit addend first		
	Duration of frame	Duration of digit 1	Duration of digit 2	Duration of frame	Duration of digit 1	Duration of digit 2
Easy tens, easy ones	581	385	421	579	341	399
Easy tens, hard ones	721	529	389	670	434	360
Hard tens, easy ones	647	355	420	521	346	413
Hard tens, hard ones	654	491	388	607	456	351

dend was presented first rather than second (626 vs 594 ms), $F(1,23) = 20.92$, MSe = 2263. This finding is similar to the results we reported in the latency data and it replicates those of Brysbaert et al. (1998). Participants also took longer to articulate the frame when the tens column was *easy* rather than hard (638 vs 582 ms), $F(1,23) = 38.52$, MSe = 3904, which replicates the odd effect we observed in the latency data. We also found a large effect of ones difficulty: Participants spoke more slowly when the ones were difficult to calculate rather than when they were easy (663 vs 557 ms), $F(1,23) = 59.68$, MSe = 9087. There was also a significant interaction between addend order and difficulty of the ones column, $F(1,23) = 6.70$, MSe = 2106. Difficulty of the ones had a larger effect when the addends were presented in the more difficult order, one-digit addend before two-digit addend.

Analysis of the duration of the first digit of the sum revealed only one main effect: When the ones were difficult, participants tended to take much longer to utter the preceding digit (477 vs 356 ms), $F(1,23) = 15.31$, MSe = 45801. The duration of the second digit was longer for the *xx* + *x* order than for the *x* + *xx* order (405 vs 381 ms), $F(1,23) = 7.00$, MSe = 3922, and longer when calculation of the ones (i.e., itself) was more difficult (413 vs 372 ms), $F(1,23) = 13.10$, MSe = 6259.

Clearly, the duration data overall reveal that even though speakers planned some aspects of even the very final word of their utterances before they began to speak, they did not prepare well enough to enable them to speak confidently without engaging in more planning during articulation. This pattern is critical, because it makes clear that the production system has both incremental and nonincremental aspects. In addition, even when people perform in a situation that almost forces them to speak as incrementally as possible, they still engage in some planning. Finally, the data make clear that it is possible for people to speak and calculate simultaneously.

Analyses of Disfluencies, Both Problem Types

Disfluencies were categorized into three major types, based on the type of editing term the speaker used (Clark & Wasow, 1998): “uh”, “um”, and other (e.g., words such as “geez” and profanities). There were five potential locations in which a disfluency might occur, before the utterance, after “the”, after “answer”, after “is”, and after the first digit of the sum. (No disfluency occurred inside a word.) If a disfluency occurred during an incorrect trial or during a trial in which a correction was made, that trial was not counted.

First, in this experiment as in Experiment 1, disfluencies overall were quite rare. Of the 2688 utterances that were analyzed (24 participants × 112 trials per participant), only 21 utterances contained any type of disfluency at all. Thus, it is important to keep in mind that these observations are based on a very small subset of all the data. The first point to make is that all but two of the 21 disfluencies occurred after “is” (right before the articulation of the first digit of the answer). The other two disfluencies occurred after the articulation of the first digit. Second, we note that of the 19 disfluencies in the postframe location, all but two occurred with problems involving a difficult calculation of the ones. Therefore, the disfluency data are compatible with the results observed in Experiment 1 and with the main findings of this second experiment.

Comparison of Results from Experiments 1 and 2

By comparing the latency and duration results from the two experiments, it is possible to assess in what ways people speak differently when they can plan carefully versus when they must grab the floor quickly. Figure 4 allows the comparisons to be made easily for the different problem types in both experiments. For the Experiment 1 results, all data are from the frame-sum condition only. The top set of four bars depicts the data for the carry problems from the first experiment. Clearly, speakers wait a very long time before beginning to speak when they are confronted with these difficult arithmetic problems. The next set of four bars represent the no-carry problems from the first experiment. As can be seen, latencies to speak are much shorter. And, as described in the Results for Experiment 1, utterance durations do not change in response to problem type, because speakers in the first experiment dealt with problem difficulty before

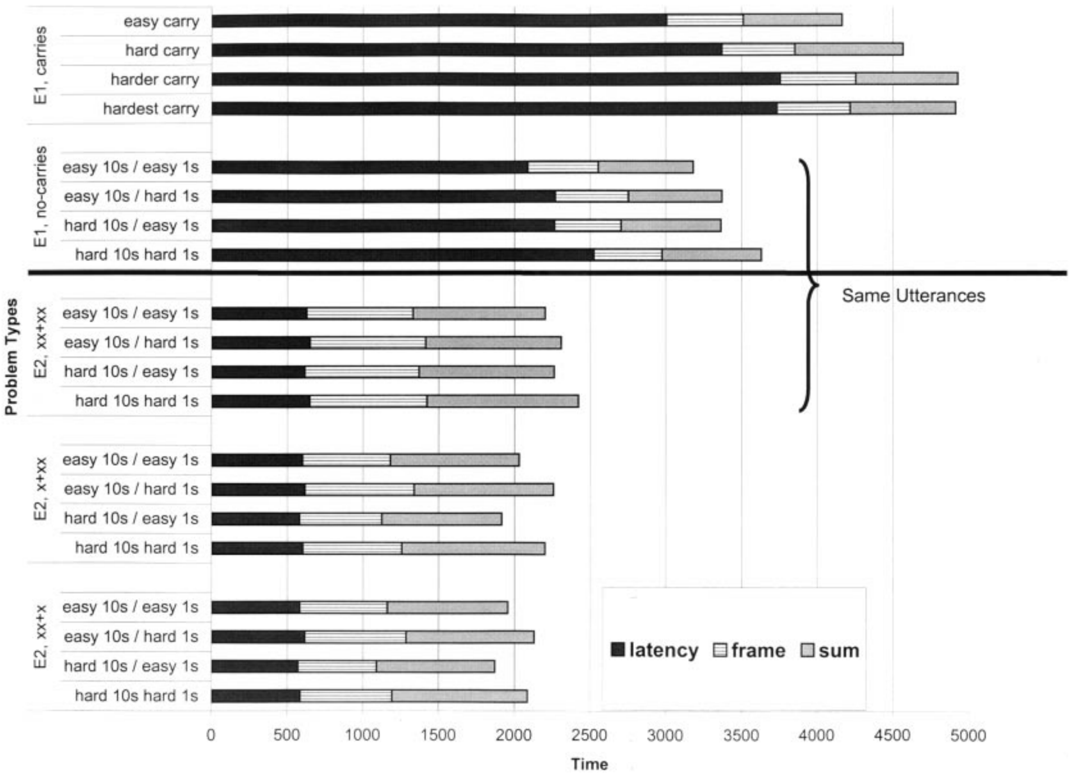


FIG. 4. Time speakers spent preparing and producing frame-sum utterances, both experiments.

they began to speak, not while they were speaking. Notice too how similar the striped and gray parts of the bars (the duration of *the answer is* and the sum) are for the top four bars and for the four bars immediately below. This visual impression is simply the result we reported in Experiment 1: Utterance durations were essentially the same in the carry and no-carry conditions.

The middle set of four bars shows the data for the no-carry problems from the second experiment. It is interesting to compare them with the four bars immediately above. The problems are the same (so the utterances are identical in content); what is different between the two experiments is that in the second, speakers were pressured to begin to speak quickly. Clearly, latencies to speak are much longer in the first experiment, where speakers had the luxury of planning. In addition, utterance durations are longer in the second experiment as well, indicating that some of the work speakers did not do before initiating speech ended up being carried out during articulation. In Experiment 1, the average duration of the entire utterance was 1232 ms; in Experiment 2, it was 1771 ms. This difference is statistically significant, $F(1,48) = 51.89$, MSe = 68795. Yet, although speakers extended the duration of the utterance by about 500 ms when they were pressured to begin to speak quickly, it is also clear from Fig. 4 that, overall, speakers in the second experiment spent less time planning and articulating their utterances overall than they did in the first experiment—about 1 s less. Recall also that speakers were no less accurate in Experiment 2, nor were they any less fluent. Thus, it can be argued that the deadline causes speakers to become more efficient, in that they can perform the same amount of work in less time. Interestingly, even though this appears to be true, it is also clear that, left to their own devices, speakers do not choose to speak in this maximally efficient manner.

Finally, the bottom two sets of four bars show the data for the mixed addend problems used in the second experiment. The level of resolution permitted by this graph does not allow the differences in performance to be seen easily. What the figure does reveal is that the differences due to problem difficulty in Experiment 2 are of a much smaller magnitude than the difference between carry and no-carry problems in Experiment 1 (and also the difference in performance due to the deadline).

The most important conclusion that can be drawn from these comparisons is that the difference in performance for the two experiments is not quite a pay-now/pay-later tradeoff: People in Experiment 2 did not need to extend their utterances by the same amount of time as they required for initiation in Experiment 1. People pay later if they choose not to plan, but their debt is smaller.

All in all, Experiment 2 yielded remarkably informative data about how incremental speakers are or can be. The one result that is less cooperative is the odd effect of tens difficulty for the mixed addend problems. (For the two-digit addend problems, this odd effect of tens difficulty was not observed: In the first experiment, latencies *and* durations were longer when the tens were difficult to calculate; in the second experiment, latencies were unaffected but durations were longer for difficult tens.)

This unexpected result is likely due to an artifact of the problems used in the mixed-addend condition. The easiest way to see this property is to examine Appendix B, which includes all the stimuli from Experiment 2. Consider only the mixed addend problems—the ones that yielded this unexpected result. Keep in mind as well that for these mixed addend problems, no calculation is required for the tens column. In the easy tens, easy ones condition, the digits in the addends for the tens place were 2, 3, 4, and 5; the digits in the addends for the ones place were 1, 2, 3, and 4. In the easy tens, hard ones condition, the digits in the addends for the tens place were 2, 3, 4, and 5; the digits in the addends for the ones place were 3, 4, 5, 6, and 7. Thus, in both of the easy tens conditions, the numerals that had to be added together were similar for the tens and ones. Now look at the hard tens problems. In the hard tens, easy ones condition, the digits used for the tens were exclusively 6 and 7; the digits used for the ones were 1, 2, 3, and 4. And in the hard tens, hard ones condition, the digits used for the tens were 6 and 7; the digits used for the ones were 2, 3, 4, 5, 6, and 7. Thus, there was much less overlap between the two columns; indeed, the tens column in the difficult tens condition virtually “popped out”. Recall the other data for the mixed addend problems demonstrating that one of the difficulties participants had in trying to answer them is that they had some trouble parsing the tens and ones columns properly. In the cases in which the tens were always 6 or 7, this task becomes much easier, and thus latencies and utterance durations were shorter. Therefore, this finding for the mixed addend problems is likely of little theoretical significance; instead, it appears to be a result of the unique properties of the stimuli used for these particular problems. Moreover, this property of the tens items did not dilute the effect of ones difficulty: Latencies and utterance durations were longer when the ones were more difficult to calculate, as would be expected.

GENERAL DISCUSSION

This section will be organized into three major parts. First, we will summarize the main results and discuss their implications for the incremental hypothesis. Second, we will describe a model of language production (F. Ferreira, 2000) that can account for why radical incrementality of the sort proposed by Wheeldon and Lahiri (1997), for example, was not observed in the experiments described here. Finally, we will briefly consider issues relating especially to the arithmetic task we required our participants to perform, focusing particularly on how our results compare to those obtained by Brysbaert et al. (1998).

Incrementality in Language Production

According to an incremental model of language production, people do not plan their utterances completely before they begin to talk. Instead, they "time-share" the tasks of planning and articulating, and therefore speak once they have prepared the earlier part of their utterance without necessarily having formulated later elements. According to a radically incremental model, speakers are incremental to the extent that they initiate articulation once they know the first phonological word of their utterances. Planning of the next phonological word takes place during that first one's articulation, and so on through to the end of the utterance.

The two experiments described here, taken together, support a limited type of incremental production. Incrementality is not an architectural property of the language production system; instead, it is a parameter of production that is under speaker control. In addition, even when a speaker has every incentive to initiate speech quickly, production latencies reveal influences of utterance-final material (in the simple, one-clause sentences we had our speakers produce). Thus, language production is not radically incremental—that is, speakers do not initiate production before having formulated at least some aspects of the utterance that go beyond the current phonological word.

The results that support these conclusions are the following.

(1) The task used in Experiment 1 provided little incentive for people to speak incrementally. Results showed that participants did not initiate speech until they had planned the sum carefully. The extent of planning was the same regardless of whether the sum occurred at the beginning of the utterance, at the end, or by itself. Utterance durations were not influenced by problem difficulty, suggesting that speakers did not engage in arithmetic calculations while articulating.

(2) The task used in Experiment 2 strongly encouraged people to speak incrementally. We found that utterance durations were influenced by problem difficulty. Therefore, it appears that speakers were calculating as they articulated. At the same time, latencies were also affected by problem difficulty, suggesting that speakers planned at least some aspects of the sum before speaking, even though the sum occurred at the end of the utterance. Planning effects were less extensive than they were in the first experiment, however. In Experiment 2, latencies were influenced by the difficulty of the ones column but not that of the tens; in Experiment 1, the difficulty of both column calculations affected latencies. For both experiments, more difficult problems overall were associated with longer utterance initiation times: In Experiment 1, utterances for no-carry problems were initiated sooner than those for carry problems; in Experiment 2, utterances for mixed addend problems (a one-digit and a two-digit addend) were initiated sooner than those for two two-digit addend problems.

(3) It might be argued that Experiment 2 did encourage people to speak *more* incrementally than they did in Experiment 1, but that people might be capable of speaking even more incrementally still. This argument cannot be rebutted definitively, but given that latencies in Experiment 2 were reduced by over 70%, and particularly given that they were about 650 ms on average, it is hard to imagine that people could initiate speech any faster. After all, when people are asked to simply name a single word on a computer monitor, their latencies are often not much lower than 600 ms or so (for example, Duffy, Henderson, & Morris (1989) reported latencies of 608 ms on average for participants to name the word that occurred at the end of a semantically neutral sentence, e.g., *the woman saw the MOUSTACHE*). Thus, the average latency in Experiment 2 is not much more than the time it would have taken participants simply to read and say out loud the phrase *the answer*. Yet, even with such short latencies, the reaction times managed to display influences of problem difficulty. Furthermore, accuracy was not compromised: Participants were as accurate in the second experiment as they were in the first. Neither were speakers any less fluent: Disfluencies occurred on 1.78% of trials in the first experiment, but on less than 1% of trials in the second.

(4) The phrasing *the answer is SUM* does not appear to be an unnatural way for speakers to state the sum of an arithmetic problem, as indicated by the results of the pilot study conducted in association with Experiment 2. Participants were free to use any utterance form they wished

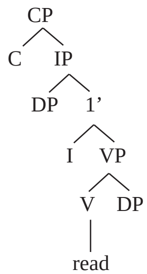
in which to state the sum, as long as they used a full sentence. Participants spontaneously hit on the form *the answer is SUM*, and they never wavered from it. This result should be viewed as tentative, because the experiment was not designed to test directly whether utterance form and content vary in response to online demands of utterance formulation. The result may be viewed as suggestive, however. It hints at the possibility that the following general story might be right: Speakers might sometimes change the form and content of an utterance online as they become aware that the utterance they are producing is not likely to end happily. For example, a speaker might change from *No one has to give money to a group that they don't like what they're doing* to something like *No one has to give money to groups whose actions they don't approve of* in order to avoid a subjacency violation (Chomsky, 1977). But they might be less likely to change the form and content of the utterance in order to give themselves more time merely to formulate some later part of an utterance; instead, the results of Experiment 2 suggest that they would choose instead to stretch out the utterance to give themselves the needed time.

Clearly, the language production system is not structured such that "processing at all levels occurs in an incremental fashion with a processor being triggered by any piece of characteristic input from the processors that feed into it" (Wheeldon & Lahiri, 1997, p. 361). In this "triggering" view, which is what we have termed "architectural incrementality" (and which permits radical incrementality), a module of the language production system that encounters information stated in its vocabulary is called into action automatically, so it performs its computations obligatorily (Fodor, 1983). Architectural incrementality is ruled out in favor of what might be termed strategic incrementality: A decision that every speaker must make is how to strike the appropriate balance between planning and initiating speech quickly. The finding that speakers are in principle capable of speaking incrementally without any cost in accuracy suggests that, in fact, speakers prefer to plan more carefully than they absolutely need to. Indeed, recall from Fig. 4 that utterances with identical content were overall shorter in the second experiment than in the first (adding together initiation plus articulation time). Thus, people are actually more efficient when they speak incrementally: They accomplish more in less time. For whatever reason, however, most people are not inclined to speak in this manner. Perhaps speaking this incrementally is simply too taxing (just as one may be capable of running a 7-min mile but may prefer a more comfortable 9-min pace).

Moreover, although the language operation can (but does not have to) operate incrementally, it does not appear that the system is radically incremental, in the sense that speech may be initiated once the content of just the first phonological phrase is known. Even in the second experiment, in which participants initiated utterances in about 600 ms, people clearly were engaging in arithmetic calculations as they articulated, and in general participants spoke in a maximally efficient manner; even so, latencies were still influenced by problem difficulty. These results are in line with what Griffin and her colleagues (1999, 2000) have observed as well: When participants had to produce utterances consisting of a stock carrier phrase plus an object name, latencies were influenced by the frequency of the pictured object's name even when the object name occurred at the end of the sentence (i.e., *They saw the OBJECT*).

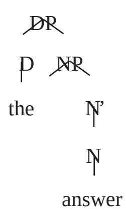
TAG-Based Model of Language Production

The model of language production proposed by F. Ferreira (2000) predicts that speakers will be unable to produce utterances in the radically incremental manner proposed by Levelt (1989) and Wheeldon and Lahiri (1997). This approach assumes that the representational format for syntactic information is a version of a Tree-Adjoining Grammar (TAG; Frank, 1992; Joshi, 1985; Joshi, Levy, & Takahashi, 1975; Kroch & Joshi, 1985). The basic unit of a TAG is the *elementary tree*, which consists of a lexical head and the arguments the head licenses. For example, access of the word *read* would result in activation of not just the word but its associated elementary tree as well, as shown below.



This structure assumes the analysis of clauses presented in Chomsky (1986), according to which a clause is an Inflectional Phrase (abbreviated as IP), and a full clause including the node for a complementizer is a Complementizer Phrase (CP). The abbreviation DP stands for Determiner Phrase (Abney, 1988). Other abbreviations are fairly standard, NP for noun phrase, PP for prepositional phrase, and VP for verb phrase.

The verb *read* is the lexical head and it licenses two arguments—a subject and an object. TAG assumes that a verbal head projects not only its own VP node but all the clausal projections as well (IP and CP). Thus, elementary trees are prototypically clause-like. Indeed, they are often described as corresponding roughly to a simple clause (Kroch, 1987), and as being similar to Chomsky’s (1955) original kernel sentences (Frank, 1992). Now let us examine the amount of syntactic structure that is retrieved when a noun such as *answer* becomes activated.



The determiner *the* brings along its DP node and licenses an NP argument. The noun *answer* comes with its NP structure, and by a process known as *substitution* (Frank, 1992), plugs into the NP slot provided by the DP. The critical point for our purposes here is that neither *the* nor *answer* can project any further; in particular, no clausal nodes are licensed by these lexical items. The implication is that this DP cannot be more than a DP—it cannot acquire a grammatical role such as subject or object, because it can receive no clausal assignment. The only way to get clausal nodes is for a verb to become available. Thus, once *is* is formulated, then a clausal tree like the one shown previously is accessed along with the lexical item, and then the DP *the answer* can be plugged into its subject position. It is only at that point that phonological encoding and articulation may begin.

This TAG-based approach to sentence production, then, predicts that incrementality cannot be more radical than the subject plus verb sequence, because a DP cannot get a grammatical role in the sentence until a verb is formulated. Radical incrementality is ruled out on this model. Of course, the results of Experiment 2 suggest that the system is even less incremental than the TAG model allows, because initiation times showed influences of postverbal material—i.e., the sum. This much planning occurred even though the task gave speakers strong incentives to speak incrementally. In the Introduction and in F. Ferreira (2000), a number of studies are discussed demonstrating that, under some circumstances, speakers initiate speech once they have formulated just the sentence’s subject and verb (e.g., Lindsley, 1975). However, Meyer’s (1996) study provided evidence that grammatical encoding requires the simultaneous formulation of both preverbal and postverbal material. A critical difference might be the type of verb that heads the clause: Lindsley’s study employed regular agent-theme verbs such as *touch*, whereas the sentences that speakers produced in Meyer’s study were headed by the copula *to be*. It is worth noting that TAGs treat copular verbs differently from other verbs, such that both “arguments” are in essence preverbal. Thus, the results reported here and in Meyer indicating that speakers plan beyond even the verb might have been observed because speakers used *to be* as the main verb in both studies. Of course, this hypothesis will not account for Griffin’s (1999, 2000) results (the utterances speakers produced in her study centered around verbs such as *saw*), so clearly this issue needs further investigation. What is important for our purposes here is that the TAG model puts a lower bound on incrementality: Speakers cannot initiate a sentence until they have formulated its subject and verb. Under some circumstances—perhaps when the verb is a copula, or perhaps when a demanding task such as arithmetic is required—speakers must formulate and encode even postverbal material.

Arithmetic and Language Production

One of the most important inspirations for the experiments reported here was the study by Brysbaert et al. (1998), which seems to support radical incrementality. Brysbaert et al. observed that speakers are faster to produce the answers to arithmetic problems consisting of a two-digit addend and a one-digit addend when the addends are arranged so that the first column calculated was also the first phonological word articulated. Thus, speakers of Dutch prefer the single digit addend to precede the double-digit addend, because numbers in Dutch are spoken so that the ones come before the tens. Speakers of French prefer the opposite order, because French works just like English: A number such as 44 is spoken so that the tens column precedes the ones column. This result seems consistent with radical incrementality.

The results of Experiment 2 are entirely consistent with the data (but not quite the interpretation) found in Brysbaert et al. (1998). We also found that, for mixed addend problems, speakers preferred the arrangement in which the two-digit number came first. Of course, we would give our result a different explanation, particularly because the preference for the $xx + x$ order emerged in the latency data, many syllables before speakers had to articulate the sum itself. It might be, then, that the Brysbaert effect concerns planning: Speakers want to formulate and encode their utterances roughly in the order in which the constituents will be articulated. In other words, even if people planned the entire utterance *The answer is SUM* before beginning to speak (to some extent), there is a question about the order in which that planning took place: Did they plan the sum and then *the answer is*, or did they plan in the other order? The finding that people preferred to receive the addends in the $xx + x$ order suggests that people might have planned the utterances roughly in the order in which they would be articulated. The Brysbaert et al. finding, then, can be viewed as demonstrating that, in general, speakers are inclined to plan in this manner. Indeed, this is the version of incrementality argued for by V. Ferreira (1996) in his work demonstrating that people speak more efficiently when they have choices regarding syntactic form than when they are constrained to just one syntactic structure. On this view of incrementality, active items grab earlier syntactic positions, and as a result, accessibility of concepts influences syntactic form. Thus, syntactic plans are preferentially built up in the order in which words become available. This type of incrementality does not imply that phonological encoding will take place on the smallest unit possible (viz. the phonological word), and it is compatible with our results showing that speakers plan the sum even in utterances in which the sum occurs at the end of the utterance.

Furthermore, it is critical to note that Brysbaert et al. (1998) did not provide any evidence that speakers were not influenced by the difficulty of both the tens and the ones calculations for the sum before beginning to speak. Their study was not designed to address this question, so the difficulty of the tens and the ones calculations were not independently manipulated as they were here. Indeed, we should emphasize that the Brysbaert et al. study was not originally designed or reported as a study of language production; instead, it is an important contribution to the literature on numerical cognition and linguistic relativity hypothesis (Sapir, 1941; Whorf, 1956). Their finding that speakers preferred the articulation order is consistent with the idea of radical incrementality but, as we argued above, is not mandated by it. The prediction that speakers of English would be influenced by just the difficulty of the tens column when producing utterances requiring the calculation of two-digit sums was an inference we drew from their reported work. As we have seen, it was not supported, probably because even in the original Brysbaert et al. study, speakers planned the entire sum before beginning to speak—they simply did so in the order in which they would articulate.

Another quite different possibility is that speakers in the Brysbaert et al. (1998) study were in fact speaking incrementally, rather than just planning in a left-to-right order. It might be that when speakers produce utterances that consist simply of nominals (i.e., just the sum) or other types of fragments, then the TAG-based clausal constraint that we described above does not operate. The constraint would be irrelevant because the utterance fragment consisting of only the sum does not need to be put into a clausal structure. In this situation, perhaps a single phonological word can be formulated and immediately produced, because the speaker is not producing a clause, so grammatical relations such as subject and object are undefined. Of course, this line of reasoning must come to terms with the finding from our first experiment that speakers in the sum-only condition did not show evidence of being influenced by just the tens column for the no-carry problems. Again, we can only speculate, but one possibility is that speakers were unable to behave as incrementally as they did in Brysbaert et al. because in our experiment participants had to add both columns, and for some problems they had to engage in carry operations as well. Perhaps this combination is critical: Many problems required carrying, making it difficult for the participants to be confident that they could deal with just the tens independently of the ones; *and*, all the problems required participants to add the tens column as well as the ones column. These two characteristics together might have led participants to believe that their most efficient strategy was to deal with the problems as a whole before beginning to speak. Let us emphasize that we are simply speculating at this point: Further research must be done on these topics before any definitive arguments can be made.

Conclusions

The two experiments that we have reported shed light on the important question of whether language production is incremental. The answer that is most consistent with the results reported here and in previous work is that the system is not architecturally incremental. Instead, the extent to which planning occurs is at least partly under speakers' control, and it depends on the intentions that motivate the speech. Moreover, even when speakers have incentives to initiate speech quickly, they still appear to engage in planning that goes beyond the immediate phonological word. Therefore, the system seems to be architecturally constrained to require planning beyond the initial phonological word, particularly for clausal utterances. At the same time, it is important to stress that language production can be incremental: The results of these experiments demonstrate that speakers are capable of planning upcoming portions of an utterance as they are articulating. This finding is especially striking given that the planning they engaged in was arithmetic calculation, because it might have been supposed that addition of two-digit numbers could not be carried out concurrently with utterance articulation. Apparently, the two can go on in parallel, at least for problems that do not require carrying operations. Thus, a fundamental premise of the incremental view is supported: The language production system is capable of interleaving planning processes and articulation.

APPENDIX A

TABLE A1

Problems Used in Experiment 1

No-carry problems			Carry problems			Filler problems	
Prob	Sum	Condition	Prob	Sum	Condition	Prob	Sum
21 + 22	43	EE	23 + 28	51	easy	30 + 11	41
21 + 26	47	EH	23 + 38	61	hard	60 + 15	75
21 + 62	83	HE	23 + 58	81	harder	13 + 79	92
21 + 66	87	HH	23 + 68	91	hardest	70 + 17	87
23 + 21	44	EE	29 + 23	52	easy	75 + 13	88
27 + 21	48	EH	39 + 23	62	hard	12 + 15	27
63 + 21	84	HE	59 + 23	82	harder	66 + 30	96
67 + 21	88	HH	69 + 23	92	hardest	54 + 11	65
21 + 24	45	EE	24 + 27	51	easy	18 + 36	54
21 + 28	49	EH	24 + 37	61	hard	12 + 58	70
21 + 64	85	HE	24 + 57	81	harder	18 + 72	90
21 + 68	89	HH	24 + 67	91	hardest	12 + 61	73
31 + 21	52	EE	28 + 24	52	easy	43 + 13	56
35 + 21	56	EH	38 + 24	62	hard	73 + 12	85
71 + 21	92	HE	58 + 24	82	harder	77 + 10	87
75 + 21	96	HH	68 + 24	92	hardest	77 + 19	96
21 + 32	53	EE	24 + 29	53	easy	19 + 77	96
21 + 36	57	EH	24 + 39	63	hard	17 + 49	66
21 + 72	93	HE	24 + 59	83	harder	10 + 78	88
21 + 76	97	HH	24 + 69	93	hardest	18 + 46	64
33 + 21	54	EE	27 + 25	52	easy	68 + 10	78
37 + 21	58	EH	37 + 25	62	hard	62 + 11	73
73 + 21	94	HE	57 + 25	82	harder	36 + 17	53
77 + 21	98	HH	67 + 25	92	hardest	36 + 18	54
21 + 34	55	EE	25 + 28	53	easy	12 + 51	63
21 + 38	59	EH	25 + 38	63	hard	40 + 43	83
21 + 74	95	HE	25 + 58	83	harder	60 + 12	72
21 + 78	99	HH	25 + 68	93	hardest	30 + 42	72
31 + 22	53	EE	29 + 25	54	easy	31 + 60	91
35 + 22	57	EH	39 + 25	64	hard	76 + 10	86
71 + 22	93	HE	59 + 25	84	harder	77 + 17	94
75 + 22	97	HH	69 + 25	94	hardest	51 + 30	81
22 + 32	54	EE	26 + 27	53	easy	13 + 55	68
22 + 36	58	EH	26 + 37	63	hard	18 + 13	31
22 + 72	94	HE	26 + 57	83	harder	30 + 33	63
22 + 76	98	HH	26 + 67	93	hardest	17 + 73	90
33 + 22	55	EE	28 + 26	54	easy	17 + 14	31
37 + 22	59	EH	38 + 26	64	hard	62 + 14	76
73 + 22	95	HE	58 + 26	84	harder	19 + 16	35
77 + 22	99	HH	68 + 26	94	hardest	16 + 11	27
23 + 22	45	EE	26 + 29	55	easy	18 + 17	35
23 + 26	49	EH	26 + 39	65	hard	50 + 16	66
23 + 62	85	HE	26 + 59	85	harder	11 + 61	72
23 + 66	89	HH	26 + 69	95	hardest	10 + 40	50
31 + 23	54	EE	28 + 27	55	easy	66 + 18	84
35 + 23	58	EH	38 + 27	65	hard	39 + 15	54
71 + 23	94	HE	58 + 27	85	harder	12 + 70	82
75 + 23	98	HH	68 + 27	95	hardest	55 + 14	69

TABLE A1—*Continued*

Problems Used in Experiment 1.

No-carry problems			Carry problems			Filler problems	
Prob	Sum	Condition	Prob	Sum	Condition	Prob	Sum
23 + 32	55	EE	27 + 29	56	easy	18 + 54	72
23 + 36	59	EH	27 + 39	66	hard	19 + 69	88
23 + 72	95	HE	27 + 59	86	harder	40 + 43	83
23 + 76	99	HH	27 + 69	96	hardest	12 + 42	54
31 + 24	55	EE	29 + 28	57	easy	14 + 17	31
35 + 24	59	EH	39 + 28	67	hard	45 + 30	75
71 + 24	95	HE	59 + 28	87	harder	71 + 10	81
75 + 24	99	HH	69 + 28	97	hardest	65 + 30	95

Note. EE, easy tens, easy ones; EH, easy tens, hard ones; HE, hard tens, easy ones; HH, hard tens, hard ones. The terms “easy”, “hard”, “harder”, and “hardest” refer to the size of the sum of the carry problems.

APPENDIX B

TABLE A2

Problems Used in Experiment 2

Mixed addend problems			Two-digit addend problems		
Prob	Sum	Condition	Prob	Sum	Condition
2 + 21	23	EE	21 + 22	43	EE
22 + 2	24	EE	23 + 21	44	EE
2 + 31	33	EE	21 + 24	45	EE
41 + 2	43	EE	31 + 21	52	EE
2 + 51	53	EE	21 + 32	53	EE
21 + 3	24	EE	33 + 21	54	EE
3 + 31	34	EE	21 + 34	55	EE
41 + 3	44	EE	31 + 22	53	EE
3 + 22	25	EE	22 + 32	54	EE
51 + 3	54	EE	33 + 22	55	EE
4 + 21	25	EE	23 + 22	45	EE
31 + 4	35	EE	31 + 23	54	EE
4 + 41	45	EE	23 + 32	55	EE
51 + 4	55	EE	31 + 24	55	EE
5 + 24	29	EH	21 + 26	47	EH
33 + 5	38	EH	27 + 21	48	EH
6 + 22	28	EH	21 + 28	49	EH
23 + 6	29	EH	35 + 21	56	EH
6 + 32	38	EH	21 + 36	57	EH
33 + 6	39	EH	37 + 21	58	EH
6 + 42	48	EH	21 + 38	59	EH
43 + 6	49	EH	35 + 22	57	EH
6 + 52	58	EH	22 + 36	58	EH
53 + 6	59	EH	37 + 22	59	EH
7 + 22	29	EH	23 + 26	49	EH
32 + 7	39	EH	35 + 23	58	EH
7 + 42	49	EH	23 + 36	59	EH
52 + 7	59	EH	35 + 24	59	EH

TABLE A2—*Continued*

Problems Used in Experiment 2

Mixed addend problems			Two-digit addend problems		
Prob	Sum	Condition	Prob	Sum	Condition
1 + 62	63	HE	21 + 62	83	HE
72 + 1	73	HE	63 + 21	84	HE
2 + 61	63	HE	21 + 64	85	HE
62 + 2	64	HE	71 + 21	92	HE
2 + 71	73	HE	21 + 72	93	HE
72 + 2	74	HE	73 + 21	94	HE
2 + 63	65	HE	21 + 74	95	HE
73 + 2	75	HE	71 + 22	93	HE
3 + 61	64	HE	22 + 72	94	HE
71 + 3	74	HE	73 + 22	95	HE
3 + 62	65	HE	23 + 62	85	HE
72 + 3	75	HE	71 + 23	94	HE
4 + 61	65	HE	23 + 72	95	HE
71 + 4	75	HE	71 + 24	95	HE
4 + 74	78	HH	21 + 66	87	HH
75 + 4	79	HH	67 + 21	88	HH
5 + 62	67	HH	21 + 68	89	HH
63 + 5	68	HH	75 + 21	96	HH
5 + 64	69	HH	21 + 76	97	HH
72 + 5	77	HH	77 + 21	98	HH
5 + 73	78	HH	21 + 78	99	HH
74 + 5	79	HH	75 + 22	97	HH
6 + 62	68	HH	22 + 76	98	HH
63 + 6	69	HH	77 + 22	99	HH
6 + 72	78	HH	23 + 66	89	HH
73 + 6	79	HH	75 + 23	98	HH
7 + 62	69	HH	23 + 76	99	HH
72 + 7	79	HH	75 + 24	99	HH

Note. EE, easy tens, easy ones; EH, easy tens, hard ones; HE, hard tens, easy ones; HH, hard tens, hard ones.

REFERENCES

Abney, S. (1988). A computational model of human parsing. Paper presented at the First Annual CUNY Conference on Human Sentence Processing.

Ashcraft, M. J. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, **44**, 75–106.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 575–586.

Bock, J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, **31**, 99–127.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.

Brysbaert, M., Fias, W., & Noel, M.-P. (1998). The Whorfian hypothesis and numerical cognition: Is “twenty-four” processed in the same way as “four-and-twenty”? *Cognition*, **66**, 51–77.

Chomsky, N. (1955). *The logical structure of linguistic theory*. Bloomington, IN: Indiana University Linguistics Club.

Chomsky, N. (1977). *Essays on form and interpretation*. New York: North-Holland.

Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.

Clark, H. H., & Clark, E. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, **37**, 201–242.

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 791–801.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, **30**, 210–233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, **100**, 233–253.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Cambridge, MA: MIT Press.

Ferreira, V. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, **35**, 724–755.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 797–827). Cambridge, MA: MIT Press.

Frank, R. (1992). Syntactic locality and tree adjoining grammar: Grammatical, acquisition, and processing perspectives. Ph.D. thesis, University of Pennsylvania.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, **47**, 27–52.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation*, **9**, (pp. 133–177). New York: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North-Holland.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production. Vol.1: Speech and talk* (pp. 177–220). New York: Academic Press.

Geary, D., Bow-Thomas, C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology*, **54**, 372–391.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Griffin, Z. M. (1999). The trade-off between lexical preparation and (dis)fluency for simple utterances. Talk presented at the USC Language Production Conference, Nov.

Griffin, Z. M., & Spieler, D. H. (2000). Speech planning in younger and older adults. Poster presented at the Annual Meeting of the Psychonomic Society, New Orleans, LA, Nov.

Henderson, A., Goldman-Eisler, F., & Skarbek. A. (1966). The common value of pausing time in spontaneous speech. *Quarterly Journal of Experimental Psychology*, **17**, 343–345.

Holmes, V. M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes*, **3**, 323–361.

Joshi, A. K. (1985). How much context-sensitivity is required to provide reasonable structural descriptions: Tree adjoining grammars. In D. Dowty, L. Kartunnen, and A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. (pp. 206–250). Cambridge: Cambridge Univ. Press.

Joshi, A. K., Levy, L., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, **10**, 136–163.

Kroch, A. (1987). Unbounded dependencies and subjacency in a tree adjoining grammar. In A. Manaster-Ramer (Ed.), *Mathematics of Language* (pp. 143–172) Amsterdam: John Benjamins.

Kroch, A., & Joshi, A. K. (1985). The linguistic relevance of tree adjoining grammar. Technical Report MS-CS-85-16, Department of Computer and Information Sciences, University of Pennsylvania.

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

Lindsley, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, **7**, 1–19.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, **35**, 477–496.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–participle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 904–921.

Sapir, E. (1941). In L. Spier (Ed.), *Language, culture, and personality: Essays in memory of Edward Sapir* Menasha, WI: Sapir Memorial Publication Fund.

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal order constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, **39**, 392–417.

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, **37**, 356–381.

Whorf, B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. New York: Wiley.

(Received July 31, 2000)
(Revision received January 22, 2001)
(Published online October 3, 2001)