BRIEF REPORT

Processes of incremental message planning during conversation

Sarah Brown-Schmidt · Agnieszka E. Konopka

Abstract Speaking begins with the formulation of an intended preverbal message and linguistic encoding of this information. The transition from thought to speech occurs incrementally, with cascading planning at subsequent levels of production. In this article, we aim to specify the mechanisms that support incremental message preparation. We contrast two hypotheses about the mechanisms responsible for incorporating message-level information into a linguistic plan. According to the *Initial Preparation* view, messages can be encoded as fluent utterances if all information is ready before speaking begins. By contrast, on the *Continuous Incrementality* view, messages can be continually prepared and updated throughout the production process, allowing for fluent production even if new information is added to the message while speaking is underway. Testing these hypotheses, eye-tracked speakers in two experiments produced unscripted, conjoined noun phrases with modifiers. Both experiments showed that new message elements can be incrementally incorporated into the utterance even after articulation begins, consistent with a *Continuous Incrementality* view of message planning, in which messages percolate to linguistic encoding immediately as that information becomes available in the mind of the speaker. We conclude by discussing the functional role of incremental

message planning in conversational speech and the situations in which this continuous incremental planning would be most likely to be observed.

Speaking begins with the formulation of a prelinguistic message—a propositional representation of the speaker's communicative intention (Levelt, 1989). To initiate the message-to-language mapping process, the message must then be shared with linguistic encoding processes. In the present research, we examined when and how this mapping proceeds for utterances produced during unscripted, interactive dialogue. We outline two accounts concerning the temporal dependencies between message-level and sentence-level planning and discuss the implications for the scope of message planning in naturalistic production.

Two early production accounts frame much of the current theorizing about message planning. According to Paul (1880), language production is a sequential process: Speakers assemble mental constructs in an order isomorphic to word order, allowing speaking to begin as soon as any message element is available. Paul's proposal was met with debate (Blumenthal, 1970), since it contrasted with Wundt's (1900) view that speakers generate a rudimentary, holistic plan for the entire communicative intention before linguistic encoding begins (Lashley, 1951; Rosenbaum, Cohen, Jax, Weiss, & van der Wel, 2007).

This conflict between incremental and holistic conceptualization persists today, although it is now clear that incremental message preparation is feasible, as hypothesized by Paul (1880). Although speakers do sometimes engage in holistic planning (e.g., when describing simple scenes: Bock, Irwin,

S. Brown-Schmidt (✉)
Department of Psychology and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: sarahbrownschmidt@gmail.com

A. E. Konopka
Max Plank Institute for Psycholinguistics,
Nijmegen, The Netherlands

Davidson, & Levelt, 2003; Griffin & Bock, 2000; Konopka & Meyer, 2014; Kuchinsky, Bock, & Irwin, 2011), a growing body of evidence demonstrates that speakers can also plan incrementally. For example, studies examining the production of noun phrases describing multiple objects have shown that speakers may sometimes plan as little as a single object (Griffin, 2001; Meyer 1996; Smith & Wheeldon, 1999) or two objects in a sentence-initial noun phrase before speech onset (Allum & Wheeldon, 2007; Konopka, 2012; Martin, Crowther, Knight, Tamborello, & Yang, 2010; see also F. Ferreira & Swets, 2002; Gleitman, January, Nappa, & Trueswell, 2007). However, these studies typically do not pinpoint the level of representation (message or sentence level) at which this incremental preparation occurs. Since speakers can encode rudimentary conceptual information in a simple visual display within a few hundred milliseconds (e.g., Dobel et al., 2007; Hafri et al., 2013), a complete message may have been set *before* speech onset, with only *linguistic* encoding proceeding incrementally thereafter.

More direct support for incremental message planning comes from a different approach to studying planning scope—experiments in which speakers prepare messages that *change* over time and thus where *new* message-level information must be planned in a separate increment from the original message. In a series of eye-tracking studies, Brown-Schmidt and colleagues (Brown-Schmidt & Konopka, 2008; Brown-Schmidt & Tanenhaus, 2006) examined planning of scalar modified NPs (*small butterfly*) to describe a highlighted object in a complex display with multiple pictures (too many to encode in one fixation). Speakers typically reserve size adjectives for situations in which the referential context contains a contrast object that differs from the target referent in size (e.g., a large butterfly; Sedivy, 2005). These studies tested the hypothesis that the part of a speaker's message that encodes size (*small*) can be planned separately from the referent (*butterfly*), and assessed how rapidly size information can be integrated into the message.

The research offered two key findings. First, speakers were more likely to include size information in their utterances (*small butterfly* vs. *butterfly*) if they had fixated the contrast object. When this fixation occurred well before speaking, speakers produced fluent expressions (*the small butterfly*), whereas delayed fixations to the contrast resulted in disfluency (*thee uh small butterfly*; *the butterfly . . . small one*). This link between fixating the contrast and mentioning size shows that the first contrast fixation reflects the time at which speakers first added size information to the message, and, importantly, suggests that modifiers can be planned separately from the object label. Second, this form of planning is not limited to disfluent utterances. Brown-Schmidt and Konopka (2008) compared planning in phrases like *the small butterfly* and *la mariposa pequeña* in Spanish–English bilinguals and showed that speakers planned size adjectives

significantly later in Spanish than in English. This suggests that the production system supports planning of some message elements in single-word increments (*lexical incrementality*).

Although these findings demonstrate incremental preparation of short messages, they leave unanswered questions regarding how message and utterance planning are coordinated for longer messages. Here we examine the production of utterances with multiple referents and ask whether message-level and sentence-level encoding *continue* to be interleaved throughout articulation of the utterance. Specifically, we ask whether, to maintain fluency, messages must be set before speaking begins (*Initial Preparation*; see below) or whether messages can be updated after speech onset and new information seamlessly integrated into the unfolding linguistic plan without a surface repair (*Continuous Incrementality*; see below).

In principle, if people can think and speak at the same time, the preparation of new message-level information should continue while linguistic encoding operates on previously planned message increments. However, concurrent articulation and message preparation poses a coordination problem for the production system, since it requires continuously distributing resources between message-level and sentence-level processes (Bock, 1982; Levelt, 1989; see Martin et al., 2010, for a review). As compared to the already resource-demanding process of single-object naming (Roelofs & Piai, 2011), the preparation of multiword utterances further increases processing requirements. In complex cognitive systems in general, distributing resources between two tasks has consequences for both processing speed and accuracy (V. S. Ferreira & Pashler, 2002; Pashler, 1994). Speakers may deal with increasing processing requirements by, for example, temporally separating the linguistic encoding of individual referents (Griffin, 2003; Meyer, Belke, Häcker, & Mortensen, 2007; Wagner, Jescheniak, & Schriefers, 2010). A similar time-sharing problem may arise at the message level: The processing requirements of message-level and sentence-level coordination during the production of longer messages may have implications for the amount of overlap in conceptual and linguistic planning that the production system can support, and, consequently, for the degree to which lexically incremental message planning is possible while articulation is underway.

To understand the mechanisms underlying concurrent conceptual and linguistic planning, we examine the planning of new message-level information in *utterance-medial* position (e.g., *the star and the small dog*). Whereas it is clear that linguistic encoding proceeds incrementally after speech onset for preplanned messages (e.g., Griffin & Bock, 2000), we ask whether speakers can add new conceptual information (size modifiers) to the second object fluently after speech onset. We propose that whether lexically incremental preparation of messages and sentences can continue during articulation depends on the way that message-level planning is coordinated with sentence-level planning.

In two experiments, we use the paradigm developed by Brown-Schmidt, in which speakers describe objects in complex displays. The target utterances were conjoined NPs (e.g., *the star and the dog*). The target objects appeared close to one another and moved in unison ("common fate") to ensure that speakers' *initial* messages would include information about both objects. The displays also contained size-contrasting objects (a large star or large dog). Both experiments test how quickly speakers are able to update their sentence plans with a modifier of the first noun (*small star*) or the second noun (*small dog*) when the first contrast fixations occur before and after speech onset.

Here we outline two views of how message encoding can be coordinated with linguistic encoding. On the one hand, the mapping of messages onto utterances may be relatively automatic and inflexible: Once speakers generate a message-level increment (the *initial* message) and pass this information on to sentence-level processes, linguistic encoding may proceed automatically and independently, without further top-down input from the message level. This *Initial Preparation* of messages is motivated by the need to alleviate the problem of coordinating message encoding with linguistic formulation: It temporally separates message-level and sentence-level processes, and thus reduces their degree of interleaving (F. Ferreira & Swets, 2002; Martin et al., 2010). Importantly, this view predicts that the coordination of message-level and sentence-level processes will be well-timed only when speakers prepare *all* message elements prior to speaking: Fluent, uninterrupted production of a conjoined NP requires that speakers encode information about both objects before speaking.[1] Adding new message information (object size) after articulation has begun would involve costly revision of the original linguistic plan, resulting in high rates of disfluency or late mention of the new information.

On the other hand, lexically incremental message planning may involve continuous, incremental preparation of messages and sentences throughout formulation—both before and during articulation. According to this *Continuous Incrementality* view, linguistic encoding begins before conceptual preparation of the entire message is complete: Speakers continuously incorporate new message-level information into the developing sentence as soon as it becomes available, including during the linguistic encoding of earlier message elements and during articulation itself. To allow for rapid and incremental updating, *Continuous Incrementality* assumes that linguistic processes remain sensitive to changes in the content of message-level increments and permit modification of individual referents up to the point of articulation.

Figure 1a and b outline the hypothesized relationship between size-contrast fixations and utterance form for both

Initial Preparation and *Continuous Incrementality*. Both views predict that pre-speech planning of size information is necessary for fluent modification of the first noun in a conjoined NP (NP1). Pre-speech planning of size information for the second noun (NP2) also results in fluent modification. The views make different predictions for the modification of NP2 when size-contrast fixations occur after speech onset. *Initial Preparation* predicts that fixations to the size contrast after speech onset should result in NP2 disfluency, whereas *Continuous Incrementality* predicts that size contrast fixations made after speech onset but during the articulation of NP1 should still result in fluent NP2 modification.

Experiment 1

In Experiment 1, we examined the formulation of conjoined NPs in sentences like *the (small) star and the (small) dog are flashing*.

Method

Participants A group of 36 native English speakers with normal or corrected-to-normal vision participated for payment or partial course credit. Data from eight additional participants were excluded due to technical problems ($n = 3$) or failure to complete the study ($n = 5$).

Procedure Participants completed an interactive task with one of four experimenters (the experimenters in both studies were familiar with the procedure but unaware of the experimental hypotheses). The experiment was programmed in MATLAB using the Psychophysics Toolbox (Brainard, 1997). Participants' eye movements were recorded with an EyeLink 1K tracker (drift checks were performed every other trial), and productions were recorded via headset microphones.

The participant and experimenter sat in the same room at separate computers (positioned so they could not see each other's screens) and, over 360 trials, took turns instructing each other to click on objects. On each trial, either one object or two adjacent objects flashed or shifted in unison on the speaker's computer at scene onset. On shifting trials, the picture(s) shifted left-to-right ~10 pixels; on flashing trials, the picture(s) flashed on and off every ~200 ms. The speaker described this movement so that the listener could duplicate it on their own screen by clicking the target picture once (flashing) or twice (shifting; the target pictures on the listener's screen were randomly placed). The listener was given three chances to click the right object(s), and trials concluded when the listener completed the action. The language used in the experiment was conversational; there were no instructions

---

[1] We used speech onset as a reference point, since it unambiguously marks the time when the pre-speech encoding of NP1 must be completed.
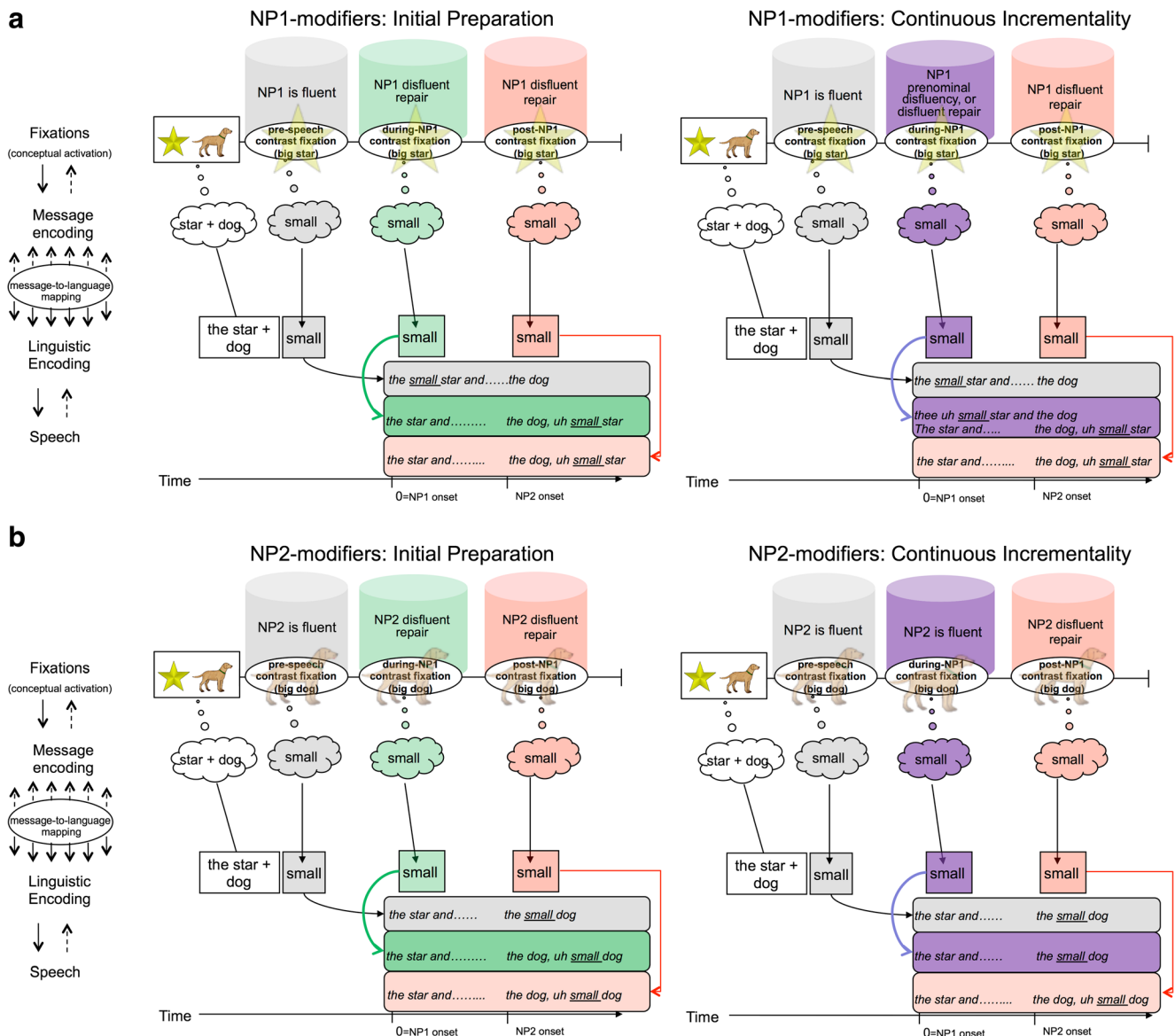
Fig. 1 (**a**) NP1-modified expressions: Schematic of the hypothesized relationship between size-contrast fixations and utterance form, under the *Initial Preparation* and *Continuous Incrementality* accounts. Time = 0 is the onset of the conjoined NP. Message formulation begins with an initial fixation to the target referents (small star and dog), followed by a fixation to the size contrast (large star). According to both accounts, fixations to the NP1 size contrast *before* the onset of NP1 result in fluent NP1 modification, whereas fixations to the NP1 size contrast *after* NP1 production result in disfluent repairs. Although the focus of the present article is on NP2 planning, the two views also make different predictions regarding NP1 modification when fixations to the size contrast occur *during* NP1: *Initial Preparation* predicts disfluent NP1 repairs, but

Continuous Incrementality predicts a combination of prenominal NP1 disfluencies and disfluent repairs. (**b**) NP2-modified expressions: According to both accounts, fixations to the NP2 size contrast *before* the onset of NP1 result in fluent NP2 modification, whereas fixations to the NP2 size contrast *after* NP1 production result in disfluent NP2 modification. The two views make different predictions regarding NP2 modification when fixations to the size contrast occur during NP1: *Initial Preparation* predicts disfluent NP2 repairs, but *Continuous Incrementality* predicts fluent NP2 modification. In short, when the message is revised after articulation begins, *Initial Preparation* predicts higher rates of disfluency than for preplanned messages, whereas *Continuous Incrementality* predicts fluent speech.

on how or when to respond, and no restrictions were placed on what either partner could say, other than that they were to say whatever was needed to get their partner to perform the action.

Materials Each trial showed a 5 × 4 grid with 20 easily nameable images from Rossion and Pourtois (2004) and similar clip-art pictures (Fig. 2). Each grid contained two pairs of pictures in size contrasts, plus 16 unique pictures. The critical

Fig. 2 Example scenes from (**a**) the listener's and (**b**) the speaker's screens. The small star and the dog (target objects) would either flash or shift on the *speaker's* screen
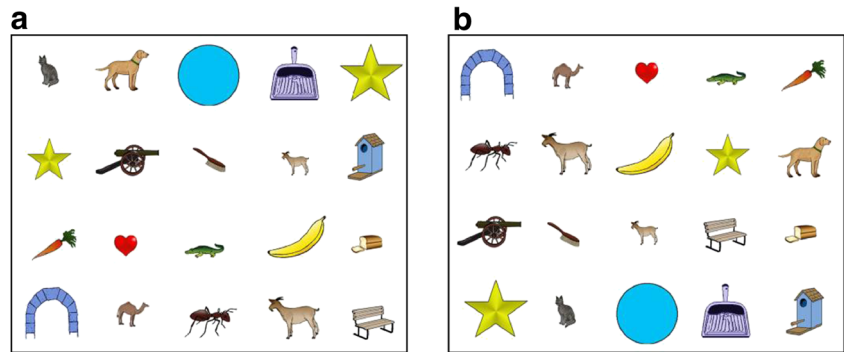
object pairs were semantically unrelated, and 58/60 pairs had different onset phonemes. The images were arranged in different locations on the participant's and the experimenter's computer screens, so as to elicit descriptive NPs (e.g., *the star*) rather than locatives (e.g., *bottom left one*).

In total, 339 unique pictures were used. Each picture appeared in one of three sizes and was repeated ~18 times during the experiment (~six times per size). Participants described 300 objects in total, 60 of which (20 %) appeared in a scalar contrast set.

Each participant played the role of *speaker* on 180 trials. Sixty of these were critical trials, on which they described two moving objects (e.g., a star and a dog), one of which was in a size contrast. On the speaker's screen, the size contrast (e.g., larger star) was at least three spaces away from the target (smaller star). Four experimental lists counterbalanced which of the two critical objects was in a size contrast and their left-to-right arrangement. On the remaining 120 filler trials, the participant described one or two objects, neither of which was in a contrast set.

The participant played the role of *listener* on 180 additional trials. These trials were similar to the target trials but were used to make the task more conversational.

Predictions According to both views (Fig. 1a–b), fluent NP1 modification (*the small star and the dog*) should occur when the size contrast is fixated before speech onset; post-speech fixations should result in disfluency. For NP2 modification (*the star and the small dog*), *Initial Preparation* predicts that contrast fixations occurring during NP1 and post-NP1 articulation should increase disfluency rates relative to pre-speech fixations. *Continuous Incrementality* predicts that only post-NP1 contrast fixations should increase disfluency rates for NP2 modifiers.

Speech onsets were analyzed as a secondary measure of planning. If messages are planned entirely before speaking, speech onsets should not differ between NP1-modified and NP2-modified utterances. If, on the other hand, message planning is continuously incremental, speech onsets should be slower when NP1 is modified, since the pre-speech message would contain more content than when NP2 is modified.

Analyses Target responses were transcribed and the word onsets coded. The timing of the *first* fixation to the size contrast relative to the onset of the conjoined NP was identified. Trials without a contrast fixation in the first 30 s were treated as if they did not have a contrast fixation (this cutoff was twice the longest speech onset time).

All data were analyzed using mixed-effects models. Unless noted, fixed effects were coded with mean-centered contrast codes. A backward-fitting procedure was used to identify models with the largest random-effects structure that converged (Barr, Levy, Scheepers, & Tily, 2013).

Typical participant utterances included the conjoined NP and verb:

1. Fluent NP1-modified: *the small airplane and the elephant are flashing*
2. Disfluent NP1-modified: *thuh . . . small arrow and the motorcycle are shifting*
3. Disfluent NP2-modified: *thuh chair and thee small peanut are flashing*
4. Disfluent NP2-repair: *thee elephant and thee plane . . . the smaller plane are flashing*

The relationship between the presence of a size contrast and scalar-modification rates are summarized in Supplement 1. Consistent with previous work, speakers used scalars when the display contained a size contrast (Sedivy, 2005) and when they had fixated the contrast (Brown-Schmidt & Tanenhaus, 2006).

Results

Eye gaze The first analyses focused on the relationship between contrast fixations and utterance fluency (Fig. 3). A logit-link mixed-effects model included NP modifier position and contrast-fixation time (pre-speech, during-NP1, post-NP1) as fixed effects. Fixation time was dummy-coded (during-NP1 fixations were the reference level; see Supplement 2 for fine-grained fixation distributions). The primary analysis treated NP fluency (fluent vs. disfluent) as a binary dependent measure (prenominal disfluency and repairs were grouped together).

Fig. 3 Proportion of fluent, disfluent, and repaired modified expressions by the timing of fixation on the size contrast (pre-speech, during-NP1, post-NP1) in (**a**) Experiment 1 and (**b**) Experiment 2. All of the bars within a fixation time window add to 1.0 (e.g., all of the pre-speech bars for NP1 add to 1.0)
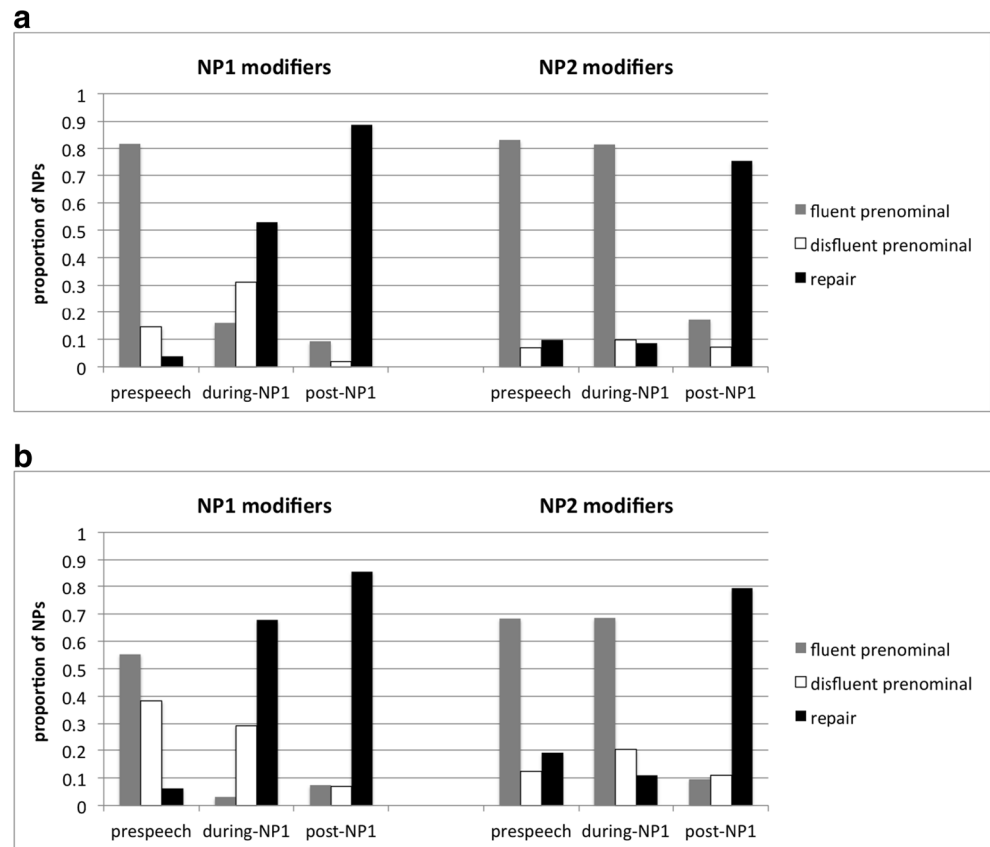
Interactions between NP position and fixation time qualified several main effects (Table 1). The pattern of disfluency rates for pre-speech versus during-NP1 contrast fixations differed by NP position ($z = 7.43$, $p < .0001$). For NP1 modifiers, disfluency was more likely when speakers first fixated the contrast during NP1 rather than before speech onset ($z = -10.60$, $p < .0001$), but during-NP1 fixations *did not* increase disfluency rates for NP2 modifiers ($z = -0.34$, $p = .73$). In other words, NP2 was fluently modified even when the NP2 contrast was fixated during NP1. When speakers fixated the contrast post-NP1, disfluency was common for both NP1 and NP2. An interaction between NP position and fixation time (during-NP1 vs. post-NP1 fixations; $z = 2.81$, $p < .01$) was due to a smaller fixation-time effect at NP1 ($z = 3.35$, $p < .001$; post-NP1 fixations resulted in primarily disfluent repairs for NP1, but fixations during NP1 resulted in a combination of prenominal disfluency and repairs) than at NP2 ($z = 13.68$, $p < .0001$).

Speech onsets The second analysis modeled conjoined-NP onsets using a Gaussian-link mixed-effects model with referential form and NP position as fixed effects (Table 2, Fig. 4). The first contrast for referential form compared fluent to disfluent expressions (prenominal disfluency and repairs);

the second contrast directly compared prenominal disfluency and repairs. An NP Position × Fluency interaction ($t = 2.54$, $p < .05$) was due to an onset delay for fluent NP1-modified expressions relative to fluent NP2-modified expressions ($t = -3.26$, $p < .01$), suggesting that onsets were influenced primarily by the complexity of NP1. There was no NP-position effect for disfluent NPs ($t = 1.56$, $p = .13$).

Discussion

The results were consistent with *Continuous Incrementality*, which proposes that messages can be prepared and updated incrementally throughout formulation. Fluency was preserved for NP1-modified expressions only when NP1-modifier planning occurred before speaking. By contrast, speakers maintained fluency when the message underlying NP2 modifiers was planned before as well as after speech onset (but before NP2 onset), showing that the initial message can be updated after articulation begins without compromising fluency. Finally, the combination of prenominal disfluency and repairs for NP1 modifiers planned during NP1 articulation suggests that in some cases, prenominal disfluency supports successful interfacing of on-the-fly message adjustments (e.g., adding a modifier) with linguistic planning (see Fig. 1a).

Table 1 Model results for the effects of contrast-fixation time and noun phrase (NP) position on fluency for both Experiment 1 (1,502 observations) and Experiment 2 (1,555 observations)

Experiment 1	Estimate	SE	z Value	p Value
(Intercept)	0.10	0.20	0.47	
NP position (NP1 vs. NP2)	−3.50	0.34	−10.25	<.0001
Contrast-fixation time (post-NP1)	2.64	0.30	8.71	<.0001
Contrast-fixation time (pre-NP1)	−1.81	0.24	−7.55	<.0001
NP Position × Fixation Time (post-NP1)	1.62	0.58	2.81	<.01
NP Position × Fixation Time (pre-NP1)	3.47	0.47	7.43	<.0001
Random effects	Variance	SD		
(Item)	5.12E–09	7.16E–05		
NP position	3.46E–09	5.88E–05		
Contrast-fixation time (post-NP1)	1.41E+00	1.19E+00		
Contrast-fixation time (pre-NP1)	3.89E–01	6.24E–01		
NP Position × Fixation Time (post-NP1)	3.37E+00	1.84E+00		
NP Position × Fixation Time (pre-NP1)	9.40E–01	9.70E–01		
(Subject)	6.23E–01	7.89E–01		
NP position	6.49E–01	8.06E–01		
Experiment 2	Estimate	SE	z Value	p Value
(Intercept)	1.9016	0.3506	5.423	
NP position	−5.368	0.6341	−8.465	<.0001
Contrast-fixation time (post-NP1)	1.1545	0.3508	3.291	<.001
Contrast-fixation time (pre-NP1)	−2.3665	0.3399	−6.963	<.0001
NP Position × Fixation Time (post-NP1)	5.2242	0.7466	6.997	<.0001
NP Position × Fixation Time (pre-NP1)	4.7333	0.6353	7.451	<.0001
Random effects	Variance	SD		
(Item)	0.736	0.858		
NP position	0.739	0.859		
Contrast-fixation time (post-NP1)	0.051	0.225		
Contrast-fixation time (pre-NP1)	0.871	0.933		
NP Position × Fixation Time (post-NP1)	3.129	1.769		
NP Position × Fixation Time (pre-NP1)	0.565	0.752		
(Subject)	0.820	0.905		
NP position	1.706	1.306		

Trials that did not have a contrast fixation or that were not modified were excluded. During-NP1 fixations are treated as the baseline condition for contrast-fixation time. The dependent measure is the fluency of the NP: fluent or disfluent (prenominal disfluency and repairs are categorized together as "disfluent")

Experiment 2

The aim of Experiment 2 was to replicate these findings. Given the low NP2 disfluency rates in Experiment 1, in Experiment 2 we employed a priming manipulation to increase disfluency rates.

Method

Participants A group of 36 speakers participated for payment or partial course credit. The data from 13 additional participants were excluded due to technical problems with the eyetracker and audio equipment ($n = 12$) and experimenter error ($n = 1$).

Procedure and materials The materials and procedure were identical to those of Experiment 1, with one exception. Participants were randomly assigned to one of three between-subjects priming conditions in which the experimenter intentionally produced a prenominal disfluency or a disfluent repair on 45/60 of contrast-present trials (e.g., *thee uh small star*; *the star, uh small one*). In the control condition and in all other trials, the experimenter produced expressions naturally, as in Experiment 1. The priming manipulation was only partially successful (weak priming was observed only for postnominal repairs) and will not be discussed further. Priming did not affect the time course of planning.

Table 2 Analysis of the onset of the conjoined noun phrase (NP), for modified NPs in Experiments 1–2

Experiment 1	Estimate	SE	t Value	p Value
(Intercept)	2,676.32	51.33	52.14	
NP position	–43.54	25.99	–1.68	.10
Fluency (fluent vs. disfluent utterances)	–22.28	57.26	–0.39	.71
Disfluency type (prenominal disfluency vs. repairs)	–174.25	106.58	–1.63	.11
NP Position × Fluency	136.66	53.86	2.54	<.05
NP Position × Disfluency Type	56.25	89.16	0.63	.56
Random effects	Variance	SD		
(Item)	10,915	104		
NP position	2,109	46		
Fluency	14,925	122		
Disfluency type	28,920	170		
(Subject)	82,117	287		
NP position	4,423	67		
Fluency	63,613	252		
Disfluency type	248,182	498		
Residual	177,800	422		
Experiment 2	Estimate	SE	t Value	p Value
(Intercept)	2,630.471	59.2832	44.37	
NP position	0.1455	31.2737	0	.829
Fluency (fluent vs. disfluent utterances)	–53.4339	25.9758	–2.06	<.05
Disfluency type (prenominal disfluency vs. repairs)	–28.4294	38.1277	–0.75	.38
NP Position × Fluency	35.0093	51.0014	0.69	.8
NP Position × Disfluency Type	201.1133	66.1787	3.04	<.01
Random effects	Variance	SD		
(Item)	17,680	133		
NP position	16,361	128		
Fluency	6,110	78		
Disfluency type	360	19		
NP Position × Fluency	18,246	135		
NP Position × Disfluency Type	24,083	155		
(Subject)	111,742	334		
NP position	9,815	99		
Fluency	973	31		
Disfluency type	20,809	144		
NP Position × Fluency	8,729	93		
NP Position × Disfluency Type	24,100	155		
Residual	155,810	395		

The fixed effects are NP position (NP1-modified vs. NP2-modified) and utterance form, coded with two contrasts, fluency (fluent vs. disfluent [both prenominal disfluency and repairs]) and disfluency type (prenominal disfluency vs. postnominal repairs). The data analyzed include scalar-modified NPs, regardless of whether the contrast was fixated (Exp. 1, *n* = 1,564; Exp. 2, *n* = 1,704)
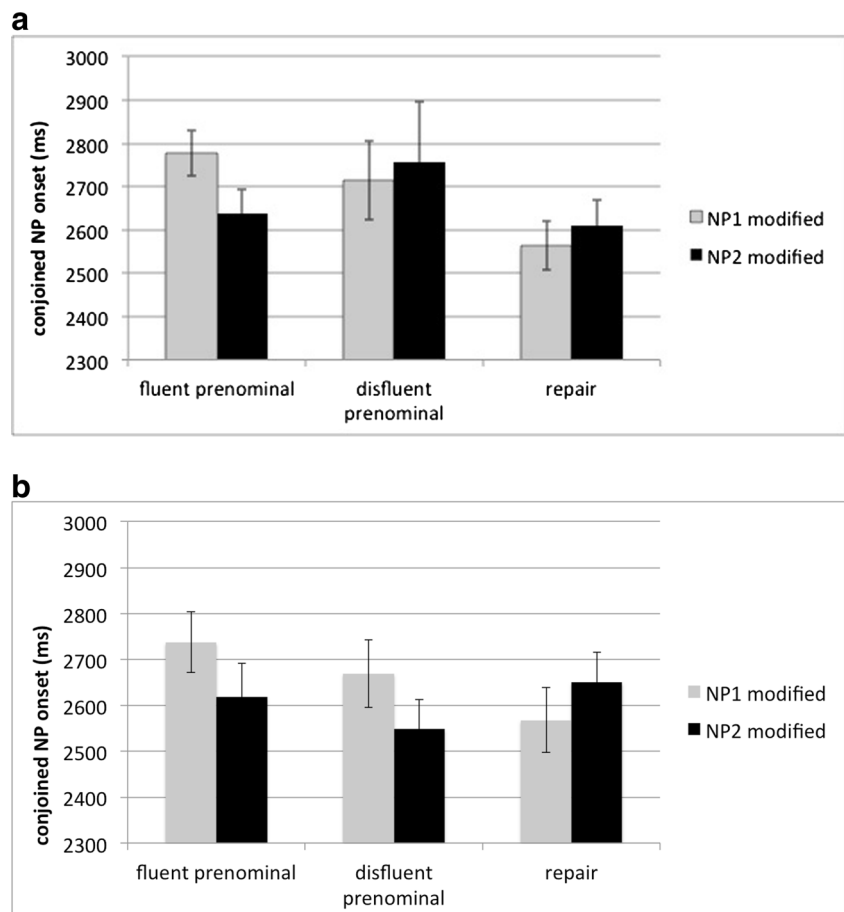
Results

Analyses were conducted as in Experiment 1. As before, NP modification was frequent when a size contrast was present in the scene (Supplement 1).

The primary analysis examined the relationship between the timing of contrast fixations and referential form (Table 1, Fig. 3). Interactions between NP position and contrast-fixation time again showed that planning was different for NP1 and NP2 modifiers. For NP1 modifiers, during-NP1 fixations resulted in increased disfluency relative to pre-speech fixations; for NP2 modifiers, during-NP1 fixations did not increase disfluency. Post-NP1 fixations resulted in increased disfluency rates at both NP positions.

The analysis of speech onsets revealed an interaction of disfluency type with NP position (Table 2, Fig. 4), due to onset

Fig. 4 Onsets of the conjoined NPs, relative to display onset, separately for modified NP1 and NP2, in (**a**) Experiment 1 and (**b**) Experiment 2. Fluent modified expressions (e.g., *the small dog*), disfluent prenominally modified expressions (e.g., *thee uh small dog*), and postnominal size repairs (e.g., *the dog, uh small one*) are shown separately

General discussion

Previous findings established that short messages can be planned and updated incrementally in lexically sized units (Brown-Schmidt & Konopka, 2008). Here, we contrasted two possible mechanisms for the planning of longer messages—*Initial Preparation* and *Continuous Incrementality*—that make different predictions about the coordination of message formulation with linguistic encoding. According to *Continuous Incrementality*, messages are continuously updated and new information integrated into the utterance plan, even after speech onset; according to *Initial Preparation*, the message-to-utterance cross-talk occurs only before articulation, so messages must be set before articulation to maintain fluency. The results of both experiments were consistent with *Continuous Incrementality*: New message elements (size modifiers) were incorporated into developing utterances

delays for disfluent, prenominally modified expressions relative to postnominal repairs for NP1; the effect was not significant at NP2.

without compromising fluency both *before* and *during* speaking. These findings allow for several conclusions.

First, messages consisting of multiple elements can be updated in a lexically incremental fashion during articulation. The ease with which speakers added size information to the message shows that the production system allows message revisions to be performed on the fly throughout formulation. The fact that the added message element corresponded to a single word (e.g., *small*) shows that the formulation process can be *lexically* incremental.

Second, the ease with which new message-level information was mapped onto language implies tight coordination between message-level and sentence-level processes. Message updates were rapidly added to the linguistic plan, indicating that linguistic encoding is not "sealed off" from the conceptual level. Instead, message implementation involves a dynamic mapping of concepts onto words: Speakers remain sensitive to small changes in the content of a message and can incorporate these changes into the sentence plan on a fine temporal scale—affording fluent articulation of late-planned message elements.

We propose that *Continuous Incrementality* may be particularly common in conversation, where feedback and interactivity demand constant message updating. When constructing messages in unprepared speech, speakers may continuously self-generate message updates prompted by changes in their communicative intent or by listener feedback; *Continuous Incrementality* allows these message updates to be fluidly integrated into ongoing linguistic encoding.[2] From a processing perspective, continuously incremental planning of small units may facilitate the coordination of message-level and sentence-level planning by reducing the need to buffer new message-level information in working memory until articulation is complete, and instead allowing speakers to make revisions to individual message units as necessary. Although production frequently involves a trade-off between planning a series of small increments and planning of larger increments (see F. Ferreira & Swets, 2002), the present research suggests that *Continuous Incrementality* may be normative in conversational settings.

An open question is whether the mechanisms of *Continuous Incrementality* extend to all types of message elements. One possibility is that key structural elements such as the verb and its arguments are planned in advance (Lee, Brown-Schmidt, & Watson, 2013) and are not easily changed or updated without interruptions or repairs. Thus, whereas *Continuous Incrementality* may be normative for modifiers and other adjuncts, *Initial Preparation* may be normative for verbs and arguments. The mechanisms by which advance and incremental planning are interwoven remain an important topic for future research.

A final point concerns methodology. These experiments examined planning during unscripted, conversational speech. In particular, the critical bit of message-level information that this research examined—referent size—was information that speakers decided to mention on their own. This methodological feature implies that we were able to track the genesis of a message (albeit a small one) all the way through to articulation. This is precisely the type of ecological validity that is necessary to inject into experimental settings to be able to ask, and answer, questions about message-level processes. The fact that this evidence came from unscripted conversation in which utterance form was varied makes it more representative of natural language production, and thus more likely to generalize to language use outside the lab.

---

[2] An alternative characterization of our findings is that speakers engaged in covert error correction when the initial linguistic plan no longer matched the updated message. According to this view, speakers must re-design the linguistic plan for NP1 or NP2. Inconsistent with this account, however, is the fact that we found no indications of linguistic re-planning or repairs in the surface form of late-planned message updates.

References

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 791–810. doi:10.1037/0278-7393.33.4.791

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. doi:10.1016/j.jml.2012.11.001

Blumenthal, A. L. (1970). *Language and psychology: Historical aspects of psycholinguistics.* New York, NY: Wiley.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89,* 1–47.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48,* 653–685. doi:10.1016/S0749-596X(03)00007-X

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition, 109,* 274–280.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54,* 592–609.

Dobel, C., Gumnior, H., Bolte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica, 125,* 129–143.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1187–1199. doi:10.1037/0278-7393.28.6.1187

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46,* 57–84.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language, 57,* 544–569.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82,* B1–B14.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10,* 603–609.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: recognition of two-participant actions from brief displays. *Journal of Experimental Psychology General, 142,* 880–905. doi:10.1037/a0030045

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66,* 143–162.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73,* 1–40. doi:10.1016/j.cogpsych.2014.04.001

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 748–756.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York, NY: Wiley.

Lee, E.-K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition, 129,* 544–562. doi:10.1016/j.cognition.2013.08.007

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

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition, 116,* 177–192.

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.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language, 57,* 210–231. doi:10.1016/j.jml.2006.10.005

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244. doi:10.1037/0033-2909.116.2.220

Paul, H. (1880). *Prinzipien der Sprachgeschichte [Principles of the history of language]*. Leipzig, Germany: Niemeyer.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2*(307), 1–14. doi:10.3389/fpsyg.2011.00307

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science, 26,* 525–554.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33,* 217–236. doi:10.1068/p5117

Sedivy, J. C. (2005). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 345–364). Cambridge, MA: MIT Press.

Smith, M., & Wheeldon, L. (1999). High-level processing scope in spoken sentence production. *Cognition, 73,* 205–246.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 423–440.

Wundt, W. (1900). *Völkerpsychologie: Vols. 1–2. Die Sprache [Language]*. Leipzig, Germany: Kröner-Engelmann.