

Grammatical planning scope in sentence production: Further evidence for the functional phrase hypothesis

LI-MING ZHAO

Chinese Academy of Sciences

F.-XAVIER ALARIO

Aix-Marseille Université and CNRS

YU-FANG YANG

Chinese Academy of Sciences

Received: November 6, 2012 Accepted for publication: October 16, 2013

ADDRESS FOR CORRESPONDENCE

Yu-Fang Yang, Institute of Psychology, Chinese Academy of Sciences, 4A Datun Road, Chaoyang District, Beijing 100101, China. E-mail: yangyf@psych.ac.cn

ABSTRACT

While there is a consensus that speakers plan their utterances before they start producing them, the scope of the initial planning unit remains controversial. In subject-initial utterances, is the planning unit the whole subject phrase or a smaller “functional phrase” within the subject phrase? Allum and Wheeldon (2007) reported that speakers show faster onset latencies in producing utterances like *The flower above the house is red*, where the subject consists of two functional phrases (*the flower and above the house*) than in producing *The flower and the house are red*, where there is a single, longer functional phrase (*The flower and the house*), both in head-initial languages like English and head-final languages like Japanese. This has been taken to suggest that the functional phrase is a preferred unit of planning, rather than the whole subject. Experiment 1 in the present study replicates Allum and Wheeldon’s study with speakers of another head-final language (Mandarin Chinese) and finds similar results. Experiments 2 and 3 investigate whether syntactic processing or visual grouping could potentially explain the pattern of responses, and find that they cannot. Together, these results provide further empirical support for the claim that the functional phrase is a primary unit of grammatical planning for speech production.

Most psycholinguistic models of language production agree in distinguishing three major processing levels: conceptualizing, formulating, and articulating (e.g., Garrett, 1975, pp. 505–529; Levelt, 1989). In a picture-naming task requiring participants to respond to a pictured object with a single noun, all processing

required for the planning of the entire utterance is complete before the articulation is initiated (e.g., Griffin, 2001; Meyer, Sleiderink, & Levelt, 1998; Meyer & van der Meulen, 2000). This may not be the case for longer utterances. Garrett argued that the planning of a whole sentence may not need to be completed at a given processing stage before that stage can release its output to a subsequent stage (cf. Garrett, 1976, pp. 234–236). Kempen and Hoenkamp (1987) developed an incremental procedural grammar hypothesis in which two stages of planning can be processed concurrently and multiple fragments can be processed simultaneously within a single stage. In considering this incremental hypothesis of articulatory processing, a central issue is to clarify how far ahead speakers plan before they start producing an utterance, a parameter referred to as the *scope of planning*. This can be assessed for each of the processing levels involved.

Garrett (1975) and Levelt (1989) hypothesized that the formulation process consists of two processing steps: grammatical encoding and phonological encoding. Grammatical encoding concerns semantic and syntactic information, while phonological encoding concerns information about the phonological form of words. The planning involved at these two levels has been investigated in some detail. There is strong experimental evidence that speakers typically use a larger scope of planning at the stages of conceptualization and syntactic/grammatical encoding, and smaller units at the stage where the phonological form of the utterance is determined (e.g., Costa, Navarrete, & Alario, 2006; Meyer, 1996; Smith & Wheeldon, 2004; Yang & Yang, 2008; for speech error evidence, see Garrett, 1975, pp. 133–177). At the grammatical level, the scope of planning is often considered to be constrained by certain grammatical units, such as a clause (a sentence fragment consisting of at least a subject and a predicate) or a subject noun phrase (the phrase including all nouns associated with the grammatical role of subject), but this remains a controversial issue. The goal of the current study is to address this controversy and to clarify the scope of planning at the grammatical level during sentence production.

In many studies, the planning scope in sentence production has been investigated by asking participants to verbally describe the relationship between two or more distinct objects and comparing onset latencies across utterance formats. Levelt and Maassen (1981) showed that utterances in which the sentence subject was a conjoined noun phrase (CNP; e.g., “*The circle and the square* move up,” the subject noun phrase is in *italics*) had longer onset latencies than those involving conjoined sentences (e.g., “*The circle* moves up and the square moves up”). They argue this onset delay is due to the necessity of retrieving the lemmas (in this context, a *lemma* is an abstract lexical entity specifying syntactic but not phonological properties of a word; see Bock & Levelt, 1994; Levelt, 1989; Roelofs, 1992; but see Caramazza, 1997) of both nouns in a CNP before utterance onset, whereas for conjoined sentences only the lemma of the first noun is retrieved. Lemma retrieval is considered to occur at the grammatical level, and the results of Levelt and Maassen (1981) led to the conclusion that the scope of grammatical planning might be the subject noun phrase. However, in that study the length of the first clause was not controlled (“The circle and the square move up” vs. “The circle moves up”), so the difference in onset latency may result from the difference in the lengths of clauses rather than their manipulation of the subject phrase.

Smith and Wheeldon (1999) used a paradigm similar to that of Levelt and Maassen (1981) but controlled the clause length while varying the subject phrase length. They showed that onset latencies for sentences such as “*The dog and the foot move above the kite*” were longer than that for sentences such as “*The dog moves above the foot and the kite.*” This indicates that the scope of grammatical planning prior to articulation does not encompass the whole sentence. Rather, this and related results led Smith and Wheeldon to reiterate that the scope of grammatical planning was the subject noun phrase. Here the nonphonological nature of planning has been assumed because of some evidence to suggest that the phonological planning unit is much smaller, limited to the first phonological word (e.g., Griffin, 2001; Schriefers & Teruel, 1999; but see Alario, Costa, & Caramazza, 2002; Schnur, Costa, & Caramazza, 2006).

Martin, Miller, and Vu’s (2004) findings also support the hypothesis that the subject phrase is the initial unit of planning. This study tested one patient with a semantic retention deficit in short-term memory and one patient with a phonological retention deficit with the materials and manipulations used by Smith and Wheeldon (1999). Martin et al. reasoned that if the phrasal planning was carried out at either the lemma or the phonological level, then a patient with a short-term memory deficit at that level should have difficulty initiating the sentences beginning with a complex noun phrase. The patient with the semantic retention deficit showed a greatly exaggerated effect of initial noun phrase complexity, while the patient with the phonological retention deficit showed an effect within the normal range. As suggested previously, these results were interpreted as evidence for phrasal planning at the grammatical level rather than the phonological level.

Language specific properties, however, raise two uncertainties. First, in these studies, the subject noun phrase only includes head nouns. In the sentence “*The dog and the foot move above the kite,*” “dog” and “foot” are both head nouns of the subject noun phrase “the dog and the foot.” An alternative structure occurs in sentences such as “*The dog above the flower is red.*” There “dog” is the head noun, while “above the flower” is a complement. Accordingly, Smith and Wheeldon’s (1999) and Martin et al.’s (2004) results do not clarify whether the grammatical planning scope may encompass the whole subject noun phrase or simply the head of the subject noun phrase. To address this first problem, Allum and Wheeldon (2007) compared onset latencies between sentences with a prepositional phrase (PP) modified subject (e.g., “The dog above the flower is red”; “PP utterances”) and sentences with a CNP as the subject (e.g., “The dog and the flower are red”; “CNP utterances”). They observed slower onset latencies for CNP utterances than for PP utterances with English speakers and materials. This result indicates that, in contrast to the conclusions stated earlier, the scope of grammatical planning is not the whole subject phrase but rather possibly the phrase consisting of the head nouns, but excluding its complement.

Allum and Wheeldon (2007) further noted that the previous research had been conducted in head-initial languages (e.g., English) where the head of the subject noun phrase is always the initial noun phrase. As a result, it remained to be clarified whether it is truly the head of the subject noun phrase that defines the grammatical planning scope, or just the initial noun or noun phrase. To address this uncertainty, they focused on the head-final characteristics of Japanese. For

example, in the Japanese sentence “The dog above the flower is red,” the noun of the modifier phrase “flower” is produced as the initial noun, before the head noun “dog” (the literal translation is “flower above dog red is”). Again, Allum and Wheeldon (2007) observed slower onset latencies for the CNP utterances than for the PP utterances. On the basis of this evidence, they suggested that the scope of grammatical planning is not defined in terms of the initial head noun or the subject noun phrase, but rather in terms of the first *functional phrase*. They defined a functional phrase as “one that likely represents a unit in the thematic representation of the utterance but is not necessarily one of the arguments of the verb or the head of a verb argument phrase” (p. 792). In this view, the subject noun phrase of a CNP utterance (e.g., “the dog and the flower”) serves a single function representing the agent. In contrast, the subject noun phrase of a PP utterance (e.g., “the dog above the flower”) consists of two smaller functional phrases, “the dog” as the agent and “above the flower” as the modifier. The faster onset latencies observed in Japanese for the PP compared to the CNP utterances are attributed to the first functional phrase being shorter in the former case. This result was further supported by the finding that lengthening the initial functional phrase resulted in an increase of onset latencies (Experiments 2 and 3 in Allum & Wheeldon, 2007). Yet more evidence comes from a task in which the lower picture (which refers to the second noun) was previewed before the presentation of the picture pair. There was a previewing facilitation effect for the CNP utterances only, suggesting that the second noun was planned in the CNPs (*the dog and the flower . . .*) but not in the PPs (*the dog above the flower . . .*; Allum & Wheeldon, 2009).

This line of research allows us to refine our understanding of grammatical planning. The functional phrase hypothesis provides a novel definition of the grammatical planning unit and calls for a modification of the phrasal scope account. Before discussing this interpretation any further, however, an important distinction should be made between *mandatory* and *preferred* planning units. While a certain planning unit may be *preferred* in certain contexts, this does not mean that it constitutes a fixed unit of encoding that is mandatorily used in any speaking situation (Konopka, 2012).

As an illustration of this distinction, note that all the experiments reviewed above have made use of largely equivalent experimental paradigms, involving the production of a few different sentence structures in response to visual displays of objects. When other experimental paradigms have been used, different boundaries have been placed on the scope of grammatical planning: the first noun in paradigms recording eye movements (e.g., Griffin, 2001; Meyer et al., 1998; Meyer & van der Meulen, 2000), or the first clause in picture–word interference paradigms (e.g., Meyer, 1996). In addition, certain studies have shown that a number of experimental factors modulate planning patterns. Examples of this modulation include time pressure in the form of a response deadline (Ferreira & Swets, 2002), cognitive load in the form of an additional conceptual task and variable utterance formats (Wagner, Jescheniak, & Schriefers, 2010), or the relative availability of the words and structures (e.g., via priming; Konopka, 2012; Wheeldon, Ohlson, Ashby, & Gator, 2013).

From this evidence, it is clear that the scope of grammatical planning may vary according to specific experimental manipulations but that under frequently

employed conditions the functional phrase is the preferred planning unit. We set out to test the reliability of this conclusion by examining how speech is elicited in three picture-naming experiments involving different sentence structures.

THE CURRENT STUDY

In the experiments reported below, we tested whether three specific features of sentence production experiments account for the observed planning patterns described above (most notably, in Allum & Wheeldon, 2007).

First, direct evidence for the functional phrase hypothesis only comes from experiments conducted in one language, Japanese, by one research group (Allum & Wheeldon, 2007, 2009). Therefore, we first replicated Allum and Wheeldon's (2007) experiment in another head-final language, namely, Mandarin Chinese (Experiment 1). This experiment provides a test of the generalizability of Allum and Wheeldon's findings, and provides the foundation for the two subsequent experiments.

Second, Allum and Wheeldon (2007) considered the possibility that the latency difference between slower CNP and faster PP utterances could be due to a difference in the syntactic complexity. While participants may use the same lexical scope in the CNP and PP sentences in the experiment, they use different sentence structures, which may require different amounts of time to retrieve and/or process. Such disparity of timing could arise from differences in syntactic complexity, relative frequency of use, and so forth. For this reason, the effects attributed to functional phrase encoding (in particular, different scopes of lexical retrieval) might be driven by the retrieval of the sentence structure, which refers to the whole sentence syntax or the syntactic planning scope (how much syntactic information speakers decide to generate prior to speech onset). It is important that Allum and Wheeldon (2007) noted that, if anything, PP utterances might be more complex than CNP utterances, and hence should be produced after longer latencies. Even though the PP structure was more difficult (or as difficult) to process than the CNP structure in the experimental circumstances, the difference in syntactic difficulty may drive a difference in lexical scope, thus causing the latency effect observed in their experiments. Experiment 2 was designed to test empirically these two processing accounts against one another. We resorted to the preview procedure introduced by Smith and Wheeldon (2001). Lemma access was factored out by providing participants with advance information about which words they had to use but not which structure. Any remaining difference between CNP and PP utterances should be driven by syntactic processing.

Third, the evidence described above suggesting the phrase/functional phrase as the preferred unit of speech planning relies on visual displays to trigger sentence production. In particular, the different sentence structures are triggered by different visual cues in the displays (e.g., two items that move together, or are of the same color, are to be produced in a conjoined phrase). One possibility is that such features of the visual displays may be the source of the onset latency differences between sentence types, the so-called visual grouping hypothesis (e.g., Allum & Wheeldon, 2007; Martin, Crowther, Knight, Tamborello, & Yang, 2010; Smith & Wheeldon, 1999). According to this account, the perceptual interference between pictures is

increased when they have a feature in common (e.g., movement or color), which is typically the case in displays prompting CNP but not PP utterances. This extra visual interference in turn may slow the retrieval of the name of the first picture. In Experiment 3, we tested the visual grouping hypothesis in the case of color cues.

EXPERIMENT 1

This experiment was designed to examine the difference in onset latencies between Chinese CNP utterances compared to PP utterances. Native Mandarin Chinese speakers were asked to name two pictures presented vertically using sentences with a CNP as the subject (CNP utterances; such as “*N1* 和 *N2* 都是红色的,” the translation equivalent of “The *N1* and the *N2* are both red”) as well as sentences with a PP modified subject (PP utterances; such as “*N1* 下面的 *N2* 是红色的,” the translation equivalent of “The *N2* under the *N1* is red”). These structures are similar to those used by Allum and Wheeldon (2007) in their experiments in Japanese. In both utterance formats, the top picture referred to the first noun and the lower picture referred to the second noun to be produced in the utterance. Mandarin Chinese is a head-final language, as is Japanese. In the PP utterances, the initial phrase is also a modifier phrase but not the major element in the clause. If the functional phrase hypothesis is reliable across languages, we should obtain longer naming latencies for the CNP utterances than for the PP utterances, as Allum and Wheeldon (2007) observed.

Method

Participants. Twenty-four undergraduate and graduate students in Beijing participated in the experiment. They were all native Chinese speakers with normal or corrected to normal vision, and they were paid for their participation.

Materials. Forty-two pictures were used in the experiment (Snodgrass & Vanderwart, 1980; Zhang & Yang, 2003). Thirty-two of them were used as experimental pictures, and the remaining 10 pictures were used as fillers. The experimental pictures were divided into two groups, matched for frequency and naming latency (Zhang & Yang, 2003). A picture pair in a trial was composed of one picture from each group. The two picture names in one pair were not phonologically related and had no obvious semantic relation. All the pictures used in the present experiment had two-character names in Mandarin Chinese (see Appendix A). Because one character in Mandarin Chinese is pronounced in one syllable (except for some special cases, such as “儿” of “花儿”), the phonological length was always the same for each item in the present experiment. Each picture appeared four times (top and lower positions crossed with PP or CNP sentence type).

Participants were asked to produce utterances based on the color of the pictures. To trigger CNP utterances, the two pictures were presented in red. Participants were asked to use noun phrases such as “*N1* 和 *N2* 都是红色的” (“The *N1* and the *N2* are both red”). To trigger PP utterances, the lower picture was colored in red and the top picture was white. Participants were asked to use a prepositional phrase such as “*N1* 下面的 *N2* 是红色的” (the translation equivalent of “The

N2 under the *N1* is red”). Note that, unlike in English, in the original Mandarin version *N1* is produced first. The *N1* referred to the object of the top picture, and the *N2* referred to the lower one. The second noun (*N2*) is the head of the subject phrase and is produced after the modifier.

Ten pictures were used in filler trials. There were two sentence structures in filler trials, the same as in Allum and Wheeldon (2007). For one sentence structure, participants were asked to produce sentences of the following form: “两个 *N* 都是灰色的” (“The two *Ns* are gray”), where *N* refers to the object of the picture, in response to two identical gray pictures. The other filler stimulus consisted of two blank gray squares of the same size as the picture squares. Each pair was also presented vertically, and participants were asked to produce the sentence “没有图片” (“There are no pictures”).

Design. The design was very similar to that of Allum and Wheeldon’s (2007) Experiment 2. The sentence type, CNP versus PP, was the only independent variable. In each block, there was one balanced set of 16 experimental pairs, 10 pairs of gray pictures as fillers, and 10 blank square items, totaling 36 trials. Each experimental picture appeared once in each block. The items within each block were presented in a pseudorandom order to ensure that the experimental items did not appear in the first two trials in any block and that trials involving identical sentence types did not appear consecutively. There were four blocks, and the order of blocks was rotated across participants.

As in Allum and Wheeldon’s (2007) Experiment 2, there were two practice blocks. Experimental items were recombined and presented in the experimental conditions (i.e., CNP or PP), equally divided between utterance types. They were also presented once in gray, as fillers. The filler pictures were combined four times to make 20 pairs, half of which appeared in each experimental condition. There were also 10 blank fillers, making a total of 78 practice trials, divided into two blocks of 39 trials. Before the practice session, there was a familiarization session, for participants to get familiar with the pictures and their names used in the practice and the experiment proper.

Procedure. Participants were tested individually and were seated in front of the computer screen, at a distance of about 70 cm. A fixation point appeared on the screen for 1000 ms. Then the pair of pictures appeared for 4000 ms. Participants were asked to name pictures with the required syntactic structures as accurately and quickly as possible. There was a blank interval of 2000 ms between trials. The screen’s background remained black. The whole experiment (including familiarization and practice) took about 45 min, and participants could have a break between sessions and blocks.

Results and discussion

Three types of responses were scored as production errors and excluded from the analyses of onset latencies: using unexpected content words, including picture names and adjectives (color); using incorrect syntax; and fluency problems (repairing, stuttering, hesitation, and production of nonverbal sounds that triggered the

Table 1. Mean latencies and percentage error rates for three experiments (participant means)

Experiment	Sentence Types	Latency (ms)		Error Rate (%)
		Mean	SD	
1	CNP	1221	191	10.9
	PP	1145	197	7.9
	Difference	76		3.0
2	CNP	758	160	7.2
	PP	743	162	4.6
	Difference	15		2.6
3	CNP	1158	181	4.8
	PP	1129	190	6.1
	Difference	29		-1.3

Note: CNP, conjoined noun phrase; PP, prepositional phrase.

voice key). Outliers were defined as latencies of less than 300 ms, more than 3000 ms, or exceeding more than three standard deviations from a participant's average. Such trials were excluded from the latency analyses. The excluded trials and recording failures amounted to 10.7% of the data. Mean latencies and percentage of production error rates are shown in Table 1.

Two separate analyses were conducted with participants and items as random variables. Paired two-tailed student *t* tests (t_1) were used in the participant analysis, and nonpaired two-tailed *t* tests (t_2) in the item analysis. There were significant differences in the onset latencies between CNP and PP utterances, $t_1(23) = 5.31$, $p < .001$; $t_2(62) = 2.74$, $p < .01$, showing longer latencies for CNP than for PP utterances. The error rates also showed a difference that was significant within participants, $t_1(23) = 2.10$, $p < .05$; $t_2(62) = 1.42$, $p = .16$, with more errors in CNP than in PP utterances.

This experiment replicated, in a different language, the significant difference in onset latencies between CNP and PP utterances reported by Allum and Wheeldon (2007). The results for the error rates were consistent with those for the onset latencies. These results indicate that the difference in production latencies between CNP and PP utterances is a robust phenomenon, reliable across languages. As discussed above, this result is consistent with the functional phrase hypothesis. In the following experiments, we test whether alternative accounts for this result can be rejected.

EXPERIMENT 2

An alternative account for the latency differences between CNP and PP utterances, as mentioned above, is syntactic processing. Participants may be using the same lexical scope for the two sentence types (e.g., encoding full subjects with two nouns in both cases), while the difference in onset latencies between those stems

from differences in the grammatical complexity of the sentences or in the syntactic scope. We explored this possibility by modifying the naming experiment to include a preview period (Allum & Wheeldon, 2009; Smith & Wheeldon, 2001).

In this experiment, the lexical items to be combined in either a CNP or a PP utterance were known in advance because the pictures to be named were seen before the cue indicating the sentence structure to be produced was displayed (both the lexical items and the sentence structure still varied randomly from trial to trial; see Methods for details). Smith and Wheeldon (2001) argued that exposing participants to the pictures prior to the trial would factor out the process of lemma access from the latencies by ensuring that it occurred prior to the response cue. Any remaining effects should reflect syntactic planning, while differences in onset latency between conditions with and without preview should reflect the time needed for lemma retrieval processes.

In this preview naming task, we divided the 4000 ms presentation time of the picture pair into two parts. First, participants viewed the picture pair depicted in white for 2000 ms and were asked to prepare the utterance as much as possible. According to the results of Experiment 1, this time window of 2000 ms is sufficient for most participants to retrieve the names of each of the two pictures. After this 2-s preview, either the lower picture or both pictures turned red. Participants were asked to name the pictures with the required syntactic structure based on the same color display rules as in Experiment 1. Response onset latency was calculated from the moment of the color change to the onset of articulation.

According to the functional phrase hypothesis, the difference in onset latencies between the present experiment and Experiment 1 should be greater for CNP utterances than for PP utterances. The functional phrase hypothesis assumes that both lemmas are processed prior to speech onset in the case of CNP while only the first lemma is processed in the case of PP. If syntactic planning processes make a significant contribution to the latency effect observed in Experiment 1 (i.e., CNP structure requires larger scope of syntactic planning or is more difficult to retrieve than the PP structure), a significant difference in onset latencies for CNP and PP utterances should remain in this preview task.

Method

Participants. Twenty-four undergraduate and graduate students in Beijing participated in the experiment. None had participated in Experiment 1. They were all native Chinese speakers, with normal or corrected to normal vision.

Results and discussion

Using the same exclusion criteria as in Experiment 1, 7.9% of the data were removed from the latency analysis. Mean latencies and percentage error rates are also shown in Table 1.

The same analyses as in Experiment 1 were conducted here. There was no significant difference in onset latencies between CNP and PP utterances, $t_1(23) = 1.49$, $p = .15$; $t_2(62) = 1.59$, $p = .12$. The difference in error rates was significant

in the analysis by participants, $t_1(23) = 2.35, p < .05$; $t_2(62) = 1.69, p = .095$, showing more errors in CNP than in PP utterances.

Comparison of Experiments 1 and 2

Two separate analyses were conducted with participants (F_1) and items (F_2) as random factors, including the factors of sentence type and experiment.

The results showed that the main effect of experiment was significant for the onset latencies, $F_1(1, 46) = 72.54$, mean squared error (MSE) = 61,757, $p < .001$; $F_2(1, 62) = 831.93$, $MSE = 7,339$, $p < .001$, and for the error rates, $F_1(1, 46) = 7.01$, $MSE = 0.4$, $p < .05$; $F_2(1, 62) = 12.23$, $MSE = 0.3$, $p < .01$. The main effect of sentence type was significant for the onset latencies, $F_1(1, 46) = 26.74$, $MSE = 1,846$, $p < .001$; $F_2(1, 62) = 9.97$, $MSE = 7,032$, $p < .01$. For the error rates, the same contrast was significant by participants but not by items, $F_1(1, 46) = 9.62$, $MSE = 0.2$, $p < .01$; $F_2(1, 62) = 3.32$, $MSE = 0.8$, $p = .073$. The interaction between sentence type and experiment was also significant for the onset latencies, $F_1(1, 46) = 11.64$, $MSE = 1,846$, $p < .01$; $F_2(1, 62) = 3.94$, $MSE = 7,339$, $p = .05$, while the interaction for the error rates was not significant ($F_s < 1$).

Discussion

Responses in Experiment 2 were significantly faster and less error prone than in Experiment 1. This suggests that the picture names were retrieved during the previewing period. Moreover, the difference in onset latencies between CNP and PP utterances was very much reduced compared to that in Experiment 1, and response latencies for the two conditions were not significantly different. Thus, the difficulty in syntactic processes for CNP and PP utterances is equal in such circumstances. To further clarify, even though the difference in syntactic difficulty may drive a difference in lexical scope (Konopka, 2012; Wagner et al., 2010), this is not the case in Experiment 1. In one word, this is *evidence* against the hypothesis that syntactic processes are the main origin of the latency effect observed in Experiment 1. The preview effect in the onset latencies (the difference between Experiment 2 and Experiment 1) was larger for CNP than for PP utterances (463 vs. 402 ms). This is in keeping with Allum and Wheeldon's (2009) finding that previewing the second noun facilitated the speech onset of the CNP utterances but not of the PP utterances.

The analysis of the error rates was not so revealing. There were fewer errors in Experiment 2 than in Experiment 1, which can be readily attributed to the retrieval of object names during the preview period. Because the difference in error rates between CNP and PP utterances was not significant by items (in neither experiments), and there was no significant interaction between sentence format and experiment, we refrain from further interpretation of the error rate pattern.

This experiment provides two positive pieces of evidence. First, the finding that a picture preview response reduces onset latencies for CNP utterances to a greater extent than latencies for PP utterances supports the functional phrase as

the preferred unit of lexical planning. Second, the difference in onset latencies observed in Experiment 1 between CNP and PP utterances can be primarily attributed to lemma access processes (respectively, two lemmas or one lemma accessed before speech onset) rather than syntactic planning.

EXPERIMENT 3

An uncertainty remains about the latency differences observed in Experiment 1. According to the visual grouping hypothesis, the longer latencies for CNP compared to PP utterances could be due to the perceptual interference from the lower picture that slows the retrieval of the name of the top picture (the first noun). Allum and Wheeldon (2007) reported a “checking” experiment (not reported in full, but cited in their General Discussion) in which CNP utterances were produced in response to two display types: either two pictures colored similarly or one picture colored and one white (the latter being similar to the PP triggering stimulus we have used). They observed no difference in the onset latencies for these two conditions, and they argued that the significant difference between CNP and PP responses could not be accounted for by properties of the stimuli such as perceptual interference between pictures in the same color. However, this manipulation introduced another visual cue, orientation, instead of the color cue to trigger CNP and PP utterances separately, thus complicating the experimental task. This has a bearing on participant’s visual attention and may pollute the comparison between conditions.

Allum and Wheeldon (2009) also investigated the characteristics of visual displays by comparing the preview effect in different utterance forms of coordination (a listing structure or a CNP utterance) in response to the same stimuli (the CNP triggering stimuli), instead of comparing the different displays eliciting CNP and PP utterances. The finding of different preview effects in CNP and listing utterances suggested different planning scopes of lexical retrieval for these two forms, despite them being elicited by the same visual display. Note however that this finding alone cannot clarify the contribution of the manipulation of the display to the differences observed between CNP and PP utterances. This is because they did not compare different displays. Martin et al. (2010) explicitly tested this possibility in the cases where movement is the cue for sentence structures. Their results led them to reject the visual grouping hypothesis in this context and to speculate that it may also be false for color cues (like those used by Allum & Wheeldon, 2007).

We believe that more evidence is needed to evaluate the visual grouping hypothesis in the case of color cues. Following the same logic as Martin et al. (2010), we manipulated the utterances that had to be produced in response to the visual displays (either sentences or word lists). This manipulation allowed us not only to compare the CNP and PP displays with the same utterances but also to estimate the difference of utterances in response to the same displays (through the comparison of results from Experiments 1 and 3). In Experiment 3, the visual displays of Experiment 1 were used, but participants produced both items in an unstructured list (*N1 N2*), regardless of the picture’s color. The filler items were still named in sentences as in Experiment 1. If the latency differences observed in Experiment 1

were caused by the influence of visual grouping to the lemma access, then the same pattern of effects should be observed for sentences (Experiment 1) and lists (Experiment 3). Although there were no CNP or PP utterances in the present experiment, we still use these terms to refer to the presentation displays, in order to compare performance here with that from the previous experiments.

Method

Participants. Twenty-four participants were from the same population as Experiments 1 and 2, and none of them had participated in the earlier experiments. They were all native Chinese speakers with normal or corrected to normal vision.

Results and discussion

Using the same exclusion criteria as in Experiment 1, 7.9% of the data were removed from the latency analysis. Mean latencies and percentage error rates are shown in Table 1.

The same analyses as in Experiment 1 were conducted here. The difference in onset latencies between CNP and PP utterances was significant only by participants, $t_1(23) = 2.29, p < .05$; $t_2(62) < 1$. The difference in error rates was not significant, $t_1(23) < 1$; $t_2(62) = 1.06, p = .29$.

Comparison of Experiments 1 and 3

Two separate analyses were conducted with participants (F_1) and items (F_2) as random factors, including the factors of sentence type and experiment.

The results showed that the interaction in onset latencies between sentence type and experiment was significant, $F_1(1, 46) = 5.79, MSE = 2,193, p < .05$; $F_2(1, 62) = 5.32, MSE = 4,001, p < .05^2$; the interaction in error rates was significant by participants and marginally significant by items, $F_1(1, 46) = 4.86, MSE = 0.2, p < .05$; $F_2(1, 62) = 3.59, MSE = 0.4, p = .063$.

Discussion

The pattern of results in Experiment 3 differs markedly from that observed in Experiment 1. Differences in latency were only significant within participants. Even if these differences were taken to be real, the differences in onset latencies between CNP and PP conditions were much greater in Experiment 1 (76 ms) than in Experiment 3 (29 ms). The interaction between the experiment and utterance format was significant. This result and the significant interaction between sentence format and experiment serve to undermine the visual grouping hypothesis as the sole origin of the theoretically critical difference observed in Experiment 1 between CNP and PP utterances. With respect to error rates, the marginally significant interaction between sentence format and experiment would seem to argue in the same direction.

GENERAL DISCUSSION

We conducted three experiments to test the functional phrase hypothesis that describes the scope of grammatical planning. In Experiment 1, we compared the performance of two types of sentences (CNP and PP) that vary on their functional phrase structure. Our results replicate the latency differences reported in Allum and Wheeldon (2007). Thus, this latency effect is not limited to a particular language. We then conducted two experiments with the same materials and similar designs. In Experiment 2, the picture pair was shown without a color cue prior to the syntactic information necessary to construct a response, and the display of the syntactic information 2 s later acted as the response cue. Under these circumstances, processing related to object-name retrieval had been completed in advance of sentence construction, and was thus assumed to be removed from sentence production latencies. Because there were no latency differences between the two sentence types, the effect observed in Experiment 1 can be attributed to word retrieval rather than syntactic processing. However, it was unclear whether response latency differences in Experiment 1 were due to different scopes of grammatical planning prior to speech onset (the functional phrase hypothesis) or influences of visual grouping to the first lexical retrieval (the visual grouping hypothesis). This last possibility was tested in Experiment 3 with a list-naming task. In this task, the latency differences were very much reduced compared to Experiment 1. This suggests that visual grouping was not the sole contributing process to the latency differences observed in Experiment 1 and that planning scope also contributes. This series of experiments provides further evidence for the functional phrase as a preferred unit of grammatical planning scope (Allum & Wheeldon, 2007) and supports a modification of the phrasal scope suggested in a number of studies (e.g., Levelt & Maassen, 1981; Martin et al., 2004; Smith & Wheeldon, 1999).

The evidence in favor of the functional phrase hypothesis seems robust within the type of experiments reported here, although as we note in our introduction, the preferred planning unit may depend on the circumstances. Typical arguments from evidence in speech error corpora present a different planning dynamic. Word exchange errors occurring in natural speech (e.g., “my *boy* bites the *dog* next door”) usually involve words from the same grammatical category (e.g., nouns exchange with nouns) and, most important for this discussion, words from different phrases (Bock & Levelt, 1994). Garrett (1993) interprets this observation as evidence that clausal scope drives lemma retrieval during sentence encoding processes.

It has also been noted that there are limits to how speech error patterns can be used to constrain accounts of error-free production (e.g., Meyer, 1992). Word exchange errors may not occur within the course of standard sentence planning, precisely because there were too many word candidates readily active or available. In support of this view, it has been reported that lexical availability can drive word order in experimental settings involving error-free speech production. Bock (1986) showed that semantically primed nouns tended to be produced earlier in sentences than nonprimed nouns. In this broader context, the results of speech error corpora are informative, but they do not invalidate the conclusion that the functional phrase is the preferred unit of encoding in the largely error-free speech situations we used.

As a second example, consider the radically incremental hypothesis, which is based primarily on studies using eye-tracking methodology (Griffin, 2001; Meyer et al., 1998; Meyer & van der Meulen, 2000). It states that the advance grammatical planning comprises only the first noun of the CNP. As discussed by Allum and Wheeldon (2007) and Martin et al. (2010), eye-tracking studies of less formulaic language production have revealed that an initial scan of the entire scene takes about 300 ms. In studies where the utterance format presented in each block is constant (i.e., only one utterance format was used in the study or the different utterance formats used were never mixed within a single block of trials) this initial scan is absent (e.g., Griffin, 2001; Meyer & van der Meulen, 2000). It is possible that this initial scan supports not only conceptual encoding of the scene but also the first stages of grammatical encoding, including lemma access for words in initial (functional) phrases. This may explain why the effect of the functional phrase as the preferred planning unit was not observed in the eye-tracking data. Other studies indicate that lexical retrieval is not entirely synchronized with the eye tracking. For example, Morgan and Meyer (2005) reported that while still fixating on the first object to be named, participants simultaneously process the name of a second object up to the level of phonological retrieval.

Meyer's (1996) picture-word interference experiments provide contrasting evidence, showing a semantic interference effect in the onset latencies for the second noun of phrases (e.g., "the dog and the flower") and sentences (e.g., "The dog is next to the flower"). Because the semantic interference effect is often interpreted as difficulty in lexical selection during word production (e.g., Levelt, Roelofs, & Meyer, 1999) and occurs during the grammatical planning stage of processing, this result was interpreted as evidence for clausal grammatical planning units. As discussed by Meyer (1996), one way to explain the difference between her results and the evidence of phrasal grammatical planning units is that the onset latency effect between utterances in the latter cases did not arise during lexical selection but during the generation of the syntactic structure. In Experiment 2, we tested this hypothesis directly and found no latency effect when the picture pair was previewed, suggesting that the syntactic retrieval per se did not make a significant contribution to the latency effect.

An alternative explanation to the difference between Meyer's (1996) results and the evidence of phrasal grammatical planning units is that speakers choose a clause as the planning unit under the specific experimental circumstances she used. As Konopka (2012; see also Wagner et al., 2010) observed, the lexical planning scope could be influenced by the structural accessibility, such that the scope of grammatical planning might be larger when the sentence structure is easier to retrieve. In Meyer's (1996) study, the utterance type was constant in each experiment. As a result, the phrase/sentence structure was highly primed and could be prepared even before the stimulus presentation. Moreover, picture pairs were presented on the screen for only 800 ms, shorter than the mean response onset latency observed in her study. This means that in many trials, the picture pairs disappeared before participants began to articulate. All these manipulations would make participants focus on the lexical retrieval and attempt to encode the second noun more thoroughly or earlier to comply with the experimenter's request for fast and accurate responses.

In studies comparing different types of utterances, including the present study, there were at least two different utterance formats as experimental ones, and fillers with different utterance structures from the experimental ones were added in many cases. Therefore, participants need to vary utterance formats from trial to trial within each block. The presentation of the stimuli was also long enough for speakers to begin their articulation, thus speakers need not retrieve both lemmas before launching a response and could adopt the functional phrase as a preferred planning unit for fluent sentence production. Once again, such suggestions must be stated carefully, considering the sensitivity of the preferred grammatical planning scope to diverse factors (time pressure: Ferreira & Swets, 2002; structural accessibility: Konopka, 2012; Wagner et al., 2010). The flexibility of grammatical planning scope is an important avenue to extend the current research.

To conclude, our study extends the empirical support for the hypothesis that the functional phrase is a preferred grammatical unit of speech planning. The evidence has been generalized from Japanese (Allum & Wheeldon, 2007) to a novel language: Mandarin Chinese. Syntactic processing and visual grouping may be excluded as main factors driving the latency differences between conjoined and prepositional utterances. In summary, the three experiments that we reported provide further evidence that, under the kind of experimental conditions used here, the functional phrase is the preferred planning unit at the grammatical level.

APPENDIX A

Experimental pictures used in Experiments 1, 2, and 3

Set 1		Set 2	
凳子 /deng zi/	stool	大衣 /da yi/	coat
茶杯 /cha bei/	cup	纽扣 /niu kou/	button
书桌 /shu zhuo/	desk	熨斗 /yun dou/	iron
镰刀 /lian dao/	hook	凉鞋 /liang xie/	sandal
毛笔 /mao bi/	paintbrush	围巾 /wei jin/	shawl
礼物 /li wu/	parcel	毛衣 /mao yi/	sweat
书包 /shu bao/	pocketbook	浴缸 /yu gang/	bathtub
小号 /xiao hao/	trumpet	箱子 /xiang zi/	suitcase
花瓶 /hua ping/	vase	盒子 /he zi/	box
风车 /feng che/	windmill	夹子 /jia zi/	clothespin
篮子 /lan zi/	basket	锤子 /chui zi/	hammer
笛子 /di zi/	flute	钳子 /qian zi/	pliers
算盘 /suan pan/	abacus	铁链 /tie lian/	chain
帆船 /fan chuan/	sailboat	教堂 /jiao tang/	church
插头 /cha tou/	plug	梯子 /ti zi/	ladder
冰箱 /bing xiang/	refrigerator	栅栏 /zha lan/	fence

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (31070989). We thank Thomas McKeeff and Dashiel Munding for native proofreading and Alfonso Caramazza for very helpful comments and support on an earlier version of the manuscript.

NOTES

1. Except for the previewing period, the experimental displays and utterances were the same in Experiments 1 and 2, so the factor of experiment was considered as a within-item factor.
2. The displays were taken as “items” in the item analysis as in the comparison of Experiments 1 and 2, and therefore, the factor of experiment was also considered as a within-item factor.

REFERENCES

- Alario, F.-X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299–319.
- 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.
- Allum, P. H., & Wheeldon, L. R. (2009). Scope of lexical access in spoken sentence production: Implication for the conceptual–syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1240–1255.
- 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., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). New York: Academic Press.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.
- Costa, A., Navarrete, E., & Alario, F. X. (2006). Accessing object names when producing complex noun phrases: Implications for models of lexical access. *Cognitiva*, *18*, 3–23.
- 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.
- Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (pp. 133–177, 505–529). New York: Academic Press.
- Garrett, M. F. (1976). Syntactic processes in sentence production. In R. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–255). Amsterdam: North-Holland.
- Garrett, M. F. (1993). Errors and their relevance for models of language production. In G. Blanken, J. Dittmann, H. Grimm, & J. C. Marshall (Eds.), *Linguistic disorders and pathologies* (pp. 72–92). Berlin: Walter de Gruyter.
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1–B14.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.
- 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.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the linguistic boundaries* (pp. 221–252). Dordrecht: Reidel.

- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., II, & Yang, C. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, 116, 177–192.
- Martin, R. C., Miller, M., & Vu, H. (2004). Lexical–semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, 21, 625–644.
- Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, 42, 181–211.
- Meyer, A. S. (1996). Lexical access in phrase and sentence production. *Journal of Memory and Language*, 35, 477–496.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66, B25–B33.
- Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychological Bulletin and Review*, 7, 314–319.
- Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 428–442.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, 35, 189–213.
- Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, 11, 17–50.
- Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73, 205–246.
- Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production: An online study. *Cognition*, 78, 123–164.
- Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 675–686.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experiment Psychology: Human Learning Memory*, 6, 174–215.
- 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.
- Wheeldon, L. R., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and advanced planning in spoken sentence production. *Quarterly Journal of Experimental Psychology*, 36, 423–440.
- Yang, J. C., & Yang, Y. F. (2008). Horizontal flow of semantic and phonological information in Chinese spoken sentence production. *Language and Speech*, 51, 267–284.
- Zhang, Q. F., & Yang, Y. F. (2003). The determiners of picture-naming latency. *Acta Psychologica Sinica*, 35, 447–454.