



Phonological planning during sentence production: beyond the verb

Tatiana T. Schnur*

Department of Psychology, Rice University, Houston, TX, USA

Edited by:

Bradford Mahon, University of Rochester, USA

Reviewed by:

Evelina Fedorenko, Massachusetts Institute of Technology, USA

Niels Janssen, Universidad de La Laguna, Spain

***Correspondence:**

Tatiana T. Schnur, Department of Psychology, Rice University, 6100 Main Street, Houston, TX 77005, USA.

e-mail: tschnur@rice.edu

The current study addresses the extent of phonological planning during spontaneous sentence production. Previous work shows that at articulation, phonological encoding occurs for entire phrases, but encoding beyond the initial phrase may be due to the syntactic relevance of the verb in planning the utterance. I conducted three experiments to investigate whether phonological planning crosses multiple grammatical phrase boundaries (as defined by the number of lexical heads of phrase) within a single phonological phrase. Using the picture–word interference paradigm, I found in two separate experiments a significant phonological facilitation effect to both the verb and noun of sentences like “He opens the gate.” I also altered the frequency of the direct object and found longer utterance initiation times for sentences ending with a low-frequency vs. high-frequency object offering further support that the direct object was phonologically encoded at the time of utterance initiation. That phonological information for post-verbal elements was activated suggests that the grammatical importance of the verb does not restrict the extent of phonological planning. These results suggest that the phonological phrase is unit of planning, where all elements within a phonological phrase are encoded before articulation. Thus, consistent with other action sequencing behavior, there is significant phonological planning ahead in sentence production.

Keywords: language production, sentences, phonological planning, verbs, picture–word interference paradigm, lexical frequency

PHONOLOGICAL PLANNING DURING SENTENCE PRODUCTION: BEYOND THE VERB

When we speak, we perform a series of actions. We generate an idea. We select the words to convey that idea and order them according to the grammatical rules of the language. Finally, we retrieve the sounds in an order corresponding to those words and perform the motor movements to begin speaking. Because speech is produced sequentially over time, the idea we want to convey has to be translated into components that can be produced in linear order. Earlier stages of speech production involve larger representations (e.g., the idea) but at later stages the representation becomes smaller (i.e., it corresponds to the word being produced at that moment in time).

In advance of articulation, to what extent must the utterance be phonologically planned? Previous evidence suggests that planning is either fully incremental (one phonological word at a time; Meyer, 1996; Wheeldon and Lahiri, 1997; Levelt et al., 1999), or alternatively, encompasses larger units, such as a phrase (Smith and Wheeldon, 2004; Schnur et al., 2006). However, results are also consistent with the extent of planning being driven by the syntactic importance of the verb (Ferreira, 2000; Ferreira and Swets, 2002). In the following, I present evidence in spontaneous sentence production demonstrating that phonological planning is not incremental, is not restricted by the verb, but encompasses a full phonological phrase. Following evidence for other behaviors that require the sequencing of actions like motor planning (Rosenbaum, 2010) and problem solving (Catrambone, 1998),

these results suggest that planning in sentence production is not incremental, but encompasses larger chunks of information. Planning larger phonological units may facilitate comprehension for the listener while helping the speaker avoid hesitations (Fox Tree, 1995).

Theories concerning the extent of phonological planning can be generally divided into one of two types, incremental or non-incremental. In one of the most fully articulated models of speech production, Levelt (1989) proposed that grammatical and phonological encoding is highly incremental in order to facilitate fluent speech: “Each processing component will be triggered into activity by a minimal amount of its characteristic input” (Levelt, 1989, p. 26). For planning at the phonological level, for articulation to begin, the minimal (and thus most incremental) unit required is a grouping of phonological segments known as a phonological word (PW), a unit of prosodic structure (Wheeldon and Lahiri, 1997; Levelt et al., 1999).

The creation of prosodic structure is part of phonological encoding and occurs when multiple words are produced in a sequence. A PW is a content word (which automatically receives stress) and any unstressed function word such as auxiliaries, determiners, conjunctions, and prepositions¹. For example, because [gAte] receives stress (stress indicated by capitalized vowels), it is a PW when produced alone and when combined with a non-stressed

¹For discussion as to which words group with a content word, see Selkirk (1984), Nespor and Vogel (1986), Selkirk (1986), Ferreira (1993), and Selkirk (1996).

word like “the” or “a,” e.g., [the gAte]_{PW}. PWs are grouped into larger structures, such as phonological phrases. Phonological phrases are created from groups of PWs and derived from syntactic structure. Specifically, all PWs that fall within a major grammatical phrase up to the phrase’s right boundary are grouped together to form a phonological phrase (PP; the X-max algorithm, Selkirk, 1986; also see Levelt, 1989; Ferreira, 1991, 1993). For example, the sentence, “He opens the gate,” is comprised of two PWs which form one phonological phrase [[He Opens]_{PW} [the gAte]_{PW}]_{PP}. This is because the right boundary of the syntactic phrase is the end of the verb phrase (VP). Although “he” has special syntactic status because it is the lexical head of the subject noun phrase (NP), because it is a monosyllabic function word (produced with reduced stress), it is not identified as a head of phrase when phonological phrases are constructed (Selkirk, 1984, 1986, 1996)². As a result, when the PWs are combined, only one phonological phrase is created in [he Opens the gAte]_{PP}.

Because of the type of utterances investigated, evidence of incremental phonological planning (a PW) also can be interpreted as non-incremental, extending to either a phonological phrase, or alternatively to the verb of the sentence. Meyer (1996) using picture naming in the picture–word interference paradigm, found faster naming times when the distractor word was phonologically related to the first object (e.g., the arrow) of a sentence, which is both a PW and phonological phrase (PP). Phonological planning was not found for the second object and phonological phrase (e.g., is next to the bag) which occurred after the verb (“[The Arrow]_{PW/PP} [is nExt to the bAg]_{PP}”). Similarly, Oppermann et al. (2010a) found that participants were faster to produce sentences when phonologically related distractors were related to the first PW and PP before the verb, but no effects were found for the object after the verb (e.g., “Die Maus frisst den Käse”: [The mOUse]_{PW/PP} [[EAts]_{PW} [the chEEse]_{PW/PP}; the verb was not tested). Thus, phonological planning encompassed both a PW and PP, and extended up to the verb of the sentence.

The non-incremental view of phonological planning suggests that more is planned in advance of articulation, where the unit of planning may be a phonological phrase (Miozzo and Caramazza, 1999; Alario and Caramazza, 2002; Alario et al., 2002; Costa and Caramazza, 2002; Smith and Wheeldon, 2004; Schnur et al., 2006; Damian and Dumay, 2007). In Schnur et al. (2006), participants were faster to produce sentences when distractors were phonologically related to the verb (e.g., [The Orange gIrl]_{PP} [wAlks]_{PW/PP}) showing that planning extended two phonological phrases. Smith and Wheeldon (2004) found that phonological planning encompassed a phonological phrase using a different paradigm to study sentence production. Participants were faster to begin speaking sentences when words were phonologically related to each other in a single phonological phrase (e.g., [[The flAg and the bAg]_{PP} [move up]_{PW/PP}) compared to when objects were named across phrase boundaries [The flAg]_{PW/PP} [mOVes]_{PW/PP} [abOve the bAg]_{PP}). In both these cases, phonological planning extended multiple PWs, encompassing entire phonological phrases.

²For a full discussion of the prosodic status of function words, see Selkirk (1984, 1986, 1996) and Nespor and Vogel (1986).

Evidence from speech errors is supportive of planning of a phonological phrase as sound errors largely occur within phrase boundaries. For example, based on an analysis of sound-exchange errors Garrett (1975) found that sound exchanges occur within a phrase, as opposed to across clauses, approximately 87% of the time (where a phrase is defined as a simple NP or VP). For the remaining 13% of errors, Garrett (1975) found that 11% of those occurred between a verb and its direct object NP (14 of 19 errors; e.g., “he was slowing shides”; Garrett, 1975) and the remaining 3% were between a verb and its subject NP (5 of 19 errors; e.g., “you should have your brÅkes chÅked”; Garrett, 1975).

An alternative non-incremental view is that lexical heads of phrase, in particular verbs, drive the extent of planning for sentence production. In a model of syntactic parsing, Ferreira (2000) proposed that verbs are encoded first in order to establish syntactic structure for the sentence (or, the verb is encoded early in sentence planning; cf. Bock, 1987; Griffin, 2000). Syntactic encoding of the verb is required to assign grammatical roles such as subject and object. Under this proposal, only when subject roles are assigned as a result of the syntactic encoding of the verb can phonological encoding begin (Ferreira and Swets, 2002). If we assume that grammatically encoded representations automatically access their phonological information (e.g., Levelt et al., 1999), then phonological planning may depend not on phonological phrase boundaries, but instead, on the syntactic encoding of the verb. Evidence of phonological planning in sentence production is consistent with this account as planning extended to the verb but not beyond, in all cases (e.g., Meyer, 1996; Smith and Wheeldon, 2004; Schnur et al., 2006; Oppermann et al., 2010a).

Thus, it is unclear what drives the extent of phonological planning in spontaneous sentence production. Sentence production evidence that planning is minimal (a PW) is also consistent with non-incremental planning. Under a non-incremental view, the unit of phonological planning is larger, either a phonological phrase, or driven by the syntactic importance of the verb. If the phonological phrase is a minimal unit of phonological encoding, then following Levelt’s (1989) notion of incrementality, all elements of that unit should be phonologically encoded before articulation. Alternatively, if the verb drives encoding, then phonological encoding may not extend beyond the verb, as previous results suggest.

CURRENT EXPERIMENTS

The goal of the experiments presented here was to determine whether for sentence production, phonological planning extends only up to the verb, or whether the unit of planning is a phonological phrase, where all components of the phrase are encoded before articulation. In the experiments presented here, participants described pictures in the format “He (She) opens the gate,” prosodically defined as single phonological phrase. This utterance format allows the examination of whether all components of a single phonological phrase, especially components following the verb, are phonologically encoded at sentence onset.

Experiments 1 and 3 used the picture–word interference paradigm which elicits spontaneous speech while controlling the nature of what is produced. In picture–word interference, participants describe pictures, while ignoring visually presented

distractor words. Picture naming is accelerated when the word sounds like the picture name in comparison to an unrelated word. This is referred to as the phonological facilitation effect (Lupker, 1982; Rayner and Springer, 1986; Meyer and Schriefers, 1991). This acceleration of speech may reflect the distractor's influence on both the distractor and the to-be-produced target word-form/morphemes (Starreveld and La Heij, 1995, 1996; Starreveld, 2000). Alternatively, it may reflect the influence on the target segments that are inserted into the metrical frame (Meyer and Schriefers, 1991). Like many others (e.g., Meyer, 1991; Meyer and Schriefers, 1991; Costa and Caramazza, 2002; Roelofs, 2002; Damian and Dumay, 2007), I assume that the bulk of the phonological facilitation effect on a produced word is a result of the word's phonological encoding before articulation, i.e., the retrieval of its phonological representation. Recently, Oppermann et al. (2010a) suggested that the phonological facilitation effect is a by-product of picture-viewing processes (further discussed following Experiment 1). Given the debate concerning the locus of the phonological facilitation effect, I provide converging evidence of advance planning of phonological representations using a simpler paradigm (picture description without word distractors) where I manipulate a lexical and phonological property of to-be-produced words, lexical frequency (see Experiment 2).

To my knowledge, this is the first set of experiments that addresses whether all components of a phonological phrase (including post-verb elements) are phonologically encoded upon sentence initiation. These experiments also offer an improvement over previous sentence production studies in approximating real-world speech by using more verbs (16–28) and objects (32) than in previous work³. Lastly, using two different chronometric methods, the experiments provide converging evidence of phonological encoding of post-verb representations.

EXPERIMENT 1

In Experiment 1, phonologically related and unrelated distractors to the verb were presented during production of utterances similar to “He opens the gate”. During sentence production, if phonological planning extends through the first PW (Meyer, 1996; Smith and Wheeldon, 2004; Oppermann et al., 2010a) and/or to the verb (Schnur et al., 2006), then participants should be faster to produce utterances in the presence of a phonologically related distractor to the verb in comparison to an unrelated distractor.

METHOD

Participants

Sixteen Harvard University undergraduate students were paid or received credit for an introductory psychology course. All were native English speakers and none participated in other experiments.

Materials

Twenty-eight line drawings depicting actions were used as target stimuli (modified from the materials used in Masterson and

Druks, 1998; see **Table A1** in Appendix). All pictures depicted an actor performing transitive actions. Although named as “he” or “she,” an actor was depicted as either a boy, girl, man, or woman so that 7 of the 28 actions fell into each category. No agent, action, or object shared initial phonemes for any picture. Each picture was presented with four distractor words: (a) phonologically related to the verb (e.g., client for climb); (b) phonologically unrelated to the verb (e.g., peer for climb); (c) a baseline condition (a string of 6 X's printed inside each picture); and (d) a filler condition. The filler condition was not analyzed. The pictures and the distractors were paired so that each distractor appeared once in the phonologically related and once in the unrelated conditions. This design controlled for unintentional pairing effects between different sets of distractors beyond the phonological relatedness with the verbs. Distractors were chosen so that they did not sound similar to the agent or object of the sentence. Phonologically related distractors shared the first two segments with the verb of the picture. At the beginning of each block, four pictures were included as warm-up trials.

The experimental stimuli were presented in 4 different blocks of 32 trials each for a total of 128 trials. The trials were randomized so that (a) the same picture did not occur twice in the same block; (b) the same distractor condition occurred no more than three times in a row; and (c) no agent occurred more than twice in a row. Care was taken so that no item from one trial was semantically or phonologically similar to an item in the following trial. Four different block orders were designed and presented to participants according to a Latin-square design.

The distractors were shown in 28-point boldface capital letters in Geneva font, superimposed on the pictures. Pictures were centered at fixation. Word position varied randomly in the region around fixation to prevent participants from systematically fixating the portion of the picture not containing the distractor. However, for an individual picture, the position of all its distractors was the same.

Before the experiment proper, participants had two practice series. In the first series participants were presented with all the pictures with a series of X's printed inside each picture, to train the subject to use the correct name for each picture. In the second practice series, they were presented with all the pictures with practice distractors printed inside every picture. These practice distractors were not used during the experiment.

Apparatus

The pictures were presented on a Macintosh using the PsychLab program (Bub and Gym, University of Victoria, British Columbia, Canada). Response times (RTs) were measured to the nearest millisecond by means of a voice key (KOSS headset/CMU voice box) from appearance of the picture until the voice key was triggered.

Procedure

Participants were asked to produce complete sentences, naming the subjects (using the pronouns “he” or “she”), the action (using the third person singular verb form) and the object (using the object's name; e.g., “He opens the gate”). Participants were tested individually in a darkened testing room. They were instructed to name pictures as quickly and as accurately as possible. When

³Other studies of phonological planning in sentence production used fewer objects and verbs (Meyer, 1996: 1 verb/72 objects; Schnur et al., 2006: 28 verbs; Oppermann et al., 2010a: 20 verbs/objects; Smith and Wheeldon, 2004: 1 verb/20 objects).

participants made mistakes during the practice session, they were asked to name the picture correctly. Each trial proceeded as follows: A fixation point (+) was shown for 700 ms, followed by presentation of the stimulus 300 ms later. There was a 2000 ms pause between trials. The experimenter remained in the testing room in order to record incorrect responses and when voice key malfunctions occurred. A session lasted approximately 25 min.

Analyses

Three types of responses were classified as errors: (a) production of the wrong word; (b) verbal disfluencies (stuttering, utterance repairs, etc.); and (c) voice key malfunctions. Responses faster than 300 ms and 3 SDs from a participant's condition mean were also eliminated. Separate analyses were carried out on the RTs using either the means per subject or means per item as dependent variables yielding $F1$ and $F2$ statistics, respectively. One variable was analyzed: Type of distractor (phonologically related, unrelated, baseline). Type of distractor was considered a within-subject and within-item variable. I report three different ANOVA's: an error analysis, a baseline analysis, and the principal analysis. For the error analysis, all three conditions were included. The baseline analysis compared RTs for the unrelated condition vs. baseline. The baseline condition was not included in further analyses. The principal analysis compared RTs for the phonologically related vs. unrelated conditions.

RESULTS AND DISCUSSION

Table 1 reports a summary of the data. **Figure 1** shows the magnitude of the phonological facilitation effect and the 95% confidence intervals. The naming latencies from two items were removed because they elicited a high percentage of errors (more than 30%). Error rates consisted of 12.2% of the data before outliers were removed and 12.5% of the data after outliers were eliminated. There was no significant difference in the number of errors produced across the three conditions ($F_s < 1$).

Response times in the phonologically related condition (782 ms) were significantly faster in comparison to the unrelated condition [806 ms; $F1(1, 15) = 5.84$, $MSE = 93862$, $p = 0.02$; $F2(1, 25) = 4.29$, $MSE = 109404$, $p = 0.04$]. The XXX condition (792 ms) did not produce statistically faster naming latencies in comparison to the unrelated condition [806 ms; $F1(1, 15) = 3.02$, $MSE = 27752$, $p = 0.10$; $F2(1, 25) = 1.74$, $MSE = 37721$, $p = 0.19$] although the effect (14 ms) was in the predicted direction.

Table 1 | Experiment 1. Mean RTs (ms), SD, and percentage of errors (Error %), for phonologically related, unrelated, and baseline conditions.

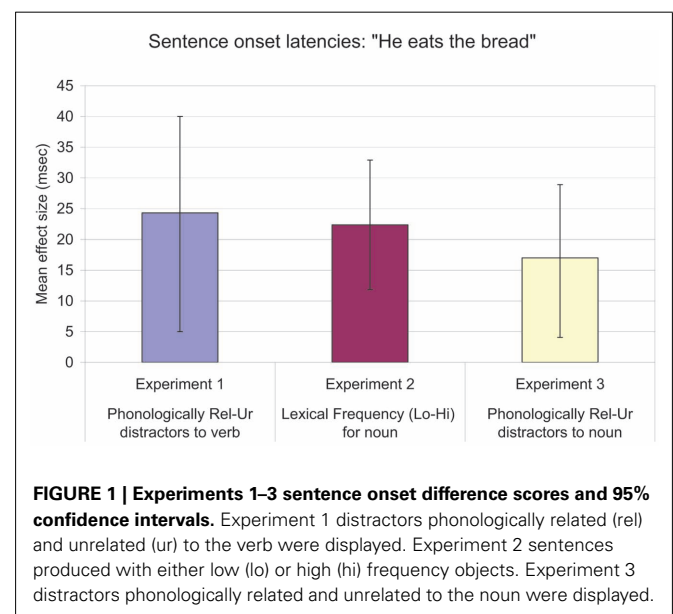
Type of distractor	Mean	SD	Error (%)
Phonologically related	782	99	11.0
Phonologically unrelated	806	107	12.7
XXX's	792	112	13.7
Phonological effect (related-unrelated)	-24*		

A significant difference of $p < 0.05$ is indicated by an *.

The experiment showed that production of a sentence was facilitated with a distractor phonologically related to the verb compared to an unrelated distractor. This suggests the verb, part of the first PW, was phonologically encoded before articulation. Phonological encoding of the first PW before articulation of sentences is consistent with previous evidence that phonological planning extends at least one PW in advance (Levelt, 1989; Wheeldon and Lahiri, 1997) and/or planning extends to the verb in sentence production (Smith and Wheeldon, 2004; Schnur et al., 2006; Oppermann et al., 2010a).

Having replicated previous results showing phonological planning through the first PW and verb in a different set of sentences, the critical question to be addressed is whether in sentence production phonological planning is non-incremental, extending an entire phonological phrase. If phonological planning in sentence production is defined by phonological phrase boundaries then I expect phonological effects to the direct object NP at articulation of the sentence for sentences like, "He opens the gate." If the PW and/or verb limit the extent of phonological planning, I expect no phonological effects for representations following the verb.

Recently, Oppermann et al. (2010a) suggested that the phonological facilitation effect in the picture-word interference paradigm is not a measure of the retrieval of phonological representations and thus is an invalid method for measuring advance phonological planning in sentence production. Specifically, they suggest that the phonological facilitation effect is a by-product of picture-viewing processes. The evidence for this position is mixed, and is derived from measuring effects of unnamed pictures on the naming latencies of pictures or words. In some cases, unnamed pictures phonologically related to the to-be-named picture or word facilitate naming which suggests that participants may automatically access the phonological representations of unnamed pictures (Morsella and Miozzo, 2002; Navarrete and Costa, 2005; Humphreys et al., 2010). However, this effect may only occur when pictures are semantically related (Oppermann et al., 2010b),



thematically related (Oppermann et al., 2008), as a result of attentional effects (Roelofs, 2008; Malpass and Meyer, 2010), or strategic effects (Bloem et al., 2004). In some cases, the effect is not found at all (Jescheniak et al., 2009).

Given the unique situations required to create the picture–picture phonological effect, I am in agreement with others (e.g., Jescheniak et al., 2009; Oppermann et al., 2010b) that simply viewing pictures does not activate their phonological representations. However, to provide further evidence against this possibility I used an additional measure of advance phonological planning during sentence production. In Experiment 2, participants simply described pictures (no word distractors presented) and the frequency of the object name was varied. Lexical frequency effects are a “litmus test” for evidence of phonological encoding (Kittredge et al., 2008), but may reflect retrieval at both lexical and phonological levels (Knobel et al., 2008; Strijkers et al., 2009). Thus, in Experiment 3, I returned to the phonological facilitation effect to provide converging evidence of the phonological encoding of the direct object NP.

EXPERIMENT 2

In Experiment 2, participants produced sentences similar to Experiment 1, “He opens the gate,” consisting of a phonological phrase with two PWs. Phonological encoding of the end of the phonological phrase, the direct object NP was investigated using the frequency effect. I tested whether varying the frequency of the direct object in the production of transitive sentences [e.g., He opens the door (HF) vs. He opens the gate (LF)] affected RTs in initiating sentence production.

It is well established that the frequency of a picture’s name affects the speed of naming the picture (Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994; Griffin and Bock, 1998; Levelt et al., 1998). Naming of low-frequency items is slower than the naming of high-frequency items. There is some controversy as to the production level on which the frequency effect in picture naming occurs. Three possible loci for the frequency effect are proposed: at the recognition (input) level, the lexical level, and at the phonological level.

Low-frequency objects may be named more slowly than high-frequency objects because it takes longer to recognize an object of low-frequency because the object itself does not appear frequently in our environment (Kroll and Potter, 1984). Using an object-decision task where no vocal response was required (participants pressed a button as to whether a displayed object represented a real or a non-existent object) Kroll and Potter (1984) found responses were slower to low-frequency objects compared to high-frequency objects. Assuming the task was sensitive to recognition processes, these results suggest that the frequency effect is based on the speed at which an object is recognized. However, other researchers using similar paradigms where a vocal response was not required do not replicate this effect (Wingfield, 1968; Jescheniak and Levelt, 1994; Griffin and Bock, 1998; Levelt et al., 1998; Meyer et al., 1998). Using a different paradigm (delayed picture naming) Almeida et al. (2007) further confirmed that lexical frequency effects in naming do not arise from input level processes.

In general, the frequency effect is thought to be located either at both lexical and phonological levels (Knobel et al., 2008; Strijkers

et al., 2009) or primarily at the phonological level (Jescheniak and Levelt, 1994; Jescheniak et al., 2003; Kittredge et al., 2008; although see Caramazza et al., 2001; Caramazza et al., 2004). Convincing evidence that the bulk of the frequency effect arises during selection of the phonological representation comes from speech error data. In a large-scale study of 50 speakers with acquired language deficits, Kittredge et al. (2008) found that phonological errors during picture naming (e.g., pillow for pineapple) occurred more often when pictures had low-frequent vs. high-frequent names (also seen in a smaller-scale study of speech errors by Stemmer and McWhinney, 1986). Importantly, Kittredge et al. found that this effect of picture name frequency was significantly larger for phonological than semantic errors (e.g., apricot for pineapple) which localizes the bulk of the effect at the level of phonological retrieval.

To test whether the direct object NP of a transitive sentence is phonologically encoded before articulation I manipulated the lexical frequency of the object (high vs. low). If the direct object is phonologically encoded before articulation begins, I expected RT differences between sentences with high- and low-frequency direct objects, where high-frequency completions should be produced more quickly than low-frequency ones.

To rule out a conceptual/input interpretation of the frequency effect I included several control experiments (described in Experiment 2 Materials). First, the degree to which an object is consistently named by one name (name agreement) has an effect on the time it takes to name an object independent of lexical frequency (e.g., Snodgrass and Yuditsky, 1996; Barry et al., 1997; Ellis and Morrison, 1998). To address this concern, name-agreement probabilities were collected for both the action and the direct object with which it was depicted. Second, RTs may be slower for actions with low-frequency objects in comparison to actions with high-frequency objects because of some inherent unnaturalness of the action and object occurring together. To address this concern, participants judged the “naturalness” of action–object pairings. I also measured object-recognition latencies (following Jescheniak and Levelt, 1994) to control for a pre-lexical locus of the frequency effect. Lastly, to ensure that the objects could reliably produce a frequency effect independent of sentence context, participants viewed the same pictures, but named the direct object in isolation (Experiment 2b).

METHOD

Participants

One hundred six Harvard University undergraduate students participated in Experiment 2. Twenty-four were assigned to Experiment 2a (sentence production) and 24 to Experiment 2b (noun production). Twenty-two participated in verb-name picture agreement ratings, 22 participated in noun-name picture agreement ratings, and 20 participated in assessment of action–object picture naturalness. Fourteen subjects participated in an object-recognition experiment to control for input processing differences between low and high-frequency objects. All were paid or received credit for an introductory psychology course. All were native speakers of English. Participants who judged action–object naturalness participated in either Experiments 2a or 2b before completing the naturalness rating. Participants who named objects

did so as part of the practice trial for Experiment 3. Otherwise, none participated in other experiments.

Materials

Sixteen line drawings depicting actions were used as target stimuli (modified from the materials used in Masterson and Druks, 1998; see **Table A2** in Appendix). All pictures depicted an actor performing a transitive action. An actor was depicted as either a female or male so that 8 of the 16 actions fell into each category. Each action was presented with a high-frequency object (average: 156 occurrences per million in the Celex word frequency count (Baayen et al., 1998) range: 67–386; and 109 in the Francis and Kucera (1982) count of written words) and separately with a low-frequency object (average: 23 occurrences per million in the Celex Baayen et al., 1998; range: 2–72; and 15 occurrences per million in the Francis and Kucera, 1982 word count). The 32 common objects were selected from a combination of picture collections (Cycowicz et al., 1997; Masterson and Druks, 1998). Objects were inserted into the action scenes in the same position in order to maintain visual similarity between action–object pairs. The names of the two objects for each action picture were equated as closely as possible in number of syllables (HF group = 1.3, LF group = 1.5). Selection of objects was restricted so that: (1) no object began with the same sound as any other word in the sentence it was part of; (2) no object rhymed with any other word in the sentence; (3) for each action, objects were maximally different in terms of written name frequency ($p < 0.001$), but conceptually plausible, i.e., the action could reasonably occur on both objects.

Each picture was presented once per block. Each picture was repeated three times, so that an action was repeated six times, three times with a low-frequency object, and three times with a high-frequency object. At the beginning of each experimental block, four pictures were included as warm-up trials. The experimental stimuli were presented in six different blocks of 20 trials in each block (16 experimental and 4 filler trials) for a total of 120 trials. The presentation of the pictures in each block was randomized with the following restrictions: (a) before repetition of a picture, all other pictures were presented at least once; (b) the pictures' names (actions and objects) from trial to trial were neither semantically nor phonologically related; (c) the pronouns (he or she) were not phonologically related to an action or object from trial to trial; (d) no pronoun was named more than three times in a row. Six different block orders were designed and presented to participants according to a Latin-square design.

Before the experiment proper, participants had one practice series. Participants were presented with all the pictures. After naming the picture, the correct response was displayed in order to verify the correct utterance for each picture.

CONTROLLING FOR NON-LEXICAL EFFECTS OF ACTION–OBJECT PICTURE PAIRS

To ensure that an action was consistently named as the same action whether it occurred with a high- or low-frequency object name, participants ($n = 22$) named the action depicted in every item to collect name-agreement ratings for the verb. Similarly, name-agreement ratings were collected for high- and low-frequency

objects ($n = 22$ participants). Lastly, to investigate whether low-frequency object sentences were produced more slowly, not because of selection of the object's name, but instead because the actions paired with the low-frequency objects were in some way unnatural or strange, judgments of action–object pairings in terms of “naturalness” were collected [on a seven-point scale, where one indicated unnatural action–object pairs (e.g., kiss frog), and seven highly natural action–object pairs (e.g., paint wall); $n = 20$ participants].

Overall, low and high-frequent object sentences did not differ in verb-name agreement (paired t -test: $p > 0.20$) or object-name agreement (paired t -test: $p > 0.40$). The average naturalness rating for actions depicted with high-frequency objects was 5.3 vs. 4.1 for actions depicted with low-frequency objects. In a paired t -test, this difference was significant ($p < 0.01$). As such, naturalness ratings were used as a covariate in the frequency analysis in sentence production (Experiment 2a). See **Table A2** in Appendix for individual verb and object name-agreement ratings, as well as naturalness ratings for action–object pairs.

In order to test whether the frequency effect was due to speed of object recognition instead of lexical frequency, I used a picture recognition task following Jescheniak and Levelt (1994). A word was displayed immediately followed by a picture. Participants ($n = 14$) decided whether the word denoted an object in the scene by pushing a yes or no button. For example, after reading the word “gate,” participants would respond yes if the subsequent action scene depicted a gate and no if a carrot were depicted. All target objects were yes trials where the word matched the object depicted in the picture. RTs did not differ between the high- and low-frequency objects depicted in each scene ($F_s < 1$). In order to ensure that words did not prime responses and thus mask an RT difference between high- and low-frequency objects, I included a set of 28 filler items interspersed between target items (the same 16 actions but with a new set of high and low-frequent objects where filler objects were maximally different in terms of written name frequency ($p < 0.001$)). Responses to filler items were always no. There was no difference between object-recognition RTs for filler action pictures depicted with high- and low-frequency objects [$F_1(1, 13) = 3.65$, $MSE = 51495$, $p = 0.08$; $F_2 < 1$]. Thus, the results demonstrate that high and low-frequent objects were equally “recognizable/perceivable.”

Apparatus

The pictures were presented on a Macintosh using the PsychLab program (Bub and Gym, University of Victoria, British Columbia, Canada). RTs were measured to the nearest millisecond by means of a voice key (KOSS headset/CMU voice box) from onset of the picture until the voice key was triggered.

Procedure

In Experiment 2a, participants were asked to produce full sentences (e.g., He opens the gate). In Experiment 2b, participants were asked to name the object being acted upon (e.g. gate). Participants were tested individually in a darkened testing room. They were instructed to name the pictures as quickly and as accurately as possible. Each trial proceeded as follows: A fixation point (+) was shown for 500 ms, followed by presentation of the stimulus 300 ms

afterward. Pictures were centered at fixation. Pictures were displayed until either the voice key was activated or for a maximum of 2000 ms. The experimenter remained in the testing room in order to record incorrect responses and when voice key malfunctions occurred. A session lasted approximately 25 min.

Analyses

Three types of responses were classified as errors: (a) production of the wrong name; (b) verbal disfluencies (stuttering, utterance repairs, etc.); and (c) voice key malfunctions. Responses faster than 300 ms and 3 SDs from a participant's mean were also eliminated. Separate ANOVAs were carried out on the RTs using either the means per subject or means per item (sentence) as dependent variables yielding $F1$ and $F2$ statistics, respectively. Two variables were analyzed: Frequency (low-frequency object vs. high-frequency object in an action scene), and Repetition (of object) and both were treated as within-subject and within-item variables. There were no *a priori* predictions concerning repetition and its interaction with frequency. However, repetition is normally included in analyses of frequency (e.g., Jescheniak and Levelt, 1994; Griffin and Bock, 1998). I report two different ANOVAs: an error analysis and the principal analysis.

RESULTS

Table 2 presents a summary of the data for Experiments 2a and 2b. **Figure 1** shows the magnitude of the difference in the lexical frequency effect for sentences (Experiment 2a), with 95% confidence intervals. Error rates for Experiment 2a consisted of 11.4% of the data before outliers were removed and 12.7% of the data after outliers were eliminated. Error rates for Experiment 2b consisted of 4.5% of the data before outliers were removed, and 6% of the data after outliers were eliminated. For Experiments 2a and 2b error rates were not significantly different across high-frequency and low-frequency conditions [Experiment 2a: $F1$ and $F2 < 1$; Experiment 2b: $F1(1, 23) = 3.45$, $MSE = 0.2296$, $p = 0.07$; $F2(1, 15) = 1.10$, $MSE = 0.2296$, $p = 0.31$].

EXPERIMENT 2A (SENTENCES)

Overall, RTs were faster for sentences with high-frequency objects (711 ms) vs. sentences produced with low-frequency objects [732 ms; $F1(1, 46) = 31.69$, $MSE = 212405$, $p < 0.001$; $F2(1, 30) = 24.16$, $MSE = 217997$, $p < 0.001$]. Over the three repetitions, RTs decreased [750, 714, and 702 ms; $F1(2, 46) = 51.45$, $MSE = 344840$, $p < 0.001$; $F2(1, 30) = 44.06$, $MSE = 397556$, $p < 0.001$]. The interaction between frequency and repetition approached significance [$F1(2, 46) = 3.20$, $MSE = 21419$, $p = 0.05$; $F2(2, 30) = 2.77$, $MSE = 24961$, $p = 0.07$], suggesting that the size of the frequency effect changed with the number of repetitions. Jescheniak and Levelt (1994) found no change in the magnitude of the frequency effect over repetitions. However, Griffin and Bock (1998) did find that the frequency effect was reduced over repetitions. This may be due to the number of practice sessions included before the experiment.

Given that the low and high-frequency object sentences significantly differed in the “naturalness” of the action and object (see Experiment 2 Materials) I used these ratings in an analysis of covariance on Experiment 2a RTs. The condition effect (low-frequency vs. high-frequency) remained significant with naturalness ratings included as a covariate [$F1(1, 1982) = 6.86$, $p < 0.01$; $F2(1, 913) = 3.95$, $p = 0.04$]. Thus, the frequency effect in sentence production persists when naturalness of the action–object scenes is statistically controlled.

Lastly I examined the materials for individual items that were poorly designed or named. One action–object scene was judged as highly unnatural: RIDE. The item “ride horse” had an average rating of 6.8 while its match “ride cow” had an average rating of 2.6. The difference (4.2) was more than 2 SDs away from the mean. Also, two items were misnamed at levels at or below 50%: cigarette (50%) and catch (~27%; see **Table A2** in Appendix). When these three items were removed from the analysis, the frequency effect remained significant (p 's < 0.001), and also significant for the analysis of covariance (p 's < 0.01).

Table 2 | Experiments 2a (sentence naming) and 2b (object naming). Mean RTs (ms), SD, and percentage of errors (E%), for high-frequency and low-frequency objects, across three repetitions, for both sentence and object naming.

Noun frequency		Type of utterance					
		Sentence			Object		
		Mean	SD	E (%)	Mean	SD	E (%)
High		711	107	12.0	704	65	4.9
	Repetition 1	733	147	15.3	723	77	7.5
	Repetition 2	705	145	10.4	696	55	3.6
	Repetition 3	697	144	10.4	693	61	3.6
Low		732	111	13.2	724	64	6.9
	Repetition 1	768	143	15.6	743	73	9.8
	Repetition 2	724	149	12.7	716	67	6.7
	Repetition 3	708	158	11.4	712	49	4.1
Overall frequency effect (low–high)		21 *			20 *		

A significant difference of $p < 0.05$ is indicated by an *.

EXPERIMENT 2B (OBJECT ALONE)

Overall, RTs were faster for high-frequency objects (704 ms) in comparison to low-frequency objects [724 ms; $F(1, 46) = 18.47$, $MSE = 210176$, $p < 0.001$; $F(1, 30) = 9.10$, $MSE = 209934$, $p < 0.001$]. Over the three repetitions, RTs decreased [733, 706, and 703 ms; $F(2, 46) = 17.03$, $MSE = 193851$, $p < 0.001$; $F(2, 30) = 10.36$, $MSE = 238950$, $p < 0.001$] although the frequency effect was not affected by repetition as indicated by the lack of interaction between frequency and repetition ($F_s < 1$).

DISCUSSION

Experiment 2 demonstrates that participants are faster to produce sentences with high-frequency direct objects compared to sentences with low-frequency direct objects. This result suggests that the second PW (“the gate”) in utterances like “He opens the gate” is phonologically encoded before articulation. This difference in RTs was not due to name uncertainty as name-agreement ratings were high and similar across materials for verbs and nouns. Nor was the difference due to speed of object recognition as evidenced by the lack of RT difference between high- and low-frequency objects in an object-decision task. Although actions paired with low-frequency objects were rated as more “unnatural” than actions paired with high-frequency objects, when this variable was included as a covariate, the frequency effect remained statistically significant. Together, this pattern of results demonstrates that the difference in RTs to name sentences was due to the lexical frequency of the object at the time of articulation, and not other factors.

Previous results investigating effects of lexical frequency on planning diverge depending on whether utterances varied from trial to trial. Consistent with Experiment 2 Results, when utterances varied from trial to trial, during NP production (e.g., the red car) frequency effects for the second PW of the phonological phrase were observed on response latencies (Alario et al., 2002). However, when similar utterances were produced on all trials such as “the A and the B are above the C,” no effects of frequency on utterance times beyond the first object were observed (Griffin, 2001; Spieler and Griffin, 2006). Participants in these studies may have initiated speech with only the first PW planned as the same words “and the” were used in every trial (cf. Martin et al., 2010). I return to this point in the Section “General Discussion”.

Experiment 2 demonstrates that the lexical frequency of the direct object affects sentence initiation times, and this is the first time phonological facilitation effects have been found for representations following the main verb of a sentence (e.g., more frequent words facilitate sentence initiation). This result suggests that the extent of phonological encoding is not restricted by the verb of the sentence (e.g., Smith and Wheeldon, 2004; Schnur et al., 2006; Oppermann et al., 2010a), but instead is driven by phonological phrase boundaries.

In Experiment 3, to provide converging evidence that phonological planning extends beyond the verb, I used the phonological facilitation effect in the picture–word interference paradigm (as in Experiment 1) and tested whether the direct object NP of a transitive sentence is phonologically encoded before articulation. Phonologically related distractors to the object were displayed in comparison to unrelated distractors.

EXPERIMENT 3

METHOD

Participants

Twenty-eight Harvard University undergraduate students participated. They were paid or received credit for an introductory psychology course. All were native English speakers. None participated in other experiments.

Materials

The same line drawings were used from Experiment 2. Each picture was presented with two distractor words: (a) phonologically related to the noun (e.g., dorm for door) and (b) phonologically unrelated to the noun (e.g., cog for door; see **Table A3** in Appendix). Following Experiment 1, the pictures and the distractors were paired so that each distractor appeared once in the phonologically related and once in the unrelated conditions. This design controlled for unintentional pairing effects between different sets of distractors beyond the phonological relatedness with the nouns. An additional set of distractors was included for another experiment, and treated as filler trials for the purposes of this experiment. Distractors were chosen so they did not sound similar to the agent or verb of the sentence except in three cases. At the beginning of each block, four pictures were included as warm-up trials.

The experimental stimuli were presented in 8 different blocks of 20 trials each for a total of 160 trials (16 target trials and 4 filler trials per block). Each picture was repeated four times, twice for the experimental conditions and twice for the fillers. The trials were randomized as in Experiment 2 and included the following restrictions: (a) the same distractor condition occurred no more than three times in a row; (b) the distractor was not phonologically or semantically related to the distractor in a following trial; (c) the distractor was not phonologically or semantically related to the agent, action, or object in a following trial. Four different block orders were designed and presented to participants according to a Latin-square design.

The distractors were shown in 28-point boldface capital letters in Geneva font, superimposed on the pictures. Pictures were centered at fixation. Word position varied randomly in the region around fixation to prevent participants from systematically fixating the portion of the picture not containing the distractor. However, for an individual picture, the position of all its distractors was the same.

Before the experiment proper, participants had one practice series. Participants were presented with the pictures with a series of X's printed inside each picture. After naming, the correct response was displayed in order to verify the correct name for each picture. Apparatus and procedure were the same as in Experiment 1.

Analyses

Errors were treated as in Experiment 1. F_1 and F_2 statistics were created as in Experiment 1. For the error and principal analyses, two separate ANOVAs were carried out with the following within-subject and within-item variable, Type of distractor (phonologically related vs. unrelated).

RESULTS AND DISCUSSION

Table 3 reports a summary of the data. **Figure 1** shows the magnitude of the phonological facilitation effect to the direct object,

Table 3 | Experiment 3. Mean RTs (ms), SD, and percentage of errors (E%), for phonologically related and unrelated conditions.

Type of distractors	Mean	SD	E (%)
Phonologically related	733	69	12.6%
Phonologically unrelated (related – unrelated)	745	74	11.6%
	–12*		
	[F1 $p < 0.05$; F2 $p < 0.20$]		

A significant difference of $p < 0.05$ is indicated by an *.

with 95% confidence intervals. The item “pull” was removed due to a high amount of error (30%). Error rates consisted of 12.3% before outliers were removed and 13.5% of the data after outliers were eliminated. There were no effects of error rates ($F_s < 1$).

When participants produced sentences, RTs were 12 ms faster in the phonologically related (733 ms) vs. the unrelated condition [745 ms; $F_1(1, 27) = 5.89$, $MSE = 50439$, $p = 0.02$; $F_2(1, 29) = 1.67$, $MSE = 31719$, $p = 0.20$]. However, the effect was only significant in the subject analysis. When I discarded the items named and rated poorly from Experiment 2 (e.g., ride, light, and catch, see **Table A2** in Appendix) the phonological facilitation effect was significant both in the subject and item analyses [17 ms; $F_1(1, 27) = 7.21$, $MSE = 80040$, $p = 0.01$; $F_2(1, 23) = 4.00$, $MSE = 64631$, $p = 0.05$]. This suggests that the phonological facilitation effect in sentence naming is a true effect, showing that the direct object NP was phonologically encoded before articulation.

GENERAL DISCUSSION

These experiments demonstrate that when producing sentences comprised of a single phonological phrase like “He opens the gate” phonological planning extends across the entire phonological phrase, to both the verb and the following direct object NP. Using the picture–word interference paradigm, I found phonological facilitation to the verb (Experiment 1). I found that the direct object was phonologically encoded when participants initiated sentences as the lexical frequency of the object (Experiment 2), and phonological relatedness of distractors (to the object; Experiment 3) affected onset latencies. These results provide the first evidence in spontaneous sentence production from different chronometric measures that planning extends to the phonological phrase boundary where the verb (syntactically important for planning; cf. Ferreira, 2000) does not restrict phonological encoding. Together with previous work, these results are consistent with non-incremental planning during sentence production, and suggest that the phonological phrase determines the extent of phonological planning.

Previous sentence production evidence demonstrates that phonological planning extends up to or includes the verb (Meyer, 1996; Smith and Wheeldon, 2004; Schnur et al., 2006; Oppermann et al., 2010a). This pattern of results is consistent with the theory that phonological encoding begins only when the verb is syntactically encoded (Ferreira, 2000; Ferreira and Swets, 2002). Costa and Caramazza (2002) also suggested that some phonological planning evidence for NPs was consistent with the lexical head of phrase determining the phonological planning boundary. However, the

results presented here rule out this interpretation as phonological planning extended beyond the verb. Evidence from languages that have freer word order is also consistent with this interpretation, as the lexical heads of NPs did not drive the extent of phonological planning for the production of complex NPs either (Miozzo and Caramazza, 1999; Alario and Caramazza, 2002; Costa and Caramazza, 2002).

If the phonological phrase is considered the planning boundary, seemingly contrasting planning results in sentence production are accounted for. The results from Meyer (1996; only the first PW and PP planned), Smith and Wheeldon, 2004; the first phrase comprised of two PWs planned), and Oppermann et al., 2010a; at SOA 0 ms only the first PW and PP planned) are explained if the phonological phrase is the relevant planning unit for phonological encoding. However, because phonological planning can extend beyond the first phonological phrase (Schnur et al., 2006), this result suggests that number of units planned is flexible.

If the phonological phrase is a unit of encoding, then as suggested originally by Levelt (1989) all elements of that unit should be encoded before the next stage of processing. This generates the prediction that if a phonological phrase is planned in advance, then all elements of that phrase should be planned before articulation of that phrase begins. Currently, in sentence production to my knowledge there is no evidence of the partial encoding of phonological phrases (e.g., Meyer, 1996; Smith and Wheeldon, 2004; Schnur et al., 2006; Oppermann et al., 2010a). Depending on how the sentences in Oppermann et al. (2010a) were prosodically produced, if the verb–noun phrase “eats the cheese” was produced as one phonological phrase there should be no partial encoding of the phrase. That is, just as no phonological facilitation was found to the direct object, none should be found for the verb either. Unfortunately, the phonological encoding of verbs was not explored, leaving this an open question.

However, for NP production, Jescheniak et al. (2003) found evidence of partial encoding of the phonological phrase. Using the picture–word interference paradigm, Jescheniak et al. found for SOA 0 ms, phonological facilitation to the first PW when participants produced phrases like “hund” (dog) or “der hund” (the dog) but phonological interference to the third PW for “der große rote hund” (the big red dog; the second PW was never tested). Given that the first two PWs varied only by two (e.g., large/small, and red/blue) phonological planning for these types of phrases may be quite different from more varied sentence production, although this needs to be tested.

An alternative non-incremental view of phonological planning is the retrieval fluency hypothesis (Griffin, 2003; cf., Martin et al., 2010). The retrieval fluency hypothesis suggests that speech initiation is influenced by the ease of phonological encoding of future words. In order to avoid hesitations, if the first word is monosyllabic (or easy to encode) speakers begin speaking only when the next word is phonologically encoded. Thus, that two PWs were planned in utterances like “the red car” (Alario et al., 2002; Costa and Caramazza, 2002; Damian and Dumay, 2007) or “The girl walks” (Schnur et al., 2006) may be due to the ease of encodability of the first PW across utterances. In these cases, the first PW was one of three or four colors, or one of four actors, which allowed for encoding of the second PW before articulation. Retrieval fluency

is also achieved with less planned in advance (e.g., a PW) if utterance content is predicted from trial to trial, because encoding of repeated material is more automatic, allowing for phonological encoding of future words after articulation begins (Martin et al., 2010). Thus, evidence of only the first PW being encoded (Meyer, 1996; Griffin, 2001; Spieler and Griffin, 2006) may be explained by the repetition of utterance format across trials (e.g., Meyer, 1996: “the A and the B”; Griffin, 2001; Spieler and Griffin, 2006: “the X and the Y are above the Z”). Critically, however, Experiments 1–3 cannot be explained by retrieval fluency, given that the first PW (e.g., “He opens”) and second PW (“the gate”) always varied from trial to trial.

A parsimonious explanation of the spontaneous sentence production results here and elsewhere is that phonological planning extends over an entire phonological phrase. However, planning of a phrase does not rule out the role of other prosodic units in the production of sentences, nor planning flexibility. Planning of larger prosodic units like intonational phrases may be required in order to account for certain prosodic phenomena. Ohala (1978) reviewed evidence where in the Japanese and Hausa languages, changes in pitch are inversely proportional to the length of the utterance. In Bengali, in order to conversationally focus a particular constituent, a high tone is assigned at the beginning of the utterance depending on where the end of the phrase occurs (Lahiri, 2000). Evidence from reading out-loud (e.g., Gee and Grosjean, 1983; Breen et al., 2010) and producing memorized sentences (Ferreira, 2007) suggests that prosodic planning operates over phonological and intonational phrases, but is flexible, affected by multiple variables. The questions to be addressed are whether these results are relevant for spontaneous speech, and exactly how to examine the planning of more complicated sentences in spontaneous speech.

What advantage does a wider extent of planning offer to sentence production in comparison to incremental planning (e.g., a PW)? Prosodic structure may help the listener understand a speaker’s message (although this point is debated, cf., Ferreira, 1993, 2007). For example, if phonological phrases are interrupted,

recognition of words is impaired (Fox Tree, 1995). Likewise, phonological phrase boundaries help the listener understand ambiguous syntactic phrases like “I like cooking apples” where the word “cooking” is ambiguous as a verb or modifier of the word “apples” (Schafer et al., 2000; Wheeldon, 2000). Because phrases correspond to information like who is involved in an action (e.g., the subject NP) or the action and what is acted upon (e.g., a VP), if phrases can be conveyed smoothly (e.g., without error) to a listener this may facilitate speech comprehension. The phonological phrase may be a preferred planning unit over larger amounts because planning of larger amounts may induce hesitations in speech. Long hesitations can convey undesirable consequences: the speaker may convey that she is “slow-witted” or has poor social skills (Fox Tree, 2000). Furthermore, during conversation, being able to hold the floor is an important aspect of the discourse process and hesitations signal to others to interrupt (Fox Tree, 2000). The ultimate goal of research in speech planning should be to understand how much is planned when more than one speaker is involved.

The results reported here and elsewhere suggest that all elements of a phonological phrase are phonologically encoded whether the elements in the phrase correspond to a simple NP (Meyer, 1996; Oppermann et al., 2010a), complex NP (Miozzo and Caramazza, 1999; Alario and Caramazza, 2002; Costa and Caramazza, 2002; Smith and Wheeldon, 2004; Damian and Dumay, 2007), VP (Schnur et al., 2006), or sentence (Experiments 1–3). These results suggest that the phonological phrase is a unit of phonological encoding, where the phonological representations of all elements are retrieved before phrase articulation begins. However, which variables influence the extent of planning is an open question for future research.

ACKNOWLEDGMENTS

Parts of these results were presented at the 45th Annual Meeting of the Psychonomic Society, Minneapolis, November 18–20, 2004 and served as partial fulfillment of the requirements for the PhD degree at Harvard University for Tatiana T. Schnur.

REFERENCES

- Alario, F. X., and Caramazza, A. (2002). The production of determiners: evidence from French. *Cognition* 28, 179–223.
- Alario, F. X., Costa, A., and Caramazza, A. (2002). Frequency effects in noun phrase production: implications for models of lexical access. *Lang. Cogn. Process.* 17, 299–319.
- Almeida, J., Knobel, M., Finkbeiner, M., and Caramazza, A. (2007). The locus of the frequency effect in picture naming: when recognizing is not enough. *Psychon. Bull. Rev.* 14, 1177–1182.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1998). *The CELEX Lexical Database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Barry, C., Morrison, C. M., and Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: effects of age of acquisition, frequency, and name agreement. *Q. J. Exp. Psychol. A* 50, 560–585.
- Bloem, I., Van Den Boogaard, S., and La Heij, W. (2004). Semantic facilitation and semantic interference in language production: further evidence for the conceptual selection model of lexical access. *J. Mem. Lang.* 51, 307–323.
- Bock, J. K. (1987). “Coordinating words and syntax in speech plans,” in *Progress in the Psychology of Language*, ed. A. Ellis (London: Erlbaum), 337–390.
- Breen, M., Watson, D. G., and Gibson, E. (2010). Intonational phrasing is constrained by meaning, not balance. *Lang. Cogn. Process.* 1–31.
- Caramazza, A., Bi, Y. C., Costa, A., and Miozzo, M. (2004). What determines the speed of lexical access: homophone or specific-word frequency? A reply to Jescheniak et al. (2003). *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 278–282.
- Caramazza, A., Costa, A., Miozzo, M., and Bi, Y. C. (2001). The specific-word frequency effect: implications for the representation of homophones in speech production. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1430–1450.
- Catrambone, R. (1998). The subgoal learning model: creating better examples so that students can solve novel problems. *J. Exp. Psychol. Gen.* 127, 355–376.
- Costa, A., and Caramazza, A. (2002). The production of noun phrases in English and Spanish: implications for the scope of phonological encoding in speech production. *J. Mem. Lang.* 46, 178–198.
- Cycowicz, Y. M., Friedman, D., Rothstein, M., and Snodgrass, J. G. (1997). Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *J. Exp. Child. Psychol.* 65, 171–237.
- Damian, M. F., and Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *J. Mem. Lang.* 57, 195–209.
- Ellis, A. W., and Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 515–523.

- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *J. Mem. Lang.* 30, 210–233.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychol. Rev.* 100, 233–253.
- Ferreira, F. (2000). “Syntax in language production: an approach using tree-adjoining grammars,” in *Aspects of language production. Studies in Cognition Series*, ed. L. Wheeldon (Philadelphia, PA: Psychology Press/Taylor & Francis), 291–330.
- Ferreira, F. (2007). Prosody and performance in language production. *Lang. Cogn. Process.* 22:8, 1151–1177.
- Ferreira, F., and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* 46, 57–84.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *J. Mem. Lang.* 34, 709–738.
- Fox Tree, J. E. (2000). “Coordinating spontaneous talk,” in *Aspects of Language Production*, ed. L. Wheeldon (East Sussex: Psychology Press).
- Francis, N. W., and Kucera, H. (1982). *Frequency Analysis of English Usage*. Boston, MA: Houghton Mifflin Company.
- Garrett, M. F. (1975). “The analysis of sentence production,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 9, ed. G. Bower (New York: Academic Press), 133–177.
- Gee, J. P., and Grosjean, F. (1983). Performance structures: a psycholinguistic and linguistic appraisal. *Cogn. Psychol.* 15, 411–458.
- Griffin, Z. (2000). How early are verbs selected in extemporaneous English sentence production? in *Architectures and mechanisms of Language Processes conference*, Leiden.
- 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. *Psychon. Bull. Rev.* 10, 603–609.
- Griffin, Z. M., and Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *J. Mem. Lang.* 38, 313–338.
- Humphreys, K. R., Boyd, C. H., and Watter, S. (2010). Phonological facilitation from pictures in a word association task: evidence for routine cascaded processing in spoken word production. *Q. J. Exp. Psychol.* 63, 2289–2296.
- Jescheniak, J. D., and Levelt, W. J. M. (1994). Word-frequency effects in speech production – retrieval of syntactic information and of phonological form. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 824–843.
- Jescheniak, J. D., Oppermann, F., Hantsch, A., Wagner, V., Madebach, A., and Schriefers, H. (2009). Do perceived context pictures automatically activate their phonological code? *Exp. Psychol.* 56, 56–65.
- Jescheniak, J. D., Schriefers, H., and Hantsch, A. (2003). Utterance format affects phonological priming in the picture-word task: implications for models of phonological encoding in speech production. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 441–454.
- Kittredge, A. K., Dell, G. S., Verkuilen, J., and Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cogn. Neuropsychol.* 25, 463–492.
- Knobel, M., Finkbeiner, M., and Caramazza, A. (2008). The many places of frequency: evidence for a novel locus of the lexical frequency effect in word production. *Cogn. Neuropsychol.* 25, 256–286.
- Kroll, J. F., and Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *J. Verbal Learn. Verbal Behav.* 23, 39–66.
- Lahiri, A. (2000). “Phonology: Structure, representation, and process” in *Aspects of Language Production. Studies in Cognition Series*, ed. L. Wheeldon (Philadelphia: Psychology Press/Taylor & Francis), 165–225.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., and Salmelin, R. (1998). An MEG study of picture naming. *J. Cogn. Neurosci.* 10, 553–567.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75.
- Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture-word interference. *Can. J. Psychol.* 36, 349–367.
- Malpass, D., and Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: evidence from extrafoveal-on-foveal effects. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 523–537.
- Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., and Yang, C. L. (2010). Planning in sentence production: evidence for the phrase as a default planning scope. *Cognition* 116, 177–192.
- Masterson, J., and Druks, J. (1998). Description of a set of 164 nouns and 102 verbs matched for printed word and frequency, familiarity and age-of-acquisition. *J. Neurolinguistics* 11, 331–354.
- Meyer, A. S. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 1146–1160.
- Meyer, A. S. (1996). Lexical access in phrase and sentence production: results from picture-word interference experiments. *J. Mem. Lang.* 35, 477–496.
- Meyer, A. S., and Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 1146–1160.
- Meyer, A. S., Sleiderink, A. M., and Levelt, W. J. M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition* 66, B25–B33.
- Miozzo, M., and Caramazza, A. (1999). The selection of determiners in noun phrase production. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 907–922.
- Morsella, E., and Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 555–563.
- Navarrete, E., and Costa, A. (2005). Phonological activation of ignored pictures: further evidence for a cascade model of lexical access. *J. Mem. Lang.* 53, 359–377.
- Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. The Netherlands: Foris.
- Ohala, J. J. (1978). “Production of tone,” in *Tone: A Linguistic Survey*, ed. V. Fromkin (New York: Academic Press), 5–39.
- Oldfield, R. C., and Wingfield, A. (1965). Response latencies in naming objects. *Q. J. Exp. Psychol.* 17, 273–281.
- Oppermann, F., Jescheniak, J. D., and Schriefers, H. (2008). Conceptual coherence affects phonological activation of context objects during object naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 587–601.
- Oppermann, F., Jescheniak, J. D., and Schriefers, H. (2010a). Phonological advance planning in sentence production. *J. Mem. Lang.* 63, 526–540.
- Oppermann, F., Jescheniak, J. D., Schriefers, H., and Gorges, F. (2010b). Semantic relatedness among objects promotes the activation of multiple phonological codes during object naming. *Q. J. Exp. Psychol.* 63, 356–370.
- Rayner, K., and Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture-word interference task. *Br. J. Exp. Psychol.* 77, 207–222.
- Roelofs, A. (2002). Spoken language planning and the initiation of articulation. *Q. J. Exp. Psychol.* 55, 465–483.
- Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 353–368.
- Rosenbaum, D. A. (2010). *Human Motor Control*. San Diego, CA: Academic Press/Elsevier.
- Schafer, A. J., Speer, S. R., Warren, P., and White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *J. Psycholinguist. Res.* 29, 169–182.
- Schnur, T. T., Costa, A., and Caramazza, A. (2006). Planning at the phonological level during sentence production. *J. Psycholinguist. Res.* 35, 189–213.
- Selkirk, E. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology* 3, 371–405.
- Selkirk, E. (1996). “The prosodic structure of function words,” in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. K. D. J. L. Morgan (Mahwah, NJ: Lawrence Erlbaum Associates), 187–213.
- Smith, M., and Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 675–686.
- Snodgrass, J. G., and Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behav. Res. Methods Instrum. Comput.* 28, 516–536.
- Spieler, D. H., and Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multi-word utterances. *Lang. Cogn. Process.* 21, 291–321.

- Starreveld, P. A. (2000). On the interpretation of onsets of auditory context effects in word production. *J. Mem. Lang.* 42, 497–525.
- Starreveld, P. A., and La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 686–698.
- Starreveld, P. A., and La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 896–918.
- Stemberger, J. P., and McWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Mem. Cognit.* 14, 17–26.
- Strijkers, K., Costa, A., and Thierry, G. (2009). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cereb. Cortex* 20, 912–928.
- Wheeldon, L. (2000). “Generating prosodic structure,” in *Aspects of Language Production*, ed. L. Wheeldon (Philadelphia: Psychology Press/Taylor & Francis), 249–274.
- Wheeldon, L., and Lahiri, A. (1997). Prosodic units in speech production. *J. Mem. Lang.* 37, 356–381.
- Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *Am. J. Psychol.* 81, 226–234.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2011; paper pending published: 28 September 2011; accepted:

19 October 2011; published online: 04 November 2011.

Citation: Schnur TT (2011) Phonological planning during sentence production: beyond the verb. *Front. Psychology* 2:319. doi: 10.3389/fpsyg.2011.00319

This article was submitted to *Frontiers in Language Sciences*, a specialty of *Frontiers in Psychology*.

Copyright © 2011 Schnur. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

APPENDIX

Table A1 | Stimuli for Experiment 1. Frequency (FREQ), letter length (LTTR), and number of syllables (SYLL) of the written distractors paired with the target pictures.

PICTURES			Phonologically related			Unrelated				
AGENT	VERB	OBJECT	FREQ	LTTR	SYLL	FREQ	LTTR	SYLL		
He	CARRY	Pumpkin	Care	108	4	1	Watch	209	5	1
She	CATCH	Fly	Cat	42	3	1	Ticket	30	6	2
She	CLIMB	Table	Client	33	6	2	Peer	32	4	1
She	CUT	String	Cup	58	3	1	Maid	44	4	1
She	DRAW	Bird	Drawl	4	5	1	Stirrup	2	7	2
She	DROP	Glass	Dross	4	5	1	Pulpit	5	6	2
He	EAT	Sandwich	Eagle	12	5	2	Opium	16	5	3
He	IRON	Shirt	Ire	1	3	2	Kin	1	3	1
She	KICK	Drum	Kid	104	3	1	Rifle	87	5	2
She	KISS	Horse	Kin	1	3	1	Ire	1	3	2
She	KNIT	Sweater	Nip	2	3	1	Drawl	4	5	1
She	LIGHT	Candle	Lice	2	4	1	Pudding	1	7	2
She	MAIL	Letter	Maid	44	4	1	Cup	58	3	1
He	OPEN	Door	Opium	16	5	3	Eagle	12	5	2
He	PAINT	Wall	Pail	8	4	2	Rinse	6	5	1
He	PEEL	Orange	Peer	32	4	1	Cat	42	3	1
She	PET	Cat	Pestle	1	6	2	Smelt	1	5	1
He	PLANT	Tree	Plasma	24	6	2	Client	33	6	2
He	PULL	Elephant	Pulpit	5	6	2	Dross	4	5	1
He	PUSH	Bus	Pudding	1	7	2	Stocky	2	6	2
She	READ	Book	Reach	324	5	1	Care	108	4	1
He	RIDE	Cow	Rifle	87	5	2	Kid	104	3	1
He	RING	Bell	Rinse	6	5	1	Pail	8	4	2
She	SMELL	Flower	Smelt	1	5	1	Pestle	1	6	2
She	STIR	Pot	Stirrup	2	7	2	Lice	2	4	1
He	STOP	Car	Stocky	2	6	2	Nip	2	3	1
He	TICKLE	Dog	Ticket	30	6	2	Plasma	24	6	2
He	WATER	Plant	Watch	209	5	1	Reach	324	5	1

Table A2 | Stimuli for Experiment 2. Frequency (FREQ), number of syllables (SYLL), noun-naming agreement rating (NNA %), verb-naming agreement rating (VNA %), and naturalness rating (NAT).

PICTURES													
AGENT	VERB	HIGH-FREQUENCY OBJECTS						LOW-FREQUENCY OBJECTS					
		OBJECT	FREQ	SYLL	NNA (%)	VNA (%)	NAT	OBJECT	FREQ	SYLL	NNA (%)	VNA (%)	NAT
He	CARRY	Chair	42	1	100	86	1.1	Pumpkin	11	1	100	82	5.1
She	CATCH	Ball	40	1	100	32	0.6	Bee	3	3	95	23	2.9
She	CUT	Star	81	1	100	100	1.6	Flag	6	2	91	100	2.6
She	DRAW	Heart	59	1	95	95	1.3	Flower	1	2	36	86	5.0
She	DROP	Glass	79	1	95	100	1.1	Spoon	7	1	100	100	5.0
He	EAT	Bread	34	1	91	100	0.8	Cheese	8	2	100	100	6.5
She	KICK	Bottle	41	2	81	100	1.7	Drum	24	2	100	95	2.5
She	LIGHT	Cigarette	35	1	50	100	1.6	Candle	4	1	100	100	6.4
He	OPEN	Door	200	2	100	95	0.5	Gate	7	1	5	100	6.6
She	PET	Cat	103	1	95	100	0.8	Rabbit	18	1	81	95	4.8
He	PLANT	Tree	30	2	91	100	0.9	Cactus	5	1	100	100	3.3
He	PULL	Table	269	1	91	91	1.4	Elephant	12	1	100	82	1.6
He	PUSH	Box	91	1	100	95	1.0	Dresser	27	2	68	100	4.7
He	RIDE	Horse	40	1	100	100	0.4	Cow	25	2	100	100	2.6
He	STOP	Car	81	1	100	91	1.1	Truck	14	3	95	100	4.8
She	TICKLE	Dog	107	1	81	95	1.1	Pig	28	1	100	91	2.0

Table A3 | Stimuli for Experiment 3. Frequency (FREQ), letter length (LTTR), and number of syllables (SYLL) of the written distractors paired with the target pictures.

PICTURES			Phonologically related			Unrelated				
AGENT	VERB	OBJECT	FREQ	LTTR	SYLL	FREQ	LTTR	SYLL		
He	CARRY	Chair	Chain	48	5	1	Track	55	5	1
He	CARRY	Pumpkin	Pummel	0	6	2	Flounder	4	8	2
She	CATCH	Ball	Bald	8	4	1	Dot	9	3	1
She	CATCH	Bee	Bead	11	4	1	Tangle	9	6	2
She	CUT	Star	Stab	8	4	1	Bog	2	8	3
She	CUT	Flag	Flat	79	4	1	Glad	64	4	1
She	DRAW	Heart	Hard	207	4	1	Treat	98	5	1
She	DRAW	Flower	Flounder	4	8	2	Pummel	0	6	2
She	DROP	Glass	Glad	64	4	1	Flat	79	4	1
She	DROP	Spoon	Spook	1	5	1	Cad	1	3	1
He	EAT	Bread	Breath	59	6	1	Cheap	63	5	1
He	EAT	Cheese	Cheap	63	5	1	Breath	59	6	1
She	KICK	Bottle	Bottom	85	6	2	Calm	91	4	1
She	KICK	Drum	Drug	47	4	1	Horn	18	4	1
She	LIGHT	Cigarette	Cinema	22	6	3	Elegant	25	7	3
She	LIGHT	Candle	Candid	2	6	2	Rabble	1	6	2
He	OPEN	Door	Dorm	5	4	1	Cog	2	3	1
He	OPEN	Gate	Gape	7	4	1	Peg	7	3	1
She	PET	Cat	Cad	1	3	1	Spook	1	5	1
She	PET	Rabbit	Rabble	1	6	2	Candid	2	6	2
He	PLANT	Tree	Treat	98	5	1	Hard	207	4	1
He	PLANT	Cactus	Cackle	1	6	2	Draper	1	6	2
He	PULL	Table	Tangle	9	6	2	Bead	11	4	1
He	PULL	Elephant	Elegant	25	7	3	Cinema	22	6	3
He	RIDE	Horse	Horn	18	4	1	Drug	47	4	1
He	RIDE	Cow	Cog	2	3	1	Dorm	5	4	1
He	STOP	Car	Calm	91	4	1	Bottom	85	6	2
He	STOP	Truck	Track	55	5	1	Chain	48	5	1
She	TICKLE	Dog	Dot	9	3	1	Bald	8	4	1
She	TICKLE	Pig	Peg	7	3	1	Gape	7	4	1