Compounds, phrases and clitics in connected speech

Hilary S.Z. Wynne [a], Linda Wheeldon [b], Aditi Lahiri [a]

[a] Faculty of Linguistics, Philology & Phonetics,
University of Oxford, Walton Street,
Oxford OX1 2HG, UK

[b] Institute of Foreign Languages and Translation,
University of Agder, Norway

ABSTRACT

Four language production experiments examine how English speakers plan compound words during phonological encoding. The experiments tested production latencies in both delayed and online tasks for English noun-noun compounds (e.g., *daytime*), adjective-noun phrases (e.g., *dark time*), and monomorphemic words (e.g., *denim*). In delayed production, speech onset latencies reflect the total number of prosodic units in the target sentence. In online production, speech latencies reflect the size of the first prosodic unit. Compounds are metrically similar to adjective-noun phrases as they contain two lexical and two prosodic words. However, in Experiments 1 and 2, native English speakers treated the compounds as single prosodic units, indistinguishable from simple words, with RT data statistically different than that of the adjective-noun phrases. Experiments 3 and 4 demonstrate that compounds are also treated as single prosodic units in utterances containing clitics (e.g., *dishcloths are clean*) as they incorporate the verb into a single phonological word (i.e. *dishcloths-are*). Taken together, these results suggest that English compounds are planned as single recursive prosodic units. Our data require an adaptation of the classic model of phonological encoding to incorporate a distinction between lexical and postlexical prosodic processes, such that lexical boundaries have consequences for post-lexical phonological encoding.

Keywords: language production, prosodic structure, phonological encoding, compounds and adjective-noun phrases, compounds and clitics

Highlights:

- First behavioural investigation of the phonological encoding of English compounds and cliticisation.
- Compounds pattern with simple words rather than phrases, both in isolation and in connected speech.
- Metrical information alone is not sufficient for determining the units of prosodification.
- We propose an adaptation of the classic model of phonological encoding to incorporate a distinction between lexical and postlexical prosodic processes.

1. INTRODUCTION

1.1 *Psychological processes involved with speaking*

Language production models agree that there are a series of cognitive stages involved in the production of speech. Each of these stages prepares corresponding representations, e.g. concepts at the semantic level, lemmas and syntactic structure during grammatical encoding, and phonological representations during phonological encoding (see Griffin & Ferreira, 2006 for a review). Numerous psycholinguistic studies have shown that the phonological encoding stage is responsible for infusing abstract lexical representations with phonological properties such as segmental ordering (e.g., Meyer 1990, 1991; Wheeldon & Levelt 1995; Wheeldon & Morgan 2002), syllabification (e.g., Ferrand et al. 1996, 1997; Morgan & Wheeldon 2003; Schiller et al. 2002), and prosodification (e.g., Ferreira 1993; Wheeldon & Lahiri 1997; Roelofs & Meyer 1998; Wheeldon & Lahiri 2002; Jescheniak et al. 2003; Cholin et al. 2004; Damian & Dumay 2007).

In early models of language production (Harley 1984, Dell 1986) the output of the phonological encoding stage was typically treated as a lexical word. In order to prepare an utterance for articulation, the phonological encoding stage accessed the output of the lexicalization stage (lexical words) and systematically built metrical frames according to the specific phonological rules of the language. A lexical word is a well-formed semantic and syntactic unit that can stand on its own, be uttered in isolation, and even be considered a full utterance. For example, the lexical word *coffee*, if uttered with a question intonation, can be construed as someone asking the listener if they would like coffee. Lexical words can be morphologically complex; e.g. *use, uses, disuse, using, user* are all lexical words each with well-defined semantic and syntactic properties. But we know that speakers do not produce connected speech in isolated units.

Multiword utterances regularly exhibit word boundaries that are not necessarily sacrosanct, as illustrated by the famous English slogan *Drinka pinta milka day*. The linguistic units in this structure correspond not to the syntactic representation of the utterance (*Drink a pint of milk a day*) but to how it sounds in regular, connected speech, i.e. its prosodic structure. The mismatch in prosodic and lexical structure points to a process in phonological encoding where features of connected speech must be prepared: that is, where *drink a* becomes *drinka*. Accordingly, more recent psycholinguistic models of phonological encoding (Roelofs 1997; Levelt, Roelofs & Meyer, 1999) no longer treat the output of the phonological encoding process as a lexical unit, but as a prosodic one.

In this paper, our focus is on the prosodic structure of two types of multiword utterances in native English speakers: compounds and phrases. We investigate two main hypotheses related to the preparation of prosodic units in English: one regarding the similarities or differences in prosodic structure of compounds and phrases, and one related to the behaviour of clitics with these items in connected speech. The first hypothesis is concerned with the prosodic structure of multi-word structures in English. Although multi-word sequences may look similar on the surface, they can be quite different depending on their prosodic structure. Linguistic theories of phonological phrasing maintain that prosodic units are not necessarily isomorphic with syntactic units (Selkirk 1980, 1986; Lahiri & Plank 2010). These theories approach the issue of asymmetry by introducing a series of hierarchically-structured prosodic units such as phonological words and phrases.

Here, compounds prove to be rather remarkable. Both the compound *White House* (the residence of the American president) and the phrase *white house* (a house painted white) contain the same number of lexical units and the same number of prosodic units; however, when spoken in regular speech they differ in stress placement suggesting that compounds and phrases differ in

structure at the prosodic level (where phonological features such as stress and intonation are assigned). Our questions are thus: are compounds and phrases processed differently during the phonological planning stage in English, and if so- to what extent does this processing reflect the difference in structure? If compounds and phrases are treated differently during the phonological planning stage, then this difference should be reflected in the time it takes the speaker to plan the utterance in which they are contained.

Our second hypothesis hinges upon the first. If English compounds and phrases are indeed treated as different prosodic structures by speakers, then this should also be observable in the behaviour of phonological clitics with these structures. Prosodic structure theory maintains that phonological clitics, such as function words such as *a* in *drink a* in the English slogan above, reduce and attach to prosodic units: this process is known as cliticisation. If the structural differences in compounds and phrases are predicted by the surface metrical stress, then such differences would also be predicted due to cliticisation.

In what follows, we first turn to the literature regarding the unit of planning during phonological encoding involving single lexical words. We then move to a discussion of the psycholinguistic evidence regarding multi-word units such as compounds, and phrases. Then we will focus on our two hypotheses as related to the current approaches.

1.2 *The Unit of Planning: lexical words and clitics*

When a speaker is planning a sequence of two words, they plan the conceptual, syntactic and phonological structure of the utterance. Evidence has accumulated from both linguistic and psycholinguistic research that prosodic rather than lexical structure governs phonological encoding processes in language production (Hannahs 1995, Wheeldon & Lahiri 1997, 2002; Vigario 2003).

Following Levelt et al. (1999), we maintain that prosodic units known as "phonological words" are built during phonological encoding. Minimally composed of at least one stressed foot, a phonological word can contain a lexical word plus any number of unstressed items, which are most often function words such as auxiliaries (is, are etc.), determiners (a, the etc.), pronouns (it, he etc.) and prepositions (to, in etc.). Following this, the number of lexical words is often different than the number of phonological words in a multi-word utterance:

(1) **Examples of lexical and phonological (ω) word formation**[1]

    (a) lexical units: $[Tim]_N$ $[is]_V$ $[sick]_{Adj}$

        prosodic units: $(Timz)_\omega$ $(sick)_\omega$

    (b) lexical units: $[Drink]_V$ $[the]_{Art}$ $[juice]_N$

        prosodic units: $(drɪŋkðə)_\omega$ $(juice)_\omega$

In the examples above (1a and 1b), we can observe how unstressed items can attach to lexical words in normal connected speech, forming a single phonological word. While the sentence *Tim is sick* (1a) is made up of three lexical words, it is only two phonological words when spoken in regular, connected speech; through reduction, *Tim is* becomes [*tɪmz*]. Similarly, in *drink the juice,* the unstressed unit *the* reduces and attaches to *drink*, forming a single phonological word: *[drɪŋkðə]*. These unstressed items are known as "phonological clitics" and the process that forms them is known as cliticisation. Further examples of this behaviour can be seen below, which reproduces the English slogan (repeated here with appropriate syntactic and prosodic bracketing):

---

[1] We use the ω symbol to notate prosodic unit (phonological word) boundaries and φ for phrase boundaries consistently throughout this work.

(2) **syntactic phrasing:**

    (a) [DRINK] [a PINT] [of MILK] [a DAY]

    (b) [ALE] [and PIE]

(3) **prosodic grouping:**

    (a) (Drink a)$_\omega$ (pint of)$_\omega$ (milk a)$_\omega$ (day)$_\omega$

       [drɪŋkə]$_\omega$ [paɪntə]$_\omega$ [mɪlkə]$_\omega$ [deɪ]$_\omega$

    (b) (Ale and)$_\omega$ (pie)$_\omega$

       [eɪlən]$_\omega$ [paɪ]$_\omega$

As mentioned above, cliticisation often involves function words such as auxiliaries, prepositions, pronouns, conjunctions, and articles. Examples of encliticisation, where weak forms undergo phonological change (e.g. colloquial German *denn* > *'n*), have regularly stood as evidence for default phonological phrasing which often violates the morphosyntactic structure of the utterance (cf. Sweet 1886; Saran 1907; Zwicky & Pullum 1983). Note that the boundaries of the prosodic units (example 3) are different than those of the morphosyntactic units (example 2) and the function words *of, a, and* encliticise to the preceding word. Thus, *pint of* in the example above is encoded as [*paɪn.tə*], a single prosodic unit in which the lexical and syllable boundaries do not coincide (cf. Selkirk 1980; Nespor & Vogel 1986; Selkirk, 1996; Peperkamp, 1997; Lahiri & Plank 2010).

Further supportive evidence for this process comes from psycholinguistic tasks employing the prepared speech paradigm, (Sternberg et al. 1978; Wheeldon & Lahiri 1997, 2002). In a series of experiments using a delayed priming task, Wheeldon and Lahiri (1997) elicited reaction time

data for utterances containing clitics in Dutch. In a series of prepared speech tasks, speakers were asked to respond to questions such as *Wat zoek je?* ("What do you seek?"). The authors found that the onset latencies for sentences containing clitics (e.g. *Ik zoek het water*, "I drink the water") were no different than those that contained none (e.g. *Ik zoek water*, "I seek water"). This indicated that clitics were attaching to the neighbouring word and forming a single prosodic unit, and that the number of prosodic words in both clitic and non-clitic sentences were the same. In contrast, sentences containing stressed elements that cannot encliticise (e.g. *Ik zoek vers water*, "I seek fresh water") had significantly longer onset latencies. This pattern held even if the first word was a full prosodic word (e.g. *Riet zoekt het water*, "Riet seeks the water" vs. *Riet zoekt water* "Riet seeks water"). The effect was also not attributable to the number of content words produced, as phrase final function words, which receive stress (e.g. *Ik zoek het*, "I seek *it*"), did not significantly differ in onset latency from the same phrases with a final content word (e.g. *Ik zoek water*, "I seek water").

Direction of cliticisation was confirmed by means of an online production task on the assumption that the size of the first prosodic unit would predict preparation time. Thus, in *Ik drink het water*, if *het* was in fact encliticising leftwards (*drinkhet)*, then the sentence containing the clitic would yield the longest latencies as the speaker would have to wait for the entire prosodic unit (*ikdrinkhet*) to be encoded before speaking. If, on the other hand, *het* encliticised rightwards, the preparation times would be similar to those for the non-clitic condition (*ik drink water*) since the speaker would only be encoding [*ikdrink*] as the first prosodic unit. As predicted, clitic sentences took significantly longer to produce in this task, confirming that the Dutch determiner *het* ("the") encliticised with the neighbouring word and formed a single prosodic word, resulting in similar naming latencies to sentences that contained no determiner.

In summary, Wheeldon and Lahiri (1997) found that neither the number of lexical words, content words, nor syllables in an utterance predicted naming latencies in clitic sentences in Dutch. Clitics attached leftwards to the initial phonological unit [*Ik zoek*], forming a single phonological word [*Ik zoek het*], which resulted in similar naming latencies to other sentences that contained the same number of phonological words, even if they contained fewer lexical or content words. Cliticisation was retained in the online task results, with speakers preferring to delay articulation before the initial phonological word [*Ik zoek het*] was ready. Crucially, the structures that predicted the naming latencies were always phonological in shape, regardless of the type of task (online or delayed).

In this section, we discussed the planning of simple lexical words along with encliticisation of weakly stressed function word. Theoretical and experimental literature provide evidence that encliticisation leads to the creation of phonological words which may result in mismatch between syntactic and phonological structure (e.g. *drinka pint* vs. *drink a pint*). Phonological encoding evidence suggests that the encliticised phonological word is treated as a single lexical word in planning. The question now arises as to the way in which multiword utterances such as compounds and phrases are treated in production planning.

1.3 *Multiword utterances: compounds and phrases*

Both compounds and phrases have received a great deal of attention (cf. Chomsky & Halle, 1968; Ladd,1984; Cinque, 1993; Fabb, 1998; Badecker, 2001; Booij, 2005; Fiorentino & Poeppel, 2007). Compounds contain two (or more) lexical words (e.g. blackboard), which means they also contain two phonological words. When treated as individual syntactic units, black and board are separate phonological units with their own stress assignment. When compounded, however, the

9

units become one phonological word (bláckbòard) with a main stress falling on the first unit and a secondary stress on the second. In essence, the two independent phonological words recursively form a single phonological word. The following examples clarify the distinction between a compound and a phrase.

(4) **structure of the compound *blackboard*** (a board used to write on with chalk):

    (4a)   **syntactic structure:** $[[black]_A [board]_N]_N$

    (4b)   **phrase stress:**   X

             p-word stress: X   x

             prosodic structure: $((bláck)_\omega (boàrd)_\omega)_\omega$

(5) **structure of the adjective-noun phrase *black board*** (a piece of wood painted black):

    (5a)   **syntactic structure:** $[[black]_{Adj} [board]_N]_{AP}$

    (5b)   **phrase stress:**   X

             p-word stress: X   X

             prosodic structure: $((blàck)_\omega (boárd)_\omega)_\phi$

The structures in (4) and (5) reveal that compounds and phrases are similar on one level but are quite different on others. For instance, both structures include independent lexical words, which, by definition, are also phonological words. However, a compound is constructed by recursive word formation and on the surface it is a single phonological word, while an adjectival phrase constitutes two separate phonological words that combine to form a phonological phrase. In terms of metrical prominence, the main stress for a compound is on the initial phonological

word (Example 4b) while for a phrase it is on the second element (Example 5b) (cf. Liberman & Prince 1979). Because the second syllable in a compound still carries stress, it cannot be considered prosodically-deficient like the determiners and auxiliaries we saw in examples (2) and (3). Note also that the syntactic head is the same for both, viz. *board.*

In addition to theoretical analyses of compounds and phrases, psycholinguistic studies have also addressed these issues. Much of the existing literature on the production of multiword structures focuses on the minimal planning unit (MPU), that is, the minimal unit that needs to be encoded before the speaker begins articulation. Studies are divided over advance planning and the minimal planning unit: Schriefers and Teruel (1999) found that naming latencies were shorter for phrases only when a distractor word was related to the first word of a phrase in a picture-word interference task. Conversely, Costa and Caramazza (2002) found that native Spanish speakers, when producing phrases containing post-nominal adjectives (e.g. *la casa azul*, "the blue house"), exhibited phonological facilitation of the adjective despite its post-position. Likewise, Jescheniak et al. (2003) found that the presentation of phonological distractor words related only to the noun caused inhibition in German adjective noun phrases (e.g. *der rote Tische*, "the red table"). Janssen and Barber (2012) argued that results from two picture-naming and interference tasks indicated that naming latencies were affected only by the frequency of the entire phrase ("token phrase frequency", which they drew from number of hits on Google), rather than individual word frequency. More recently, Lange and Laganaro (2014) elicited evidence that only the first word of a French adjective-noun phrase (e.g. *grand chat*, "large cat") was primed by a phonologically-related distractor word during picture-word interference; no effect was found when the second word in the phrase was primed. In an ERP study using picture naming, Lange et al. (2015) found that response latencies were significantly longer for adjectival phrases than simple nouns. Thus,

11

the evidence for the phonological encoding of noun phrases and compounds is mixed in distractor tasks.

Simple naming tasks have elicited more uniform results for the production of adjective-noun phrases: in Wheeldon and Lahiri (2002), results from a prepared speech production task indicated that Dutch speakers showed sensitivity in naming latencies when the prosodic unit changed size, rather than the lexical or syllabic unit. Noun-noun compounds (e.g. *ooglid*, 'eyelid') elicited significantly shorter results than adjective- noun phrases (e.g. *oud lid*, 'old member'). Speakers took significantly longer to plan utterances containing adjective-noun phrases. Furthermore, the latencies for the compounds were similar to monomorphemic words (e.g. *orgel*, 'organ'). This supports evidence from a delayed naming task run by Inhoff et al. (1996), who also found that naming latencies for English compounds were nearly identical (434 ms) to monomorphemic words (433 ms).

More recently, Jacobs and Dell (2014) conducted a series of implicit-priming tasks that tested the production of simple words, phrases, and nominal compounds; their findings suggested that compounds had priming behaviour similar to the monomorphemic words. That is, the onsets of the second unit of nominal compounds did not receive priming (e.g. the /d/ in *sawdust*); although the same word did when used in an adjective-noun phrase (e.g. *grey dust*).

From the evidence above it seems that compounds and phrases do differ in production planning and that compounds pattern with monomorphemic words. However, although the theoretical analysis suggests that the phonological status of compounds and phrases differ, there is no experimental evidence that the single phonological word status of compounds is similar to that of monomorphemic words. For instance, as we have seen above, phonological encliticisation of weak function words is the norm in Germanic languages (e.g. *drinka pint* vs. *drink a pint*). If

compounds have the same status as single lexical words, they should also be subject to encliticisation and thereby further resort to recursive word formation. This would lead to an encliticised compound becoming even longer which in turn should have an effect on production latencies. We lay out our predictions below.

1.5 *Predictions*

First, although compounds and monomorphemic words pattern in the same way in a variety of tasks (Wheeldon & Lahiri 1997, 2002; Dell & Jacobs 2014), nevertheless if the compound is indeed a single recursively-structured prosodic unit, it must contain at least two syllables. Thus the size of the first phonological word will always be larger when it is a compound, as compared to a phrase where the adjective could be a single syllable: e.g. *blackboard* vs. *black board.* Moreover, a compound would also be larger than a monosyllabic monomorphemic word. Consequently, a task where latencies would be affected by the size of the first prosodic unit should find a difference between compounds on the one hand versus an adjective phrase consisting of a monosyllabic adjective and a monosyllabic single word.

Second, equally valid would be the assumption that encliticisation of these unstressed elements will lead to recursive word formation, as clitics attach to any stressed element: main or secondary. What we would like to argue here is that the assumption that it behaves as one prosodic unit has other consequences. Namely if it is one prosodic unit it should be able to undergo encliticisation and again recursively become a single prosodic unit. If this is indeed the case, then independent of the size of the compound, an enclitic would again recursively make the compound into one prosodic unit. Therefore, the utterance *a blackboard* would also be a single phonological word because, although the determiner *a* is an important syntactic unit, it attaches to *blackboard*

13

as a weak phonological unit. Thus, the appropriate phonological structure of the utterance "blackboards are heavy" would be as in (6):

(6) *Blackboards are heavy.*

      (a) **Lexical grouping**:   [blackboards]$_N$ [are$_v$ heavy$_{Adj}$]

      (b) **Phonological grouping**: ((bláck)$_\omega$ (boàrds)$_\omega$)$_\omega$ are)$_\omega$ (heavy)$_\omega$

The prediction is different for an adjective noun phrase, which contains two separate prosodic units: in these items, encliticisation (often auxiliary or any other item following the noun) would only encliticise to the second element and never to the first. The first element (the adjective) would be unaffected. That is to say, the initial prosodic unit would remain untouched and the second element that would grow with encliticisation as in example 7 below.

(7) *Black boards are heavy.*

      (a) **Lexical grouping**:   [black$_{Adj}$ boards]$_N$ [are$_v$ heavy$_{Adj}$]

      (b) **Phonological grouping**: (bláck)$_\omega$ ((boàrds)$_\omega$are)$_\omega$ (heavy)$_\omega$

These are the central predictions that we will test in this paper. We report four experiments that were designed to elicit information about how speakers plan multi-word utterances in English during the phonological encoding process. Following the work by Wheeldon and Lahiri (1997, 2002), we investigated naming latencies in both delayed and online tasks for English noun- noun compounds, adjective-noun phrases, and two types of monomorphemic words.

The experiments were arranged into two groups. The first group (Experiments 1 and 2) tested our first prediction; namely that compounds and phrases are treated differently in English spoken word production. Experiment 1 presented the four conditions (compounds, phrases, simple words with initial stress, and simple words with final stress) in a delayed naming task. While we have seen some evidence for the prosodic structure of compounds and phrases in picture-word interference tasks (e.g. Lange & Laganaro 2014) and phonological distractor tasks (Dell and Jacobs 2014), this experiment was designed to elicit further evidence via a pure naming task. The experiment used the same design as in Wheeldon and Lahiri (2002) to confirm how English compounds and phrases are being planned when speakers have time to prepare their utterances in full. Visual targets were followed by an auditory cue, after which speakers were given time to respond using the targets they saw. Following results from previous studies, the latency to begin speaking here should be determined by the total number of prosodic units in the utterance. That is, if compounds are indeed treated as single prosodic units, then they should pattern with monomorphemic words.

Experiment 2 presented four conditions (compounds, phrases, simple disyllabic words with initial stress, and simple monosyllabic words) in an online naming task. In section 1.3, we discussed experimental evidence for the flexibility of the planning scope. We know that the response times in this type of priming task is affected only by the size of the first prosodic unit (not the number of prosodic units) since this is all that the speaker has time to plan for (Wheeldon & Lahiri 1999, Experiment 4). For this reason, in addition to the delayed task we tested how speakers planned multi-word utterances when they have no time to prepare.

The second group (Experiments 3 and 4) tests our second prediction: the phonological planning of compounds and phrases with clitics in delayed and online task conditions. We

hypothesised that, if speakers plan compounds as single phonological units, then it follows that clitics should attach to compounds in the same way as they attach to monomorphemic words. That is, if function words that follow compounds are encliticised, then they would form a single unit; in contrast, a function word should only encliticise to the second element of a phrase.

Experiment 3 followed the delayed task design of Experiment 1, giving speakers time to prepare their utterances before initiating speech. To encourage connected speech and encliticisation, this task involved longer utterances and more complex structures. Thus, the responses contained sentences such as "dishcloths are clean" (compound), "green yards are clean" (phrase), and "grapes are clean" (monosyllabic simple word). The online version of this study (Experiment 4) allowed us to test encliticisation: that is, whether the word "are" in the utterances is reducing and attaching leftwards to a neighbouring word rather than to the right. As we saw from Experiment 2, naming latencies reflected the size of the first prosodic word when speakers had less time to plan; if "are" is encliticised (i.e. attached leftwards), the size of the initial prosodic unit would be larger which would be reflected in the response latencies. The key hypothesis here is that phrases would elicit significantly shorter naming latencies compared to all other conditions since speakers are required only to encode the first prosodic unit (the adjective consisting of one syllable). In contrast, compounds and disyllabic words should elicit the longest naming latencies because the compound-plus-clitic and the disyllabic word plus clitic would both contain three syllables. Under this hypothesis, monosyllabic words, on the other hand, should be significantly longer than phrases since the clitic is expected to attach to it resulting in a prosodic unit of two syllables.

2. EXPERIMENT 1

Experiment 1 used a delayed speech task to test whether native speakers of British English planned noun-noun compounds as single prosodic units during the phonological encoding process.

2.1. *Method*

2.2. *Materials*

The experimental materials were constructed from 60 items (e.g. *groundhog*), 15 per condition type. Four experimental conditions were distinguished: noun-noun compounds, adjective- noun phrases, disyllabic initially-stressed morphologically-simple words, and disyllabic final-stressed morphologically-simple words (see Table 1 for examples of responses and Appendix A for the full list of items). Two groups of disyllabic words were chosen as control conditions to test for any effect of stress pattern on naming latencies. Words were arranged into five sets beginning with one of five phonemes: /d/, /g/, /n/, /l/, and /b/. These phonemes were chosen in order to more easily identify boundaries during analysis. 9 morphosyntactically-simple disyllabic filler words were included as well. These served to mask the pattern of manipulation between morphosyntactically-complex and morphosyntactically-simple words.

Table 1 *Experimental Condition Types and Sample Responses*

	Condition Type	**Sample Response**
1	Noun-Noun Compound	It was daytime.
2	Adjective-Noun Phrase	It was dark time.
3	Stress Initial Simple Word	It was denim.
4	Stress Final Simple Word	It was decree.

Word familiarity was confirmed by rating data from 35 participants using a native speaker judgement task. The task contained the 60 experimental items along with filler words. Participants were asked to rate each word according to their familiarity with the word on a scale from 1 to 5, where 1 represented "not at all familiar" and 5 represented "extremely familiar". The mean judgement ratings for each condition (compounds, phrases, disyllabic initially-stressed words, and disyllabic final-stressed words) did not significantly differ from each other ($p= 0.12$). The stress of the compounds, disyllabic initially- stressed words and disyllabic final-stressed words was confirmed by 35 native Southern British English participants with questionnaires. These questionnaires presented the experimental items from the compound, disyllabic initially-stressed, and disyllabic final-stressed word conditions, along with filler words. Participants were asked to select which syllable of the word received the most stress, with the options "first", "second", and "not sure". Speaker judgements for the target items were generally strong, with all items rating at least 80% for correct stress.

The individual morphemes of compounds and phrases were also matched for a number of factors that could affect lexical recognition: word-length (number of letters), imageability, and frequency measures (simple CELEX frequency, written, and spoken counts per million as well as log frequency) (see Table 2 for means of these variables). As part of our analysis, we extracted frequencies for each unit of the compound and phrase target words, as well as the frequency of the compounds as wholes. The first morpheme of the compound (e.g. *ground*) was matched for length and frequency to the adjective of the phrasal condition (*green*). Within the compound target group, the initial compound unit contained the same form for each phoneme-organised set: e.g. (*DAYbreak*, *DAYtime*, *DAYlight*). The second morpheme of the compound condition was identical to the second morpheme of the phrasal condition (e.g. *groundHOG* and *green HOG*). The phrases

were also constructed in such a way to discourage speakers from compounding the two words in the phrase: therefore, they contain some unique combinations, such as *long jaw* and *green hog*. This also meant that the words in our phrases would have extremely low co-occurrence frequencies.

Table 2 *Means of variables for the experimental words (Comp 1 and Comp 2 refer to the first and last unit of the compound, Comp is the simple compound frequency)*

Variable	Comp 1	Comp 2	Comp	Adj	N	Sim Ini	Sim Fin	Mono
CELEX freq	291.5	258	5.3	298.7	259	2.9	3.2	19.6
CELEX written	299.6	259.2	5.7	301.8	259.2	2.8	3.3	19.7
CELEX spoken	187.1	243.2	1.3	258.9	243.2	2.4	1.6	19.7
Log Freq	2.1	1.6	0.5	2.4	1.6	0.5	0.5	1
Imageability	568.4	366.1	78	541	366.1	160.2	236.1	288.3
No of Letters	4.4	3.8	8.2	4	3.8	5.9	6.5	4.1

2.3. *Design*

The experiment was designed using E-Prime 2.0 software. There were 9 blocks of 21 trials in total (189 items total including filler words) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and three different versions of the experiment were presented. Each experimental word was presented once at each preparation latency: 800 ms, 1200 ms, and 1400 ms. Therefore, each word appeared three times over the experiment.

2.4. *Procedure*

Participants were tested individually. They were seated in a sound-attenuated room facing a Dell laptop, wearing headphones that contained a microphone. They were told that they would see words on the screen, then hear the question, "What was it?". The experimental question was recorded in a soundproofed booth by a male Southern British English speaker. The participants were asked to prepare the response as fully and naturally as possible. Before the experimental trials began, they completed a set of five practice trials.

Each trial began with a blank screen, shown for 500 ms. Then a fixation cross appeared centred on the screen for 500 ms. After the fixation cross disappeared, the target item was shown centred on the screen for 500 ms. Target words were presented in lowercase, black 18-point Courier New font on a white screen. The word then disappeared from the screen and the participant immediately heard the experimental question, "What was it?" through headphones. Then three beeps of equal duration were heard: the first occurred 2 seconds after the offset of the word, the next 1 second later, and the final beep occurred at a variable latency (800 ms, 1200 ms, or 1400 ms) from the offset of the second beep. These variable latencies were chosen in order to prevent participants from guessing when the prompt would occur, and falling into patterns. Participants were instructed to answer as quickly as possible after the third beep.

2.5. *Participants*

Eighteen participants between the ages of 18 and 26 took part in Experiment 1. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation.

2.6. *Analyses*

Participants' responses were recorded as audio files on a Sony DTC-100 ES DAT-recorder. Response latencies were measured from the offset of third beep to the onset of the response (e.g. "it was daytime"). Initial measurements were made using Voicekey software designed by Henning Reetz, which automatically calculated and labelled speech onsets. The software generated time-marks for the speech editing Praat software in a 'point-tier'. Every time-mark was then hand checked (and corrected if necessary) by an impartial coder.

2.7. RESULTS AND DISCUSSION

2.7.1. General Analysis

Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. In addition, the compound "nightmare" and its matched items in conditions 2-4 were removed from the dataset from this experiment and Experiment 2. This was done in response to feedback that "nightmare" was not fully transparent in meaning and might therefore be treated differently than the other, more-transparent compounds.[2] Responses uttered before the final beep in the delayed tasks were also discarded.

All RT data were submitted to two linear mixed effect models. Model 1 analysed the difference between phrases and all other conditions using the phrasal condition as an intercept, and Model 2 was restricted to the simple word conditions using the compound condition as an intercept (compounds, disyllabic simple words, and monosyllabic simple words). In both Models, word frequency was included in the fixed effects structure as a continuous variable. Three variants of

---

[2] All analyses were also run including 'nightmare' and there were no differences with the pattern of results reported.

the adjective-noun frequencies were tested: the average of the two elements of the adjective-noun phrase, the first element (the adjective), and the second element (the phrase).

In both models, naming latency was modelled as a function of the fixed effect factors, Condition Type (cond), Frequency (freq), and Preparation Time (beep) in the delayed tasks. Each model was dummy-coded with the critical condition as intercept. Subjects and items were treated as random factors. Following Barr et al. (2013), we began by testing the maximally-appropriate random structure and continued to adjust the model until it reached convergence. Following Baayen, Davidson and Bates (2008), all t-values greater than 2 were treated as significant. The maximal model contained interactions between condition and frequency, preparation time, random slopes and intercepts for the interaction between condition, frequency and beep within subjects, and random intercepts for condition, frequency and beep within items. The models also included factor of Block (1-10) and word length, and interactions of these factors with Condition to test for effects of practice and word size over the course of the experiment. These interactions will only be reported if significant.

As part of this investigation, an error analysis was also carried out: errors were categorised as "time out" (the subject said nothing), "voice key error", "disfluency" (e.g. stuttering), "stress error" (incorrect prosody), and "wrong item". For the error analyses only subject-produced errors were included (technical errors such as voice key errors were not included in this analysis). The means of subject errors are shown in Table 3.

2.7.2 *Analysis: Experiment 1*

All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a loss of 9.1% of data (4.2% of compounds, 2.7% of phrases, 3.2% of

initially-stressed words, and 4.7% of finally-stressed words). The resulting mean RTs and percentage error rates for each condition and preparation time are shown in Table 3. As can be seen, on average the phrases show the longest speech onset latencies and the lowest error rate compared to all other conditions, which show only small differences to each other. This pattern is the same across all preparation latencies.

Table 3 *Naming Latencies (in ms) and percentage error rates (in parentheses) are shown for the four conditions of* **Experiment 1,** *on average and across the three planning durations*

	(1) Compounds	(2) Phrases	(3) Initial Stress	(4) Final Stress	Mean Lat. (ms)
PWds:	1	2	1	1	
LexWs:	2	2	1	1	
Syllables	2	2	2	2	
Beep Lat.: 800 ms: 1200 ms: 1400 ms:	452 (2.5%) 443 (1.7%) 440 (2.5%)	494 (1.7%) 476 (1.7%) 485 (0.7%)	448 (2.2%) 443 (2.1%) 435 (2.3%)	465 (2.8%) 439 (1.6%) 445 (1.7%)	465 (2.3%) 450 (1.8%) 452 (1.8%)
Mean Lat. (ms)	**445** (6.6%)	**485** (4.2%)	**442** (6.6%)	**449** (6.2%)	
SD:	137	156	136	144	

The maximal model to converge[3] for our first analysis (Model 1) contained an interaction between Condition and Frequency and Beep in the fixed effect factors, random slopes and intercepts for Condition in the by-subject analysis, and random intercepts in the by-item analysis. This model converged and its output is shown in Table 4, where significant effects are marked

---

[3] Formula: rt ~ cond*celexfreqadj + cond*beep + (1 + cond | sub) + (1 | item).

with an asterisk. This model appeared comfortably homoscedastic when fitted to residual plots. In this analysis, the adjective-noun phrase served as the intercept. The results revealed significant differences in the naming latencies for compounds ($t=3.19*$), disyllabic simple words ($t=3.0*$), and monosyllabic simple words ($t=2.43*$). There was no interaction between Condition and Frequency for any of the three variants we tested. For the sake of brevity, only the first model (which included the average frequency for the adjective-noun phrase condition) has been reported below. An analysis of the interaction of Preparation Time and Condition revealed no effect on reaction times.

In the second analysis (Model 2, Table 5), the most maximal model to converge contained an interaction between Condition and Frequency and Condition and Beep[4] in the fixed effect factors, random slopes and intercepts for Condition and Beep in the by-subject analysis, and random intercepts in the by-item analysis. Here we found no significant difference in the naming latencies between compounds and disyllabic initial stress or disyllabic final stress) words. There was no effect of Preparation Time with Condition on reaction times.

Table 4 Linear Mixed-Effects Analyses of RT Data in Experiment 1 (**Model 1: full model, with**

---

[4] Formula: rt ~ cond*beep+ cond*freq + (1+cond+beep|sub)+ (1|item).

average adjective-noun phrase frequencies, phrases as intercept)

	Estimate	Std. Error	t value
(Intercept: Adj-N phrases)	492.589905	30.035376	16.400
Compounds	-41.097588	12.902291	-3.185*
Simple initial disyllabic	-48.943127	16.335630	-2.996*
Simple final disyllabic	-30.221239	12.425949	-2.432*
Beep2	-14.392648	8.422697	-1.709
Beep3	-5.662024	8.364544	-0.677
Frequency	-0.004021	0.007175	-0.560
Compounds: beep2	8.125722	11.910867	0.682
Simple initial: beep2	9.828434	11.951682	0.822
Simple final: beep2	-9.605807	11.984476	-0.802
Compounds: beep3	-4.958837	11.942007	-0.415
Simple initial: beep3	-5.402878	11.925089	-0.453
Simple final: beep3	-15.019299	11.927666	-1.259
Compounds: frequency	-0.205878	0.377931	-0.545
Simple initial: frequency	1.617768	1.469819	1.101
Simple final: frequency	0.627682	1.181047	0.531

Table 5 *Linear Mixed-Effects Analyses of RT Data in Experiment 1 (**Model 2: Simple words only,***

compounds as intercept)

	Estimate	SE	t Score
(Intercept: Compounds)	451.90045	25.54494	17.690
Simple initial disyllabic	-4.50789	9.61706	-0.469
Simple final disyllabic	13.45188	8.92021	1.508
Beep2	-6.73827	11.80299	-1.347
Beep3	-12.77076	9.47865	-1.004
Frequency	-0.62563	0.62314	-1.004
Simple initial: beep2	6.23943	12.19398	0.512
Simple final: beep2	-16.49964	11.92412	-1.384
Simple initial: beep3	4.73011	12.21441	0.387
Simple final: beep3	-10.046	11.95707	-0.840
Simple initial: freq	-0.63634	2.62856	-0.242
Simple final: freq	2.24206	2.16009	1.038
beep2: freq	0.06101	0.89895	0.068
beep3: freq	1.00102	0.89232	1.122

Error rates were analysed using a generalised linear model (GLM) with a binomial distribution (link= logit). The analysis (Table 6) revealed no effect of preparation time, condition type, or block.

Table 6 *Generalised Linear Model of Error Data in **Experiment 1** (family=binomial, link= logit)*

	Estimate	SE	z Score
Preparation Time: (Intercept)	-3.13050	0.28092	-11.14
Beep:	0.08661	0.13320	-0.65
Condition Type: (Intercept)	-3.27002	0.23504	-13.913
Cond:	-0.01324	0.08860	-0.149
Block: (Intercept)	-3.277186	0.231794	-14.138
Block:	-0.004861	0.041422	-0.117

2.7.3 *Discussion*

In this experiment, we observed a significant effect of Condition: the size of this effect is similar (in fact, greater) to that in Wheeldon and Lahiri (2002). Adjective noun phrases resulted in reaction times that were approximately 40 ms longer than the compounds. A cross-analysis of the naming latencies of compounds with both types of simple disyllabic words did not result in any significant difference, suggesting that speakers planned these three groups of targets similarly. There was a significant difference in the time it took speakers to produce adjective-noun phrases in comparison with all other target words. This suggests that speakers treated the phrases as two

discrete phonological units, and crucially, they also treated compounds as single phonological word units for the purposes of this task.

3. EXPERIMENT 2

Experiment 1 provides evidence that, prior to articulation, speakers plan compound words as single phonological words when they are given time to prepare their utterances. But these results only tell us about phonological encoding under delayed planning conditions. What happens when planning time is taken away? Will speakers sacrifice utterance initiation speed in order to encode a fully-formed unit?

Despite evidence that speech production is affected by elimination of planning time, relatively few production studies run their experimental tasks in both delayed and online task conditions. Wheeldon and Lahiri (1997) found that Dutch speakers took significantly longer to produce sentences containing clitics when they presented their task in online naming conditions. This suggested that speakers were only able to prepare the first phonological word before articulation, and that response time was affected by the size and complexity of that unit. They also found that participants preferred to construct well-formed phonological words even at the cost of utterance initiation speed. Lange and Laganaro (2014) ran an online picture-word interference task using French noun phrases, such as (e.g. *grand chat*, "large cat"). They found that only the first element of the phrase was primed by a phonologically related distractor; no priming effect was found when the second element in the phrase was primed. Based on this evidence, they concluded that the unit of planning displayed flexibility, yet remained a well-formed phonological word.

We hypothesise that the naming latencies of the native English speakers will reflect only the first prosodic unit of the utterance. The latencies will differ depending on the size and

complexity of this initial unit. If speakers continue to plan well-formed phonological words even under online conditions, the naming latencies for compounds should reflect a single phonological word unit.

3.1. *Method*

3.2. *Materials*

The aim of Experiment 2 was to test whether removing the preparation time in Experiment 1 would affect post-lexical encoding of compound words. Therefore, Experiment 2 presented almost identical experimental items to Experiment 1. Because we hypothesise that only the first unit of the target will determine the production latency, we had to ensure that the control conditions contained corresponding structures. Therefore, four experimental conditions were again distinguished: noun-noun compounds, adjective-noun phrases, disyllabic morphologically-simple words, and monosyllabic morphologically-simple words (see Table 7 for examples and Appendix B for the full list of items). High-frequency monosyllabic words were chosen for Condition 4 due to the high frequencies of the adjectives in the adjective-noun phrases. 15 disyllabic filler words were included as well.

Table 7 *Experimental Condition Types and Sample Responses:* ***Experiment 2***

	Condition Type	Sample Response
1	Noun-Noun Compound	daytime
2	Adjective-Noun Phrase	dark time
3	Disyllabic Simple Word	denim
4	Monosyllabic Simple Word	date

3.3. *Design and Procedure*

The design of this experiment was similar to Experiment 1 except that participants were required to perform an online production task. Therefore, there were no preparation or deadline beeps and participants were instructed to respond as quickly as possible following the auditory question. In addition, it was important in this experiment that the target words occurred sentence-initially in the participants' responses. Therefore, participants were instructed to produce the target word only.

3.4. *Participants*

Twenty participants between the ages of 18 and 26 took part in Experiment 2. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation.

3.5. *Analyses*

The analyses were the same as Experiment 1. Latencies were measured from the end of the [t] plosive of the auditory prompt ("What was it?") to the onset of the target's utterance.

3.6. *Results and Discussion*

3.6.1. *RT Analysis*

Measurements were undertaken in the same way as in Experiment 1. Errors resulted in a loss of 9.5% of data. All data points beyond two standard deviations from the mean were counted as outliers and removed. This, combined with the errors, resulted in a total loss of 12.4% of data (12.6% from compounds, 12% from phrases, 11% from disyllabic words, and 13.7% from monosyllabic words). The resulting mean RTs and percentage subject error rates are shown in Table 8.

Table 8 *Naming Latencies (in ms) and percentage error rates (in parentheses) are shown for the four conditions of **Experiment 2**, on average.*

	(1) Compounds	**(2) Phrases**	**(3) Disyllabic**	**(4) Mono-syllabic**
PWds:	1	2	1	1
LexWs:	2	2	1	1
Syllables	2	2	2	1
Mean Lat. (ms)	**262** (8.3%)	**223** (8.6%)	**259** (7%)	**221** (14.3%)
SD:	96	90	97	82

The maximal model[5] to converge for our first analysis (Model 1) contained an interaction between Condition and Frequency in the fixed effect factors, random slopes and intercepts for Condition in the by-subject analysis, and random intercepts in the by-item analysis. The output of

---

[5] Formula: rt~ cond*celexfreqadj + (1 +cond| sub) + (1| item).

this analysis can be found in Table 9 below. This model appeared comfortably homoscedastic when fitted to residual plots. The results from the mixed-effect analysis of Condition on reaction time revealed a significant difference in the naming latencies for compounds (t=3.19*) and disyllabic simple words (t=4.05*). There was no difference between the average naming latencies of the phrasal condition and the monosyllabic simple word condition (t=0.79). There was no interaction between Condition and Frequency for any of the three frequency variants we tested. For the sake of brevity, only the first model (which included the average frequency for the adjective-noun phrase condition) has been reported in Table 9. On average the phrases and monosyllabic conditions showed the shortest speech onset latencies, similar to one another. The compound and disyllabic word conditions resulted in longer naming latencies, similar to one another as well.

Table 9 *Linear Mixed-Effects Analyses of RT Data in Experiment 2* **(Model 1: full model, with average adjective-noun phrase frequencies)**

	Estimate	Std. Error	T value
(Intercept: Adj-N Phrases)	222.35926	17.99206	12.359
Compounds	37.20535	10.78320	3.450*
Disyllabic	44.71232	11.04499	4.048*
Monosyllabic	8.28222	10.45821	0.792
Frequency	0.02352	0.01760	1.336
Compounds: frequency	0.71404	0.44052	1.621
Disyllabic: frequency	-1.97699	1.72465	-1.146
Monosyllabic: frequency	-0.09847	0.17756	-0.555

In our second analysis (Model 2), the maximal model[6] to converge contained an interaction between Condition and Frequency in the fixed effect factors, random slopes and intercepts for Condition for subjects, and random intercepts for items. The output of this model can be found in Table 10 below. This analysis, which used compounds as the intercept, revealed a significant difference between compounds and monosyllabic simple words only (t=-2.72*). Compounds and disyllabic words were statistically similar (t=0.78). Additionally, there was no effect of an interaction between Condition and Frequency in this analysis.

Table 10: *Linear Mixed-Effects Analyses of RT Data in Experiment 2* (**Model 2: Simple words only**)

	Estimate	SE	t Score
(Intercept: Compounds)	259.4504	15.8021	16.419
Disyllabic	7.5985	9.8032	0.775
Monosyllabic	28.6004	10.5059	-2.722*
Frequency	0.7459	0.4613	1.617
Disyllabic: frequency	-2.6988	1.8642	-1.448
Monosyllabic: frequency	-0.8188	0.4970	-1.647

Error rates were analysed using a generalised linear model (GLM) with a binomial distribution (link=logit). The analysis (Table 11) revealed no effect of Condition. We found that the number of errors (Figure 5.14) decreased in blocks 2 and 3, and rose in blocks 4 and 5; however no effect of Block was found in the analysis.

---

[6] Formula: rt~ cond*celexfreq + (1 +cond| sub) + (1| item).

Table 11 *Generalised Linear Model of Error Data in **Experiment 2** (family=binomial, link= logit)*

	Estimate	SE	z Score
Condition Type: (Intercept)	-5.0041	0.5675	-8.817
Cond:	0.3399	0.1899	1.790
Block: (Intercept)	-4.9192	0.6425	-7.657
Block:	0.2307	0.1736	1.329

3.6.2 *Discussion*

As in Experiment 1, the results from Experiment 2 showed significant effects of Condition on reaction times: responses containing compounds also showed no difference in naming latencies from disyllabic, morphosyntactically-simple words. In contrast to Experiment 1, however, speakers in this task consistently took longer to produce utterances containing compounds (262 ms) than they took to produce the adjective-noun phrases (223 ms). These results are consistent with those found in Meyer and Schriefers (1991), Meyer (1996), Wheeldon and Lahiri (1997), Schriefers and Teruel (1999), and Lange and Laganaro (2014), where on-line naming latencies only reflected the first element of adjective-noun phrases. The mean naming latency (223 ms) for adjective-noun phrases reflects encoding of the first phonological unit, resulting in similar naming latencies to the monosyllabic word condition (221 ms): this suggests that, when planning time was removed, speakers were only able to encode the first unit of the adjective-noun phrase condition. Crucially, the RT data for compounds suggests that speakers waited until the entire phonological unit was ready before beginning articulation. This indicates that speakers once again treated noun-

noun compounds as single phonological units, in contrast to the phrases (which were treated as separate phonological units).

3.7. *Interim Discussion*

In Experiments 1 and 2, we found that English native speakers treat two-word compounds as single units, with RT data statistically different from that of adjective-noun phrases. In the first set of experiments (Experiments 1 and 2), we compared production latencies of utterances containing compounds to those of utterances containing syntactic phrases (adjective-noun phrases) or morphologically simple words. In the delayed task, naming latencies reflected the total number of phonological words in the utterance. Speakers took an average of 40 ms longer to produce utterances containing syntactic phrases than noun-noun compounds. Utterances containing compounds generated statistically similar production latencies to morphologically-simple words. Conversely, when planning time was removed in Experiment 2, latencies were only sensitive to the first phonological unit of the utterance. Both of these findings are in line with those in Wheeldon and Lahiri (1997, 2002).

In both Experiments 1 and 2, we saw a significant effect of condition type on production latencies. In the delayed task (Experiment 1), compound words elicited shorter naming latencies than adjective-noun phrases, similar to those of disyllabic morphosyntactically-simple words. In this experiment, the number of syllables was held constant across all conditions, while the number of phonological words and lexical words, and the placement of stress all varied. Despite these variations, the only thing that reliably predicted naming latencies was the total number of prosodic units in the utterance. Although compounds contained two morphosyntactic words (and therefore two phonological words), they were treated as single prosodic units in this task, generating nearly

35

identical naming latencies to morphosyntactically-simple words. In the online task (Experiment 2), we saw a different pattern of effects: compound words elicited longer naming latencies than phrases, similar only to those found for disyllabic morphosyntactically-simple words. Adjective-noun phrases were associated with shorter latencies in this task, similar only to those found for the monosyllabic simple words.

Experiment 2 made clear the importance of testing the stimuli in both task types: when speakers had enough time to plan, the total number of prosodic units dictated the naming latencies. When planning time was restricted, however, the phonological encoding process was only able to prepare the first prosodic unit before articulation: this was evident in the naming latencies in Experiment 2. Crucially, speakers sacrificed utterance initiation time in order to access well-formed prosodic units, in the case of the compound words: they did not begin speaking until the entire compound structure was encoded. While the data in Experiment 1 serves to support existing evidence for the phonological word in phonological encoding, Experiment 2 offers an important additional contribution to this process: it revealed that speakers still preferred to plan compounds as a single prosodic structure. These findings support both the theories regarding the prosodic word structure of compounds and the shape of the planning unit in the word-form encoding process.

4. EXPERIMENT 3

Experiments 1 and 2 have established that English speakers are treating noun-noun compounds as single phonological units. The aim of Experiment 3 and 4 is to test the limits of this hypothesis. Our theory and accompanying empirical evidence dictates that weak function words behave like clitics and will attach to a phonological word in order to form a larger recursive phonological word. Thus, the total number pf phonological words would not increase with the

addition of a clitic. If this is indeed the case, then clitics should attach to compound words as a whole and form a single phonological unit and the cliticised compounds will elicit production latencies similar to those of cliticised disyllabic words. Cliticised phrases will elicit longer latencies, due to the total number of phonological words in the utterance being greater than all other conditions. If, however, the clitic should only attach to the last unit of the compound, then the mean naming latencies for the compound and phrasal conditions will be the same. Note, that this experiment does not actually support encliticisation – i.e., the function word attaching leftwards; it could potentially add rightwards. The crucial test here is that the function word does not add to the number of prosodic words.

4.1. *Method*

4.2. *Materials*

In order to ensure that speakers were generating well-formed sentences, we had to present target words that would easily cliticise in plural form with the verb *are*, e.g. *dishcloths are clean*. This required a new target wordlist, with different intended utterances (see Table 12). The experimental materials were constructed from 60 items, 15 per condition type, selected from the CELEX database based on their lexical frequency. Four experimental conditions were distinguished: noun-noun compounds, adjective-noun phrases, initially-stressed disyllabic simple words, and monosyllabic simple words (see Appendix C for full list of items). 10 disyllabic filler words were included as well.

Table 12: *Experimental Condition Types and Sample Responses: Experiment 3*

Condition Type	Target	Sample Sentence
N-N Compound	dishcloths	Dishcloths are clean.
Adj-N Phrase	drab cloths	Drab cloths are clean.
Disyllabic	donkeys	Donkeys are clean.
Monosyllabic	drapes	Drapes are clean.

Table 13: *Means of variables for the experimental words (**Experiment 3**)*

	Comp 1	Comp 2	Comp	Adj	N	Di	Mono
CELEX frequency	130.5	170.4	0.5	266.6	170.4	8.4	10.4
CELEX written	105.8	169.9	0.5	269.4	169.9	8.9	10.8
CELEX spoken	74	177.2	0.2	231.2	177.2	2.2	4.8
Log Frequency	1.7	1.8	0.1	2.2	1.8	0.8	0.8
Imageability	498.8	488.5	-	453.2	488.5	395.1	495.6
No of Letters	4.1	5.1	9.3	4.1	5.1	6.2	4.4

A word familiarity task was run with these targets and resulted in high ratings: mean judgement ratings for each condition did not significantly differ from each other ($p=.67$).

4.3. *Design*

The experiment was designed using E-Prime 2.0 software. There were 10 blocks of 19 trials in total (190 items total including filler words) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and five different versions of the experiment were presented, in which the sets of experimental questions were rotated. Each

experimental word was presented once at each preparation latency: 800 ms, 1200 ms, and 1400 ms. Thus, each word appeared three times over the experiment.

4.4. *Procedure*

Participants were tested individually. They were seated in a sound attenuated room facing a Dell laptop, wearing headphones that contained a microphone. They were told that they would see words on the screen, then hear five different questions (see Appendix D).

The experimental questions were recorded in a soundproofed booth by a female Southern British English speaker. The participants were asked to prepare the response as fully and naturally as possible. Before the experimental trials began, they completed a set of five practice trials. The rest of the experimental design was identical to Experiment 1.

4.5. *Participants*

Twenty participants between the ages of 18 and 29 took part in Experiment 3. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation.

4.7. *Results and Discussion*

4.7.1. *RT Analysis*

All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a loss of 8.1% of data (9.7% of compounds, 7.1% of phrases, 6.9% of disyllabic words, and 8.7% of monosyllabic words). Mean reaction times and percentage error rates for Experiment 3 are shown in Table 14.

Table 14 *Naming Latencies (in ms) and percentage error rates (in parentheses) are shown for the four conditions of **Experiment 3**, on average and across the three planning durations*

	(1) Compounds + Clitics	(2) Phrases +Clitics	(3) Disyll + Clitics	(4) Monosyll +Clitics	Mean Lat. (ms)
PWds	1	2	1	1	
LexWs	2	2	1	1	
Syllables	3	3	3	2	
Beep Lat.:					
800 ms.	343 (7.3%)	416 (4.0%)	333 (4.7%)	344 (5.0%)	359 (4.9%)
1200 ms.	328 (5.0%)	394 (4.3%)	322 (4.3%)	325 (7.3%)	342 (5.3%)
1400 ms.	322 (6.7%)	390 (4.3%)	321 (6.3%)	327 (4.7%)	341 (5.8%)
Mean Lat. (ms)	**331** (6.3%)	**400** (4.2%)	**326** (5.1%)	**332** (5.6%)	
SD	**115**	**133**	**110**	**112**	

The maximal model[7] for our first analysis (Model 1) contained an interaction between Condition and Preparation time, Condition and Frequency, and Condition and Block in the fixed effects, an interaction with random slopes and intercepts for Condition for subjects analysis, and random intercepts for items. This model appeared fairly homoscedastic when fitted to residual plots. Crucially, a log-transform did not result in a better model fit.[8] The output for this model can be found in Table 15 below. The results from this analysis, which used the adjective-noun phrases as intercept, revealed a significant difference in the naming latencies for compounds (t=-8.72*), disyllabic simple words (t=-9.07*), and monosyllabic simple words (t=-8.38*). Preparation time

---

[7] Formula: rt ~ cond*beep+ cond*celexfreqadj+ cond *block+(1+cond|sub)+ (1|item).
[8] Profile log-likelihoods of reaction times were generated via a Box-Cox power transformation from the "Car" package (Fox et al., 2009).

was significant in this experiment, but there was no interaction between any of the conditions and preparation time in the analysis. Additionally, there was no interaction of frequency with Condition in this experiment.

The second analysis (Model 2, Table 16), again removed the adjective-noun phrase condition and used compounds as the intercept. This model did not generate any significant differences between the naming latencies of compounds and disyllabic (t=-1.49) or monosyllabic words (t=3.41*). Preparation Time generated significantly different naming latencies, but there was no interaction between Preparation Time and Condition. In both analyses, adjective-noun phrases differed significantly from the other three conditions, while noun-noun compounds showed no significant difference from either of the morphologically-simple word conditions (disyllabic or monosyllabic).

Table 15 *Linear Mixed-Effects Analyses of RT Data in Experiment 3 (**Model 1: full model, with average adjective-noun phrase frequencies)***

	Estimate	Std. Error	T value
(Intercept: Adj-N phrases)	416.084900	25.374316	16.398
Compounds	-71.517928	8.198070	-8.724*
Disyllabic	-79.838564	8.804404	-9.068*
Monosyllabic	-68.644924	8.189517	-8.382*
Frequency	0.000947	0.011301	0.084
Beep2	-17.009182	5.373811	-3.165*
Beep3	-19.170199	5.408839	-3.544*
Block 10[9]	-25.3527	9.5171	-2664*
Compounds: frequency	-1.425355	2.879355	-0.495
Disyllabic: frequency	-0.084291	0.281667	-0.299
Monosyllabic: frequency	-0.263148	0.197551	-1.332
Compounds: beep2	-4.562581	7.569259	-0.603
Disyllabic: beep2	2.298580	7.575666	0.303
Monosyllabic: beep2	-1.788515	7.606821	-0.235
Compounds: beep3	-9.211473	7.580883	-1.215
Disyllabic: beep3	5.160293	7.621945	0.677
Monosyllabic: beep3	0.462791	7.628192	0.061

---

[9] Analyses of blocks 1-9, which were not significant, have been truncated due to size.

Table 16 *Linear Mixed-Effects Analyses of RT Data in Experiment 3 (**Model 2: Simple words only**)*

	Estimate	SE	t Score
(Intercept: Compounds)	345.2387	22.2422	15.522
Disyllabic	-10.4299	6.9935	-1.491
Monosyllabic	0.1741	6.8286	0.026
beep2	-17.1059	6.0986	-2.805*
beep3	-21.0011	6.1636	-3.407*
frequency	-2.5473	4.6844	-0.544
Block 10[10]	-26.5506	9.5787	-2.772*
Disyllabic: beep2	8.6856	9.4649	0.918
Monosyllabic: beep2	2.6499	9.1112	0.291
Disyllabic: beep3	4.9242	9.5559	0.515
Monosyllabic: beep3	4.1228	9.1501	0.451
Disyllabic: freq	2.6355	4.7071	0.560
Monosyllabic: freq	2.4810	4.6951	0.528
beep2: freq	0.1005	6.4282	0.016
beep3: freq	3.1444	6.3163	0.498

The results also revealed an effect of block on reaction times: latencies were significantly faster in Block 10 than Block 1 (t=-2772*); however, there was no interaction of Condition and Block on naming latencies. There was no effect of word length or prompt type.

---

[10] Analyses of blocks 1-9, which were not significant, have been truncated due to size.

The error rates were analysed using a generalised linear model (GLM) with a binomial distribution (link= logit). This analysis (Table 17) revealed a weak effect (z= 2.016*) of Condition on error rate. There was no effect of preparation time or Block on error rates.

Table 17 *Generalised Linear Model of Error Data in **Experiment 3*** (family=binomial, link= logit)

	Estimate	SE	z Score
Preparation Time:			
(Intercept)	-2.69856	0.29301	-9.21
Beep:	-0.07687	0.13874	-0.554
Condition Type:			
(Intercept)	-2.48357	0.20368	-12.194
Cond:	-0.1721	0.08538	-2.016*
Block:			
(Intercept)	-2.58408	0.20802	-12.422
Block:	-0.05294	0.03659	-1.447

4.7.2 *Discussion*

In this experiment, we observed a significant effect of Condition: the size of this effect is similar to that in Experiment 1. Cliticised adjective-noun phrases resulted in the longest mean naming latency (400 ms), while cliticised compounds elicited a much shorter mean latency (331 ms). This finding was consistent with the results from Experiment 1, which indicated that speakers were planning compounds and phrases differently. Naming latencies for the clitic + compound condition were much shorter, similar to those for both morphologically-simple word conditions.

5. EXPERIMENT 4

Experiment 4 presents the target words of Experiment 3 in online task conditions. We saw in Experiment 2 that planning time affected phonological encoding: under restricted planning, speakers only encoded the first phonological unit of the utterance. For compounds, this was the entire recursive unit; for phrases, this was only the first phonological unit (the adjective). Based on our earlier hypotheses, we predict that clitics will behave no differently in online task conditions: they will still attach to compounds and form a larger phonological word. Because naming latencies only reflect encoding of the first phonological unit, the latencies for adjective-noun phrases will only reflect encoding of the adjective. This will result in significantly shorter latencies than all other conditions, including the monosyllabic word condition.

5.1. *Method*

The materials and design were identical to those used in Experiment 3. The procedure was altered to an online production task. Participants were instructed to respond as quickly as they could, following the auditory question.

5.2. *Participants*

Twenty-five participants between the ages of 18 and 26 took part in Experiment 4. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation.

5.3. *Results and Discussion*

5.4.1. RT Analysis

Data trimming treated all data points beyond two standard deviations from the mean as outliers: this resulted in the removal of 56 data points (4.3% of compounds, 3.2% of phrases, 2.7% of disyllabic words, and 4.8% of monosyllabic words). The total loss made up only 3.8% of data. The resulting mean RTs and percentage error rates are shown in Table 18.

Table 18 *Naming Latencies (in ms) and percentage error rates (in parentheses) are shown for the four conditions of **Experiment 4**, on average.*

	(1) Compounds	(2) Phrases	(3) Disyllabic	(4) Mono-syllabic
PWds:	1	2	1	1
LexWs:	2	2	1	1
Syllables	2	2	2	1
Mean Lat. (ms)	**215** (4.2%)	**190** (3.2%)	**222** (4.0%)	**208** (4.3%)
SD:	67	63	64	66

As can be seen, on average the phrasal condition elicited the shortest speech onset latencies. The latencies for the monosyllabic word condition were longer, averaging 208 ms, and the latencies for the phrasal and disyllabic word conditions were the longest.

The maximal model[11] to converge for our first analysis (Model 1) contained an interaction between Condition and Frequency in the fixed effect factors, random slopes and intercepts for Condition in the by-subject analysis, and random intercepts in the by-item analysis. The output of this analysis can be found in Table 19 below. This model appeared comfortably homoscedastic

---

[11] Formula: rt ~ cond*celexfreqadj + (1+cond|sub)+ (1|item).

when fitted to residual plots. A log-transform did not improve fit. The results from the analysis revealed significant differences between the intercept (adjective-noun phrases) and all other conditions: compounds (t=5.63*), disyllabic simple words (t=5.57*), and monosyllabic words (t=4.09*). There was no effect of or interaction involving block, word length, or prompt type.

In our second analysis (Model 2), the maximal model[12] to converge contained an interaction between Condition and Frequency in the fixed effect factors, random slopes and intercepts for Condition in the by-subject analysis, and random intercepts in the by-item analysis. The output of this model can be found in Table 20 below. This analysis, which used compounds as the intercept, only revealed a significant difference between the naming latencies for compounds and monosyllabic words (t=-2.94*). Naming latencies for disyllabic words and compounds were statistically similar (t=0.42). Additionally, there was no effect of interaction between Frequency and Condition.

Table 19 *Linear Mixed-Effects Analyses of RT Data in Experiment 4 (**Model 1: full model, with**

---

[12] Formula: rt~ cond*celexfreq + (1 +cond| sub) + (1| item).

average adjective-noun phrase frequencies)

	Estimate	Std. Error	T value
(Intercept: Adj-N Phrases)	191.295976	11.036866	17.332
Compounds	29.800978	5.292348	5.631*
Disyllabic	27.533549	4.942097	5.571*
Monosyllabic	24.234570	5.931431	4.086*
Frequency	0.008327	0.010802	0.771
Compounds: frequency	0.006529	0.023926	0.273
Disyllabic: frequency	0.018968	0.015883	1.194
Monosyllabic: frequency	-0.024258	0.015051	-1.612

Table 20 *Linear Mixed-Effects Analyses of RT Data in Experiment 4 (**Model 2: simple words only)***

	Estimate	SE	t Score
(Intercept:Compounds)	221.840304	10.907763	20.338
Disyllabic	1.646925	3.911548	0.421
Monosyllabic	-11.667143	3.965861	-2.942*
Frequency	0.005005	0.008179	0.612
Disyllabic:freq	0.01325	0.02530	0.524
Monosyllabic:freq	-0.03079	0.02476	-1.244

The error rates were analysed (Table 21) using a generalised linear model (GLM) with a binomial distribution (link= logit). There was a weak effect ($z = -2.085$*) of Condition on errors,

with Condition 4 eliciting more errors (z= -2.055*) than other conditions. Block did not affect error rate (z= -0.466).

Table 21 *Generalised Linear Model of Error Data in **Experiment 4** (family=binomial, link= logit)*

	Estimate	SE	z Score
Condition Type:			
(Intercept)	-2.62963	0.20017	-13.137
Cond:	-0.19176	0.09196	-2.085*
Block:			
(Intercept)	-2.89287	0.23073	-12.538
Block:	-0.03014	0.06472	-0.466

5.4.2 *Discussion*

In this experiment, we again observed a significant effect of Condition. Cliticised adjective-noun phrases resulted in the shortest mean naming latency (203 ms), while cliticised compounds elicited a much longer mean latency (234 ms). A cross-analysis of the naming latencies for compound + clitic condition only showed a statistically-significant result for the disyllabic control condition. The mean naming latency for the monosyllabic word + clitic condition (208 ms) was significantly faster than either the disyllabic (222 ms) or compound condition (215 ms). Adjective-noun phrases were faster yet, with a mean naming latency of 190 ms.

6. GENERAL DISCUSSION

Classic language production models (Levelt 1989, et al. 1999, and Roelofs 1997) do not assume a lexically-shaped unit as the minimal underlying representation during phonological encoding. Both have agreed that the unit being prepared at this stage is the phonological word

(rather than the lexical word), which is generated from metrical information retrieved from the lexicon according to the phonological rules of the language. Our study supports the claim that the phonological word is the unit of phonological encoding, as our data confirm that when planning an utterance, the crucial unit can be more than a lexical word; it can be in fact a lexical word accompanied by a phonologically weak element such as determiners and prepositions.

In this study, we have also shown that metrical information alone is not sufficient for determining the units of prosodification. Our studies have gone beyond single lexical items and compounds, which are composed of at least two stressed lexical items e.g. *dóghòuse*, compared to a two word phrase, which also has two stressed lexical items, e.g. *dàrk hóuse*. Our question was: is *doghouse* treated the same way as *dark house* during phonological encoding or is it treated the same way as *dóg*? As we argued above, lexically *doghouse* is one prosodic unit indistinguishable from *dog* and indeed *dogmatic* despite its metrical similarity to *dark house*. This is because phonological words are built recursively and they are the largest phonological unit that is fed into post-lexical phonological encoding processes.

It therefore follows that post-lexical phonological processes should allow unstressed items to cliticise to compounds as a whole in the same way as with monomorphemic lexical items. In contrast the same unstressed items will cliticise to only the last element of a phrase (e.g. *house* in our example above). No current model of language production makes this distinction between lexical and postlexical prosodic processes.

We ran a delayed speech production task based on Sternberg et al. (1978) and modelled after Wheeldon and Lahiri (2002) to test prosodification in English. We compared naming latencies for noun-noun compounds, adjective-noun phrases, and monomorphemic nouns and found that naming latencies for the compounds were significantly shorter than the phrases, and

similar to the latencies for the monomorphemic words. Speakers took an average of 40 ms longer to produce utterances containing syntactic phrases than noun-noun compounds. Utterances containing compounds generated statistically similar production latencies to morphologically-simple words. This suggested that English speakers behave the same as those in the other studies. However, evidence from online planning experiments (cf. Wheeldon & Lahiri, 1997) indicate that the shape of the planning unit might change under different task conditions. Consequently, we ran an online version of the production task in Experiment 1: here we found that English speakers still planned prosodic units during online planning. As before, latencies for this task revealed sensitivity to the first phonological unit of the utterance. Both of these findings are in line with those in Wheeldon and Lahiri (1997, 2002).

In the series of experiments reported above, we found that English native speakers treat two-word compounds as single prosodic units, with RT data statistically different than that of adjective-noun phrases. The first set of experiments established that the planning unit in English speakers was also a phonological word. Crucially, we also found that speakers treated compound words as single prosodic items, at least for the purposes of planning spoken utterances.

In order to test our second hypothesis, we introduced clitics into the utterances in Experiments 3 and 4. Experiments 3 and 4 tested naming latencies for utterances containing the original four conditions, with the added feature of an auxiliary, e.g. *dishcloths are clean*. We had hypothesised that, if English compounds are really acting as single, recursive prosodic units, then the auxiliary should reduce and attach leftwards to the entire unit in normal speech, e.g. *dishcloths're nice*. The results were consistent with those in Experiments 1 and 2: in the delayed task, utterances containing cliticised compounds elicited similar naming latencies to utterances containing morphologically-simple words; i.e. latencies for *degrees're* and *dishcloths're* were no

different. However, latencies were significantly longer for the phrasal conditions as in *dark cloths're*.

Under online task conditions (Experiment 4), speakers once again generated similar naming latencies for utterances containing cliticised compounds to those containing cliticised disyllabic words. Latencies for the adjective-noun phrases were significantly shorter (30 ms), indicating that speakers were only encoding the first (uncliticised) unit of the phrase; i.e. *dark* in *dark cloths're*. Here, results again suggested that planning was restricted to the size of first phonological unit: in this case, naming latencies for noun-noun compounds were similar to disyllabic monomorphemic words, while adjective-noun phrases generated the shortest naming latencies (due to the fact that speakers were only accessing the first unit of the phrase- the monosyllabic adjective). Recall that the first prosodic unit in the utterance had three sizes.

Table 22: *Size of the Prosodic Unit in **Experiment 4***

Condition	First PWd Unit	Size of First PWd Unit
Compound + Clitic	(doorways are)$_\omega$	(σ σ σ)
Adj-N + Clitic	(dark)$_\omega$	(σ)
Disyllabic + Clitic	(donkeys are)$_\omega$	(σ σ σ)
Monosyllabic + Clitic	(dates are)$_\omega$	(σ σ)

The condition with the smallest phonological word (*dark* in the phrasal condition) generated the fastest response times, followed by the monosyllable + clitic condition (*dates are*), and then the two larger compound + clitic (*dishcloths are*) and disyllable + clitic structures (*donkeys are*). This suggested that the auxiliary was attaching leftwards to the entire prosodic

unit of the compound, yet only attaching to the noun of the adjective-noun phrase in the phrasal condition.

The results of both groups of experiments revealed important information about phonological encoding and prosodification in native English speakers. We confirmed that speakers were indeed generating prosodic frames according to phonological, not lexical, words. Furthermore, the difference in naming latencies between compounds and phrases indicated that the prosodic frames for the complex morphosyntactic structures of compounds were constructed according to the outer, recursive phonological word structure of the compound. This is remarkable as it reveals both experimental evidence for the planning of prosodic structures in English, and evidence for the makeup of the structure itself.

Both experiments indicated that compounds were being planned as a single unit, generating significantly-different mean latencies to the adjective-noun phrases. Critically, speakers in the online tasks continued to treat compounds as single, disyllabic units, suggesting that these items retained their prosodic structure even when planning time was taken away. Taken together, the results from all four experiments indicate that compounds, which contain two morphological words and therefore two phonological words, are nevertheless treated as single prosodic units for the purposes of postlexical encoding during speech production. In summary, our findings require the development of the classic model of phonological encoding to incorporate a distinction between lexical and postlexical prosodic processes, such that lexical boundaries have consequences for post-lexical phonological encoding.

Our claim raises many issues concerning the size of recursive structures both lexically and post-lexically. The current study has focused on constraints for minimal prosodic units at both levels. However, the issue of maximal units remains. For example, in agglutinating

languages such as Turkish and Finnish, a suffixed word can easily contain six to seven syllables and even English has been claimed to have affixed words of considerable length, such as *antidisestablishmentarianism* (cf Payne, 2006). Similarly compounds, which are recursive in the lexicon, are not limited to the two word structures we have examined here. Theoretically, longer compounds are possible such as *hazelnut-butterknife,* where the structure would be as in examples 8a or 8b:

(8) *Hazelnut butterknife*

    (a) **Hazelnut-butterknife** (a butter knife for cutting hazel nuts):

$$(((\text{hazel})_\omega \ (\text{nut})_\omega)_\omega \ ((\text{butter})_\omega \ (\text{knife})_\omega)_\omega)_\omega$$

    (b) **Hazelnut-butterknife** (a knife for spreading hazelnut butter):

$$((((\text{hazel})_\omega \ (\text{nut})_\omega)_\omega \ (\text{butter})_\omega)_\omega \ (\text{knife})_\omega)_\omega$$

The question therefore remains, to what extent processing limitations, rather than the structure of linguistic units per se, constrain the recursivity of both lexical and postlexical prosodic units in connected speech production.

Acknowledgments

References

Anderson, S. R. (2005). *Aspects of the Theory of Clitics*. Oxford University Press, USA.

Baayen, R., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390– 412.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (in press). Fitting linear mixed-effects models using lme4.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2):344–380.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Eigen, C., and Rcpp, L. (2015). Package 'lme4'. *Convergence*, 12:1.

Bien, H., Levelt, W., and Baayen, R. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17876–17881.

Boersma, P. and Weenink, D. (2013). Praat software. *Amsterdam: University of Amsterdam*.

Fox, J., Bates, D., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., Ogle, D., et al. (2009). "Car": Companion to applied regression, R Package version 1.2-16.

Cholin, J., Schiller, N., and Levelt, W. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50(1):47–61.

Collins, A. and Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, *46*(1), 178-198.

Damian, M. and Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, 57(2):195–209.

Davis, C. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37(1):65–70.

Delazer, M. and Semenza, C. (1998). The processing of compound words: A study in aphasia. *Brain and Language*, 61(1):54–62.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3):283.

Duñabeitia, J., Perea, M., and Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition*, 105(3):691–703.

Ferrand, L., Segui, J., and Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, 35(5):708–723.

Ferrand, L., Segui, J., and Humphreys, G. (1997). The syllable's role in word naming. *Memory & Cognition*, 25(4):458–470.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100(2):233.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions?. *Cognition*, *96*(3), 263-284.

Finkbeiner, M., Gollan, T. H., and Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition*, 9(02):153.

Griffin, Z. and Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, 38(3):313–338.

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. *Handbook of psycholinguistics*, 21-59.

Hannahs, S. (1995). The phonological word in French. *Linguistics*, 33(6):1125–1144.

Harley, T. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8(3):191–219.

Inhoff, A. W., Briihl, D., and Schwartz, J. (1996). Compound word effects differ in reading, on-line naming, and delayed naming tasks. *Memory & Cognition*, 24(4):466– 476.

Jacobs, C. and Dell, G. (2014). 'hotdog', not 'hot''dog': the phonological planning of compound words. *Language, Cognition and Neuroscience*, 29(4):512–523.

Janssen, N., Bi, Y., and Caramazza, A. (2008). A tale of two frequencies: Determining the speed

of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, 23(7-8):1191–1223.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PloS one*, *7*(3), e33202.

Jescheniak, J. D. and Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4):824.

Jescheniak, J. D., Schriefers, H., and Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):441–454.

Jordan, M. (1990). *Motor learning and the degrees of freedom problem*. Lawrence Erlbaum Associates, Inc.

Lange, V. M., & Laganaro, M. (2014). Inter-subject variability modulates phonological advance planning in the production of adjective-noun phrases. *Frontiers in Psychology*, *5*, 43.

Lange, V. M., Perret, C., & Laganaro, M. (2015). Comparison of single-word and adjective-noun phrase production using event-related brain potentials. *Cortex*, *67*, 15-29.

Lahiri, A. and Plank, F. (2010). Phonological phrasing in Germanic: the judgement of history, confirmed through experiment. *Transactions of the Philological Society*, 108(3):370–398.

Levelt, W. (1989). *Speaking*. MIT Press.

Levelt, W. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1):1–22.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–75.

Levelt, W. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98(23):13464–13471.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249-336.

MacKay, D. (1987). Constraints on theories of sequencing and timing in language perception and production. In *Language perception and production: Relationships between listening, speaking,*

reading and writing, volume 3, page 407. Academic Press London.

Mahon, B., Costa, A., Peterson, R., Vargas, K., and Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):503.

Meyer, A. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29(5):524–545.

Meyer, A. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30(1):69– 89.

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

Meyer, A. and Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(6):1146.

Morgan, J., and Wheeldon, L. "Syllable monitoring in internally and externally generated English words." *Journal of Psycholinguistic Research* 32.3 (2003): 269-296.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*, volume 28. Dordrecht: Foris.

Oldfield, R. and Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4):273–281.

Payne, T. (2006). *Exploring language structure: a student's guide*. Cambridge University Press.

Peperkamp, S. (1997). *Prosodic words*, volume 34. Holland Academic Graphics: The Hague.

Reetz, H. (2010). "voicekeyhilary.pl". [Perl Computer program].

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, pages 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, 47(1):59–87.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3):249–284.

Roelofs, A. and Meyer, A. S. (1998). Metrical structure in planning the production of spoken

words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4):922.

Saran, F. (1907). *Deutsche Verslehre*, volume 3. CH Beck'sche verlagsbuchhandlung. O. Beck.

Schiller, N., Costa, A., and Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. *Papers in Laboratory Phonology*, 7:35–59.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-prime computer software and manual. *Pittsburgh, PA: Psychology Software Tools Inc*.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, 45(1):33–54.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4):841.

Schriefers, H. and Teruel, E. (1999). The production of noun phrases: A cross-linguistic comparison of french and german. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, pages 637–642.

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11(3):563–605.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, 3(01):371– 405.

Selkirk, E. (1996). The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, *187*, 214.

Sternberg, S., Monsell, S., Knoll, R., and Wright, C. (1978). The latency and duration of rapid movement sequences: comparisons of speech and typewriting. In *Information Processing in Motor Control and Learning*, pages 117–152. Academic Press.

Sweet, H. (1886). *Elementarbuch des gesprochenen Englisch (Grammatik, Texte und Glossar)*. Clarendon Press.

Vigário, M. (2010). Prosodic structure between the prosodic word and the phonological phrase: Recursive nodes or an independent domain? *The Linguistic Review*, 27(4).

Wheeldon, L. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37(3):356–381.

Wheeldon, L. and Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition*, 85(2):B31–B41.

Wheeldon, L. R. and Levelt, W. J. (1995). Monitoring the time course of phonological encoding.

Journal of Memory and Language, 34(3):311–334.

Wheeldon, L. and Morgan, J. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes*, 17(5):503–535.

Zwicky, A. M., & Pullum, G. K. (1983). Cliticization vs. inflection: English n't. *Language*, 502-513.

Appendix A Target Items for Experiment 1

Noun-Noun Compounds	Adj-N Phrases	Simple Initial Stress	Simple Final Stress
daybreak	dark break	denim	default
daytime	dark time	debit	deceit
daylight	dark light	dagger	decree
groundhog	green hog	gravel	gazelle
groundwork	green work	griddle	gazette
groundnut	green nut	gravy	guitar
nightcap	nice cap	nibble	neglect
nightgown	nice gown	nickel	noblesse
nightmare	nice mare	nitrate	notate
lockjaw	long jaw	locker	lament
locksmith	long smith	locust	latrine
locknut	long nut	lodger	lapel
bathrobe	bad robe	banter	baboon
bathroom	bad room	ballad	balloon
bathtub	bad tub	basil	bamboo

Appendix B Target Items for Experiment 2

Condition 1: Compounds	Condition 2: Phrases	Condition 3: Disyllabic	Condition 4: Monosyllabic
daybreak	dark break	denim	dame
daytime	dark time	debit	date
daylight	dark light	dagger	daze
groundhog	green hog	gravel	grease
groundwork	green work	griddle	greed
groundnut	green nut	gravy	greet
nightcap	nice cap	nibble	nine
nightgown	nice gown	nickel	Nile
nightmare	nice mare	nitrate	ninth
lockjaw	long jaw	locker	lob
locksmith	long smith	locust	log
locknut	long nut	lodger	loft
bathrobe	bad robe	banter	bag
bathroom	bad room	ballad	band
bathtub	bad tub	basil	bat

Appendix C Target Items for Experiments 3 and 4

Condition 1: Compounds	Condition 2: Phrases	Condition 3: Disyllabic	Condition 4: Monosyllabic
doorways	deep ways	daisies	doves
dishcloths	drab cloths	dolphins	ducks
dustpans	dark pans	donkeys	drapes
graveyards	green yards	gospels	grapes
grandstands	grey stands	griddles	graphs
grindstones	good stones	goblins	gloves
nightgowns	nice gowns	noodles	nuns
neckties	neat ties	napkins	nets
nightshirts	new shirts	nickels	nodes
lampshades	low shades	lemons	leeks
logbooks	late books	lanterns	lungs
lipsticks	large sticks	lions	lanes
bookshops	big shops	barrels	bowls
bathtubs	bright tubs	blankets	blades
ballrooms	blue rooms	bankers	brooms

Appendix D Experimental questions and sample responses for Experiments 3 and 4

Question	Sample Response
What are good?	Dustpans are good. Nice gowns are good. Lemons are good. Bowls are good.
What are nice?	Lampshades are nice. Bright tubs are nice. Noodles are nice. Graphs are nice.
What are clean?	Dishcloths are clean. Green yards are clean. Napkins are clean. Lungs are clean.
What are dry?	Bathtubs are dry. Low shades are dry. Griddles are dry. Nets are dry.
What are big?	Graveyards are big. New shirts are big. Barrels are big. Lanes are big.