

Hilary Wynne, Linda Wheeldon and Aditi Lahiri

Planning complex structures in a second language: Compounds and phrases in non-native speech production

Abstract: Models of bilingual language production (e.g. Roelofs and Verhoef 2006) assume activation of the native language (L1) during speech. This assumption is tied to evidence from psycholinguistic tasks where non-native speakers regularly exhibit the cost of speaking in the non-dominant language. One of the most obvious signs of a non-native (L2) English speaker is incorrect stress assignment in polysyllabic words. This phenomenon has often been ascribed to interference from the L1 metrical structure on the construction of L2 word frames (cf. Archibald 1993, 1997). A major claim of L2 production models is that a word's prosodic and segmental components are activated separately. If this is true, then errors in stress assignment should not interfere with the construction of L2 word frames. This study presents a series of psycholinguistic experiments designed to test the production of complex constructions in English by fluent speakers of a language with similar compounding strategies but different word stress: Bengali. Over the course of these experiments, we found evidence that the planning unit for both complex and simple words retained a prosodic shape in non-native speakers of English. Furthermore, we found that native Bengali speakers, despite difficulties in assigning correct stress, were able to distinguish between compounds and phrases across all four experiments. Specifically, they were able to access the prosodic shapes of both complex structures despite errors in stress assignments. These findings support models of phonological encoding in which post-lexical information is prepared during separate subprocesses and suggest that the learnability of complex structures such as compounds in English involves the distinct preparation of prosodic and segmental components.

Keywords: Bilingualism, language production, prosody, compounds, phonological encoding

1 Introduction

1.1 Background

It has been argued that the single most obvious feature of non-native speech is prosody (Hatch 1983; Dickerson 1989). In linguistic terms, prosody encompasses the “suprasegmental” aspects of a language such as stress, intonation and rhythm (Fox 2000; Lehiste 1970). One of the most noticeable signs of a non-native (L2) English speaker is incorrect stress assignment in polysyllabic words: e.g. *'agenda* for *a'genda* in native (L1) Hungarian speakers (cf. Archibald 1998), *'gazelle* for *ga'zelle* in Bengali L1 speakers, or *Dar'jeeling* in English L1 speakers for Bengali [*dardzilin*].¹ This phenomenon has often been ascribed to interference from the L1 metrical structure on the construction of L2 word frames (cf. Archibald 1993, 1997).

In language production models, the preparation of prosodic features is assigned to a process known as phonological encoding. Unlike earlier models of language production which treated phonological encoding as a single streamlined process, contemporary L1 models (Levelt et al. 1999; Roelofs 1997) view the phonological encoding (post-lexical) level as one that consists of intermediate (“sub”) stages; that is, there is a dedicated sub-stage which first prepares the metrical information for an utterance, then prepares the prosodic frames of the utterance.

Although the modelling of L1 phonological encoding has received much attention in recent years, modelling of this process in L2 speech remains relatively unexplored. The preparation of prosodic frames of complex words such as compounds and phrases, in particular, can reveal much about how L2 speakers treat the substages of the phonological encoding process. This study presents a series of psycholinguistic experiments designed to test the production of English compounds and phrases by fluent speakers of a language with similar compounding strategies but different word stress: Bengali.

We investigate two main hypotheses. The first hypothesis is concerned with the distinction between two different prosodic structures: compounds and phrases. English compounds are often identified by special semantic, morphological, and phonological criteria, i.e. headedness, visual concatenation, pluralization marking, and stress (cf. Bauer 2009). One of the key criteria for distinguishing compounds from phrases in English is stress: the hard and fast rule is that compounds have primary stress on their initial constituent (e.g. *'blackboard*, a board for writing upon), while phrases have primary stress on the final constituent (e.g. *black'board*, a wooden board that is painted black) (cf. Liberman and Sproat 1992;

¹ Primary stress will be marked throughout this chapter by use of stress marks.

Marchand 1969). However, this criterion is not without exception; in addition to *Madison* 'Avenue'-type compounds (which feature primary stress on the second constituent), there are word pairs in which the compound-phrase stress distinction is barely discernible, e.g. *Christmas cake* – *Christmas pudding* in British English (cf. Giegerich 2015).

We maintain that the preparation of metrical information is distinct from that of prosodic structure in the phonological encoding process: critically, we expect L2 speakers, if fluent enough, to access the correct prosodic framing for compounds and phrases despite errors in stress assignment and that this will be reflected in reaction time data from psycholinguistic experiments. If L2 speakers of a language with different word stress are still able to access the correct prosodic structure of compounds and phrases and experimental evidence from phonological encoding supports this assumption, then we will establish an additional measure that prosodic structure mapped onto to syntax is not isomorphic with lexical word stress.

Our second hypothesis is concerned with the behavior of function words with the phonological word units. A prominent feature of regular, connected speech in English is the encliticization of function words (e.g. *Tim's sick*) (cf. Lahiri and Plank 2010). The question we ask here is thus: do L2 speakers of English exhibit the same behavior in regards to encliticization with compounds and phrases as native speakers? That is, is the prosodic distinction between compounds and phrases maintained? If this is the case, then this should also be evident in reaction time data. This study is arranged as follows: first we will discuss empirical and experimental evidence for the phonological word as a planning unit in native language production. We will then discuss models of L2 phonological encoding and will also further motivate our choice of using native speakers of Bengali for this study.

1.2 The unit of planning: Phonological words

Modern speech production models (e.g. Levelt et al. 1999; Roelofs 1997) maintain that the phonological encoding processes in language production prepare utterances in the form of prosodic, not lexical, units. These units, known as “phonological words” are defined in prosodic constituent theory as lexical words (minimally composed of at least one stressed foot) combined with any number of unstressed items (e.g. function words). Consequently, the number of lexical words is often different than the number of phonological words in an utterance²:

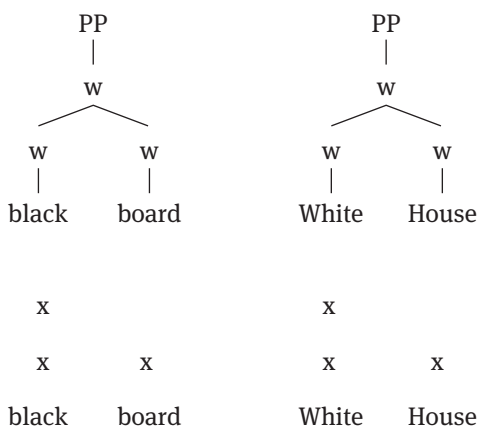
² We use the ω symbol to notate prosodic unit (phonological word) boundaries and φ for phrase boundaries consistently throughout this chapter.

- (1) a. lexical units: [Tim]_N [is]_V [sick]_{Adj}
 prosodic units: ('Timz)_ω ('sick)_ω
 b. lexical units: [Drink]_V [the]_{Art} [juice]_N
 prosodic units: ('drɪŋkðə)_ω ('juice)_ω

In the examples above (1a and 1b), the function words *is* and *the* reduce and attach to the lexical word to their left, forming (respectively) the single phonological words: [tmz] and [drɪŋkðə]. Therefore, both examples contain three lexical words but only two phonological words.

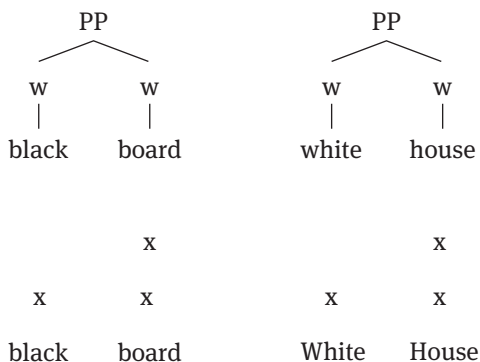
English compound words (e.g. *blackboard*) consist of two or more lexical units (e.g. *black* and *board*); therefore, by definition, they also consist of two prosodic units. When combined, the compound *blackboard* becomes a single prosodic unit taking primary stress on its first constituent ('blackboard) and inflection as a whole (e.g. **blackboards* or **nightsgowns*).

- (2) compounds “blackboard” and “White House”:



Compare the structures in (2) to those in (3) below: the same lexical items, when appearing in phrasal constructions, have a different stress assignment (end-stress) as well as a different prosodic structure: two distinctive prosodic units.

(3) phrases “black board” and “white house”:



Although evidence for the structure of compounds at the lexical level of language production is mixed (cf. Lorenz and Zwitserlood 2016; Lüttman et al. 2011), findings from psycholinguistic tasks that tap into the phonological encoding level point unequivocally towards the treatment of compounds as single prosodic units (Wheeldon and Lahiri 2002; Wynne et al. 2018). The strongest evidence for this comes from several naming tasks in which participants were required to prepare utterances containing noun-noun compounds (e.g. *nightgown*), adjective-noun phrases (e.g. *nice gown*) and monomorphemic words (e.g. *nickel*, *neglect*, or *net*). In the first task, participants heard the question *What was it?* and had to respond with the visually-presented word after three time periods interpolated by beeps (e.g. 800 ms, 1200 ms, and 1400 ms). Naming latencies in this task reflected the number of phonological words in the utterance to be produced (Wheeldon and Lahiri 1997). This difference in naming latencies between compounds and phrases was maintained even when speakers did not have time to plan their utterance in full, i.e. in an online naming task where participants were told to respond immediately after hearing the question *What was it?* In this task, naming latencies were a function of the size of the initial phonological word.

When submitted to a more complex question-answer paradigm in which participants heard a variety of questions (e.g. *What are clean?* *What are nice?*), the results once again showed that reaction times for the utterances reflected prosodically-shaped planning units. Crucially, in the online planning task, we found evidence that the auxiliary (*are*) was reducing and attaching leftwards to the compounds to form single prosodic units (e.g. (*nightgowns're*)_ω), while for the phrases the auxiliary only attached to the closest prosodic unit (e.g. (*shirts're*)_ω in *new shirts're*).

Again, naming latencies were a function of the size of the initial phonological word to be produced: in the case of the phrasal condition (*new shirts're*), the

latencies reflected only the first, mono-syllabic word of the utterance: *new*. The monosyllabic word condition (*nets're*) elicited slightly longer latencies than the phrasal condition, reflecting planning of a bisyllabic phonological word: *nets're*.

Crucially, responses for utterances containing compounds (e.g. *nightgowns are clean*) reflected that the auxiliary word reduced and attached leftwards to compounds to form single prosodic units (as in example 4 below).

- (4) Prosodic structure of compounds with an auxiliary
 Compound structure: ((night)_ω(gowns)_ω)_ω
 Compound + *are*: (((night)(gowns))are)_ω (clean)_ω

In the utterances containing the phrases, naming latencies revealed that the auxiliary only attached to the closest prosodic unit (as in example 5 below).

- (5) Prosodic structure of phrases with an auxiliary
 Phrasal structure: (nice)_ω (gowns)_ω
 Phrase + *are*: (nice)_ω ((gowns)are)_ω (clean)_ω

This result was replicated with semantically-opaque compounds (e.g. *toadstool*) and spaced compounds (e.g. *time zone*) (Wynne et al. in review). Therefore, it appears that, regardless of how compounds are built and/or accessed at the lexical encoding level, they are treated as single prosodic units at the phonological encoding level, as reflected by the significantly different naming latencies for the two structures in the production tasks cited above. Now that we have established how native speakers of English encode compound words for the purposes of production, we now turn to the planning of complex morphosyntactic structures in a second language.

1.3 Planning in a second language

There is experimental evidence that, when speaking a second language, L2 speakers are slower than L1 speakers to name pictures (Hanulová et al. 2011; Kroll et al. 2005; Roberts et al. 2002), more likely to make errors (Ecke 2004), and exhibit an overall lag in speech production (Costa & Santesteban 2004). Crucially, the cost of speaking in one's L2 does not diminish when high levels of proficiency are reached (Sholl et al. 1995). Age of acquisition, word frequency, and inhibition mechanisms have all been investigated as sources of this delay; however, the delay has not been attributed to any single production process. As such, L2 production models have varied greatly in structure: some (cf. Pavio and Desrochers 1980; Potter et al.

1984) have assumed an entirely dual system with separate rules, sets of phonemes, and words for each language, while others (cf. De Bot 1992, 2004; Green 1998; Poullisse and Bongaerts 1994; Roelofs and Verhoef 2006) propose models in which the L2 can interact with L1 at any number of stages during production.

Although the claims made in the linguistic literature differ widely as to the structure of the bilingual mental lexicon, most seem to agree that, although initial specification of the target language occurs early in the production process, L1 encoding processes are still active alongside the L2 processes in later stages of language production. Roelofs and Verhoef (2006) argue that phoneme representations, at the very least, are certainly shared between languages in L2 speakers. Their “WEAVER++” model introduces control processes: when speaking in an L2, the speaker places a goal symbol in working memory which specifies speaking in the L2, and this goal symbol is activated at every stage of production. Thus, both L1 and L2 lemmas are active throughout the production process, with control nodes at each level that remind the processor to employ the L2.

One of the claims of WEAVER++ is that word forms activate a word’s prosodic and segmental components separately, on the basis of contextual information for the utterance, and the prosodic component (the prosodic frame) is the phonological word frame. If there is indeed activation of non-target (L1) information during L2 phonological encoding, then it follows that there will also be activation of L1 components at these separate levels. In order to investigate whether prosodic and segmental information is indeed being activated separately, we need data from speakers of languages that differ in word and phrasal stress but have a similar compound structure.

1.4 Bengali

Compounds composed of two or more noun stems are common in Bengali. Like English compounds, the Bengali compound is composed of individual units that have their own stress and meaning when isolated, as in [ˈd̪uḍʰ] (*milk*) or [ˈbʰaṭ̪] (*rice*). When compounded, stress falls on the first syllable of the structure (cf. example 6).

- (6) a. ˈd̪uḍʰ bʰaṭ̪
 milk rice
 ‘rice mixed with milk’

As in English, Bengali compounds differ from noun phrases in both semantic and phonological structure; this is illustrated in the placement of the emphatic marker (clitic) =i (“only”) which attaches to the end of a phonological word. In example (7a) below, the emphatic clitic =i (“only”) attaches to the compound [matʰer matʰa], (lit. “fish-heads (for eating)”) as a single unit since the two words combine to a single phonological word. Conversely in example 7b, [matʰer matʰa] acts as a phrase (“heads of fish”), with the emphatic clitic attaching to the first element: [matʰer]=i.

- (7) a. fuḍʰu ((matʰer)_ω (matʰa)_ω)_ω=i kʰa-i
 only fish-GEN.PL head=CL eat-1SG
 ‘I really only eat fishheads’
 b. fuḍʰu (matʰer=i)_ω. (matʰa)_ω jɔl-e dekʰ-l-am
 only fish- GEN.PL=CL head water-OBL see-PAST-1SG
 ‘I saw only heads of fish (poking out of) the water’ (i.e. not heads of other creatures)

Stress in polysyllabic Bengali words is overwhelmingly word-initial (cf. Chatterji 1926; Hayes and Lahiri 1991). In phonological phrases (example 8a), the left-most non-clitic word is the most prominent, as in '[kalo] '[tʃul] ‘black hair’, unlike English phrases (example 8b), which often display iambic stress: *black 'board*.

- (8) a. (('tomar kalo)_ω ('tʃul)_ω)_φ
 you.GEN black hair
 ‘your black hair’
 b. ((black)_ω ('board)_ω)_φ

In both languages, main stress in compounds falls on the first prosodic word: in Bengali, this is always the first syllable; however, this can differ in English depending on individual words.

- (9) a. (('ḍuḍʰ)_ω (bʰaṭ)_ω)_ω
 milk rice
 ‘rice mixed with milk’
 b. (('black)_ω (board)_ω)_ω

Thus, Bengali and English, which are both Indo-European languages, have similar compound formation and employ similar cliticization of function words, but differ in terms of word and phrasal stress.

2 Goals and predictions

Our predictions for this study hinge upon evidence from existing English production experiments, as well as Bengali prosodic phonology. Overall, we predict that L1 Bengali speakers, if they are proficient enough in English, will have no difficulty differentiating between compounds and phrases. If prosodic and segmental components are indeed being activated separately during phonological encoding, then errors in stress assignment should not interfere with the construction of L2 word frames. That is, even if the speakers place incorrect stress on the targets (e.g. *'dark,time* or *'de,ceit*), they will be able to successfully differentiate between compounds and phrases based on their prosodic shape. We will begin by addressing hypotheses for Experiments 1 and 2, which were simple naming tasks containing compounds, phrases, and monomorphemic words.

2.1 Predictions for Experiments 1 and 2

Experiment 1 of this study tests planning of compounds, phrases and monomorphemic words by highly-fluent L2 English speakers whose L1 is Bengali. We hypothesize that highly-fluent speakers will have no trouble differentiating between the prosodic structures of compounds and phrases: this will be evident from the reaction time data. That is, reaction times for compounds and phrases will be significantly different from another another, with compounds eliciting reaction times similar to those for bisyllabic monomorphemic words. In this respect, findings will be similar to those in Wynne et al. (2018). Considering the Bengali stress rule, we predict an increase in stress errors for the phrasal and bisyllabic final-stressed word conditions, however we do not expect this to interfere with the preparation of prosodic frames.

In Experiment 2 (an online naming task), the stimuli also consist of four conditions: noun-noun compounds, adjective-noun phrases, bisyllabic monomorphemic words with initial stress, and monosyllabic monomorphemic words. Very little evidence exists for production in online task conditions in L2 speakers. Ahmadian et al. (2012) found that, when performing a task with an unstructured storyline (that is, a task that has a beginning, middle, and end) under online planning conditions, L2 speakers scored lowest in all three of their test areas: fluency, accuracy, and complexity. Consequently, our hypotheses here concentrate on the demands associated both with second language production and the removal of planning time: we hypothesize that, here, highly-fluent L2 English speakers will again be able to differentiate between compounds and phrases. Based on evidence from previous online experiments (Wheeldon & Lahiri, 1997; Wynne et al., 2018), we

expect reaction times to again reflect single prosodic units; however due to lack of preparation time, they will only reflect the **first** prosodic unit of the utterance.

3 Experiments 1 and 2: Production of compounds and phrases

3.1 Experiment 1: Delayed production of compounds and phrases

3.1.1 Method

3.1.1.1 Materials

The method and materials follow those of Wynne et al. (2018). The first experiment in this study consisted of a delayed production task with a simple question-answer paradigm. The experimental materials were constructed from 60 items, 15 per condition type. Four experimental conditions were distinguished: noun-noun compounds, adjective- noun phrases, bisyllabic initially-stressed morphologically-simple words, and bisyllabic final-stressed morphologically-simple words (see Table 1 for an example of the responses and the appendix for the full list of items). Nine morphosyntactically-simple bisyllabic filler words were included: 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	<i>It was daytime.</i>
2	Adjective-Noun Phrase	<i>It was dark time.</i>
3	Stress-Initial Simple Word	<i>It was denim.</i>
4	Stress-Final Simple Word	<i>It was decree.</i>

In order to avoid issues of semantic opacity, all targets were semantically-transparent words.³ Words were arranged into five sets beginning with one of

³ The compound *nightmare* and its matched items in conditions 2, 3 and 4 were removed from the dataset from this experiment and Experiment 2. This was done due to the fact that *nightmare* was not fully transparent in meaning and might therefore be treated differently than the

five phonemes: /d/, /g/, /n/, /l/, and /b/: these phonemes were chosen in order to more easily identify boundaries during analysis. Within the compound condition, the initial compound morpheme contained the same form for each phoneme-organised set (e.g. *daybreak*, *daytime*, *daylight*). The adjective-noun phrases in the phrasal condition were constructed in such a way to discourage speakers from compounding the two words in the phrase: therefore, they included some unique combinations, such as *long jaw*, *dark time*, and *bad tub*. The second morpheme of the compound condition was also identical to the second morpheme of the phrasal condition (e.g. *groundhog* and *green hog*).

The individual morphemes of compounds and phrases were matched for a number of factors that could affect lexical recognition: word-length (number of letters), imageability, and frequency measures where possible (simple CELEX frequency, written, and spoken counts per million as well as log frequency). Word familiarity for all targets was confirmed by rating data from 35 participants using native speaker judgement tasks. Stress of the compounds, bisyllabic initially-stressed words and bisyllabic final-stressed words, was confirmed by 35 native Southern British English participants with questionnaires.

3.1.1.2 Design

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.

3.1.1.3 Procedure

Participants were tested individually. They were seated in a quiet room facing a laptop, wearing headphones that also contained a microphone. They were instructed, in English, that they would see words on the screen and then hear the question, “What was it?”. The experimental question was recorded in a soundproofed booth by a male Southern British English speaker. 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.

other, more-transparent compounds. However, recent evidence for the planning of semantically opaque compounds (Wynne et al. in review) shows no difference in the prosodic treatment of these items and semantically transparent compounds.

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 thus falling into patterns. Participants were instructed to answer as quickly as possible after the third beep.

3.1.1.4 Participants

Eighteen highly-fluent non-native English speakers between the ages of 18 and 22 took part in Experiment 1. Selection for participants was based on their medium of instruction during primary and secondary school levels. In the West Bengal region of India, students follow one of two tracks: English-medium or Bengali-medium. English-medium students receive instruction in English from Class 1 (age 6), while Bengali-medium students receive all of their education in Bengali. The importance of establishing language background for studies on L2 speech is clear: without knowledge of the speakers’ backgrounds, we have no concrete way to define fluency (cf. Grosjean et al. 2004). Therefore, a language background questionnaire based on the LHQ (“Language History Questionnaire”, Li et al. 2006, 2014) was administered. This questionnaire contained questions about age of acquisition (AoA), language use, educational background, and fluency.

The participants were recruited from Gokhale Memorial Girls’ College in Kolkata, India, in March 2012. To test proficiency levels in English, all participants were given a speaking and reading exercise based on the International English Language Testing System (IELTS). The exercise measured coherence, lexical resource (vocabulary), grammatical range, accuracy and pronunciation. The researcher asked the participant to discuss simple “everyday” topics (such as holidays, friends, and family). It also contained a reading component in which the participant was required to read a paragraph aloud. Speakers were able to show a high level of oral proficiency with good pronunciation, a wide vocabulary, ease of speaking, and wide grammatical range. No participant scored lower than IELTS Band 7 in the speaking task, and most were comfortably within Band 8 or 9. They were given course credit for their participation, and did not report any hearing, speaking, or reading impairments.

3.1.2 Results and discussion

3.1.2.1 Analysis

Participants' responses were recorded as audio files on a Sony DTC-100 ES DAT recorder. Response latencies were measured from the offset of the 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.

For mean naming latencies, a mixed-effects model was used.⁴ This was an interceptless mixed effect model such that each condition had a coefficient representing that condition's mean (technically its best linear unbiased estimate). Control variables were entered first and *Preparation Time* was sum-coded. Subjects and items were treated as random factors. To determine the random effect structure, we started with the full random effect structure and proceeded in a step-wise reduction approach. Goodness of fit was established by normality of residuals. We then used t-tests to perform direct pairwise conditions between all conditions.⁵ Following Baayen et al. (2008), all t-values greater than 2 or less than -2 were treated as significant.

As part of this investigation, an error analysis using a generalized linear model (GLM)⁶ was also carried out: errors were categorized 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).

3.1.2.2 Results

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. Responses uttered before the

4 `model <- lmer(rt ~ 0+cond+beep + (0+beep|sub) + (1|word), d, contrasts=list(beep=contr.sum(3)))`

5 **T-test function:** `lmerTest <- function(model, refcoef, comparecoef){varcov <- vcov(model)
t <- (fixef(model)[comparecoef] - fixef(model)[refcoef]) / sqrt(varcov[comparecoef,comparecoef] + varcov[refcoef,refcoef] - 2*varcov[comparecoef,refcoef])}`

6 `glm(error ~ condition + preparation time, family = binomial)`

final beep in the delayed tasks were also discarded. This resulted in a loss of 13.3% of data (9.8% of compounds, 12.8% of phrases, 10.6% of initially-stressed words, and 13.4% of finally-stressed words). Stress errors were not removed. All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a further loss of 32 items (a further 1.1% of the data), resulting in 2,994 total tokens. The resulting mean RTs and mean subject percentage error rates (including stress errors) for each condition and preparation time are shown in Table 2.

Table 2: Naming Latencies (in ms) and percentage subject error rates (in parentheses) are shown for the 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:					
800 ms:	502 (9.0%)	518 (13.6%)	489 (17.7%)	523 (38.6%)	479 (19.8%)
1200 ms:	487 (10.3%)	489 (12.3%)	477 (16.3%)	464 (36.6%)	452 (18.9%)
1400 ms:	468 (10.3%)	489 (12.7%)	473 (16.0%)	476 (35.3%)	448 (18.8%)
Mean RT (ms)	455 (10.1%)	471 (13.2%)	452 (17.0%)	450 (36.9%)	
SD:	24.4	24.4	24.4	24.5	

The model⁷ contained Condition and Preparation time as fixed effects, and revealed a significant effect of Condition ($F=94.86$, $df=4$, $p < .0001^*$) and Preparation Time ($F=19.1$, $df=2$, $p =.004^*$). There was no interaction between Condition and Preparation Time ($F=1.07$, $df=6$, $p=.37$) (see Table 3).

Table 3: Mixed Effects Model of Reaction Time data in Experiment 1 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	2721587	680397	4	68.71	94.862	<.0001*
Prep Time (PT)	109786	54893	2	19.1	7.6533	0.004*
Cond*PT	46166	7694	6	2958.33	1.0728	0.376

7 $rt \sim 0+cond+beep + (0+beep|sub) + (1|word)$, d, contrasts=list(beep=contr.sum(3))

The results of paired t-tests revealed significant differences in the naming latencies between phrases and compounds ($t=3.61^*$), phrases and monomorphemic bisyllabic initially-stressed words ($t=-3.52^*$), and phrases and monomorphemic bisyllabic final-stressed words ($t=-3.41^*$). There was no significant difference in naming latencies for any other comparison (all t s between .08 and .16): compounds, monomorphemic bisyllabic initially-stressed words and monomorphemic bisyllabic final-stressed words all had statistically similar reaction times. The shortest preparation time (800 ms) resulted in significantly slower responses ($t=-3.83^*$) than the 1000 ms preparation time. All four frequency measures were tested⁸; no effect of any combination of compound word frequency was found in these analyses (see Table 4).

Table 4: Analysis of Frequency*Condition Interactions in Experiment 1 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
simple word frequency	32259	8065	4	2961.35	1.0977	0.35592
left constituents	22136	3689	6	2958.28	0.5003	0.80855
right constituents	29341	4890	6	2958	0.6636	0.6792
sum of constituents	36850	6141.7	6	2958	0.8335	0.543791

Error rates were analyzed using a generalised linear model (GLM)⁹ with a binomial distribution (link= logit) (see Table 5). The analysis revealed an effect ($\chi^2= 311.30$, $p= <.0001^*$) of *Condition* on error rate: Condition 4 (the bisyllabic final-stressed word condition) resulted in a significantly higher number of errors than the initially-stressed word conditions (all comparisons yielded $t > 3.67$). There was no

Table 5: Generalized Linear Analysis of Errors in Experiment 1 (link=logit, distribution= binomial).

	Estimate	Std. Error	z value	Pr(> z)
Condition	0.84242	0.05501	15.315	<.0001*
Prep Time	-0.09276	0.06386	-1.453	0.146

⁸ e.g. `rt ~ 0+cond*beep*sum frequency + (0+beep|sub) + (1|word)`, `d, contrasts=list(beep=-contr.sum(3))`

⁹ `glm(error ~ condition + preparation time, family = binomial)`

effect of *Preparation Time* ($\chi^2=1.97$, $p=0.15$). Aside from stress errors, the most frequent error that the non-native speakers made was “wrong item”.

Replacements for these words were often words with higher frequency that shared initial phonological segments with the intended target:

- (10) a. *bell* for *ballad*
b. *lemon* for *lament*
c. *bad road* for *bad robe*

3.1.2.3 Discussion

Similar to our findings for L1 speakers of English (Wynne et al. 2018), naming latencies for this experiment reflected the total number of prosodic units in the utterance. Reaction times revealed that utterances containing phrases took significantly longer to prepare than those containing compounds or monomorphemic words, suggesting that the speakers were able to recognise compounds as single prosodic units. Condition 4 (the bisyllabic final-stressed words) resulted in a significantly higher number of errors: specifically, stress errors. Nevertheless, reaction times for this condition did not differ from those for compounds or initially-stressed monomorphemic words.

3.2 Experiment 2: Online production of compounds and phrases

3.2.1 Materials and design

The experimental materials were constructed from 60 items, 15 per condition type. Four experimental conditions were again distinguished: noun-noun compounds, adjective-noun phrases, bisyllabic initially-stressed morphologically-simple words, and monosyllabic morphologically-simple words (see Table 6 for an example of the responses and the Appendix for the full list of items).

Table 6: Experimental Condition Types and Sample Responses.

Condition Type	Sample Response
1 Noun-Noun Compound	<i>daytime</i>
2 Adjective-Noun Phrase	<i>dark time</i>
3 Bisyllabic Simple Word	<i>denim</i>
4 Monosyllabic Simple Word	<i>dart</i>

Since evidence from previous online production tasks (Wheeldon & Lahiri 1997; Wynne et al. 2018) indicates that speakers only plan the first prosodic unit of the utterance when preparation time is limited, we had to ensure that the control conditions would contain corresponding structures. Consequently, a monosyllabic word condition (e.g. *dart*) replaced the final-stressed bisyllabic word condition used in Experiments 1 and 2. The rest of the target items remained the same. 15 new filler words were added. Word familiarity for the new condition was confirmed by rating data from 35 participants using a native speaker judgement task.

3.2.2 Design and procedure

There were 5 blocks of 15 trials (60 targets + 15 fillers) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and five different versions of the experiment were presented. Experimental stimuli were only presented once in this task in order to avoid word repetition effects (cf. Humphreys et al. 1990).

Participants were instructed to respond as quickly as possible following the auditory question. It was important in this experiment that the target words occurred sentence-initially in the participants' responses: if naming latencies are indeed only reflecting the size of the first prosodic unit, then the intended response of Experiments 1 and 2 (e.g. *it was groundhog*) would result in flat results (as reaction times would only reflect *it was*). Therefore, participants were instructed to produce the target word only (e.g. *groundhog*).

3.2.3 Participants

Twenty-one highly-fluent non-native English speakers between the ages of 18 and 22 took part in Experiment 2. Participants were native speakers of Standard Colloquial Bengali. They were recruited from Gokhale Memorial Girls' College in Kolkata, India, in March 2012. They were given course credit for their participation. They did not report any hearing, speaking, or reading impairments. No participant scored lower than Band 7 in the speaking task, and most were comfortably within Band 8 or 9.

3.2.4 Results and discussion

3.2.4.1 Analysis

Latencies were measured from the end of the [t] plosive of the auditory prompt (*What was it?*) to the onset of the participant's response.

3.2.4.2 Results

Errors (not including stress errors) resulted in a loss of 2.7% of data (2.8% of compounds, 2.8% of phrases, 1.9% of bisyllabic words, and 3.4% of monosyllabic words). All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a further loss of 43 items (a further 3.6% of the data), resulting in N = 1120. The resulting mean RTs and percentage error rates (including stress errors) for each condition and preparation time are shown in Table 7.

Table 7: 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) Bisyllabic	(4) Monosyllabic
PWds:	1	2	1	1
LexWs:	2	2	1	1
Syllables	2	2	2	1
Mean Lat. (ms)	203 (4.8%)	173 (9.3%)	195 (4.3%)	174 (4.0%)
SD:	81	76	80	79

The model¹⁰ contained the fixed effect of *Condition*. There was a significant effect of *Condition* on reaction times.

Table 8: Mixed Effects Model of Reaction Time data in Experiment 2 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	678571	169643	4	22.12	55.86	<.0001*

10 model <- lmer(rt ~ 0+cond+ (0+cond|sub) + (1|word), d

In a pairwise comparison, mean latencies for the phrasal condition differed significantly from compounds ($t=3.87^*$) and bisyllabic words with initial stress ($t=3.03$). They did not differ significantly from the monosyllabic word condition ($t=0.36$). Mean naming latencies for compounds, accordingly, were similar to those for bisyllabic monomorphemic words ($t=1.54$), but not the monosyllabic words ($t=4.47$). Frequency analyses reveal no interaction between any combination of compound frequency and condition (see Table 9).

Table 9: Analysis of Frequency*Condition Interactions in Experiment 2 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
simple	13329	6665	3	54.404	2.1939	0.1212
left	15100	5033	3	53.858	1.6573	0.18711
right	19915	6638	3	53.432	2.1858	0.10041
sum	14056	4685	3	53.577	1.5425	0.21415

Error rates were analyzed using a generalized linear model (GLM)¹¹ with a binomial distribution (link= logit). Although errors appeared higher for Condition 2 (phrases), the analysis revealed no effect of *Condition* on error rate ($\chi^2= 5.56$, $p=.12$).

Table 10: Generalized Linear Analysis of Errors in Experiment 2 (link=logit, distribution= binomial).

	Estimate	Std. Error	z value	Pr(> z)
condition	0.04227	0.06876	0.615	0.539

3.2.4.3 Discussion

In this task, speakers had less time to plan their utterances: consequently, the amount the speaker was able to encode before beginning their utterance was affected. Here, phrases took the shortest amount of time to prepare, generating mean naming latencies similar to that of the monosyllabic words. Compounds took significantly longer to prepare, similar in naming latencies to bisyllabic monomorphemic words. This suggests that, once again, compounds and phrases were

¹¹ glm(error ~ condition, family = binomial)

treated differently. Due to limited planning time, reaction times reflected only the first prosodic unit of the utterance: for phrases, this was the adjective (e.g. *bad* in *bad tub*), and for compounds this was the entire compound (e.g. *bathtub*).

4 Interim discussion

As we discussed in Section 1.2, phonological words often combine with unstressed items such as function words to form a larger phonological word. Following this, we wanted to test whether function words would combine with compound words to form a single phonological word unit. In the following experiments, we alter the naming paradigm to involve more natural questions and responses: in both delayed (Experiment 3) and online (Experiment 4) task conditions, speakers will be required to answer one of five questions (e.g. *what are good?* or *what are nice?*) using a target from one of four conditions. The intention here is to elicit regular, connected speech containing encliticization of the auxiliary verb *are* to the compound or phrase (e.g. *doorways're nice*).

4.1 Predictions for Experiments 3 and 4

In Experiment 1, naming latencies reflected the total number of prosodic units in the utterance. Phrases were treated as two distinctive prosodic units while compounds were treated as single prosodic units, eliciting similar reaction times to monomorphemic words. Experiment 3 builds on this paradigm by increasing the size of the predicted prosodic unit to include an auxiliary verb. As discussed in Section 1.2, auxiliaries and function words often cliticize to neighbouring prosodic words (e.g.: [*tɪmz*] and [*dɪŋkðə*] for *Tim is* and *drink the*). Following this, we predict that reaction times will once again reflect the total number of prosodic units in the utterance, and that a compound, when paired with an auxiliary, will constitute a single prosodic unit.

We predict that Experiment 4, an online experiment, will also elicit similar results; however, reaction times here will only reflect the first prosodic unit in the utterance. As in Experiment 3, the prosodic unit for the compound word will include the auxiliary *are*.

5 Experiment 3: Delayed production of compounds, phrases and clitics

5.1 Method

5.1.1 Materials

In Experiments 3 and 4, we produced a new target wordlist and five different auditory prompts (see Tables 11 and 12 for auditory prompts and the Appendix for the target wordlist). All target items were presented in the plural, e.g. *dishcloths*, and the auditory prompts were designed to invoke the reduction and attachment of the auxiliary word *are* to these targets.

Table 11: Auditory Prompts for Experiments 4 and 5.

What are clean?
 What are good?
 What are dry?
 What are nice?
 What are big?

Table 12: Experimental Condition Types and Sample Responses for the Auditory Prompt “What are clean?”.

Condition Type	Sample Sentence
N-N Compound	<i>dishcloths are clean.</i>
Adj-N Phrase	<i>drab cloths are clean.</i>
Bisyllabic	<i>donkeys are clean.</i>
Monosyllabic	<i>drapes are clean.</i>

The experimental materials were constructed from 60 items, 15 per condition type. Four experimental conditions were distinguished: noun-noun compounds (e.g. *dishcloths*), adjective-noun phrases (*drab cloths*), initially-stressed bisyllabic simple words (*donkeys*), and monosyllabic simple words (*drapes*). We chose monosyllabic words as a fourth control condition due to our findings in Experiment 2 (online planning of compounds), where reaction times indicated that speakers only had time to plan the first phonological word unit: in the case of the phrasal condition, this was the monosyllabic adjective of the adjective-noun combination (e.g. *dark* in *dark time*).

As in Experiments 1 and 2, targets were arranged into five sets beginning with one of five phonemes: /d/, /g/, /n/, /l/, and /b/. The adjective-noun phrases also consisted of more ‘natural’ combinations than those in Experiments 1 and 2, such as *green yards* and *new shirts* (instead of *dark light* and *green hog*). Word familiarity was confirmed by the same method as for Experiments 1 and 2: a judgement task was run on 36 native Southern British English speakers.

5.1.2 Design

There were 10 blocks of 19 trials in total (190 items total including the ten 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 and blocks were rotated.

5.1.3 Participants and procedure

Twenty participants between the ages of 14 and 21 took part in Experiment 3. They were recruited from Shri Shikshayatan School in Kolkata, India, in March 2015. These speakers all scored very high on the proficiency tasks; no participant scored lower than Band 8 in the speaking task, and most were comfortably within Band 8 or 9.

5.1.4 Results and discussion

5.1.4.1 Analysis

Response latencies were measured from the offset of the third beep to the onset of the response.

5.1.4.2 Results

Errors resulted in a loss of 4.1% of data (4.0% of compounds, 3.1% of phrases, 3.8% of initially-stressed words, and 4.9% of monosyllabic words). All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a further loss of 159 items (4.6% of the data), resulting in $N = 3311$. The resulting mean RTs and percentage error rates (including stress errors) for each condition and preparation time are shown in Table 13.

Table 13: Naming Latencies (in ms) and percentage subject error rates (in parentheses) are shown for the four conditions of Experiment 3, on average and across the three planning durations (N= 3311).

	(1) Compounds	(2) Phrases	(3) Bisyllabic	(4) Monosyllabic	Mean Lat. (ms)
PWds:	1	2	1	1	
LexWs:	2	2	1	1	
Syllables	2	2	2	1	
Beep Lat.:					
800 ms:	440 (4.7%)	453 (9.1%)	425 (4.9%)	436 (4.3%)	439 (5.8%)
1200 ms:	409 (1.9%)	433 (8.6%)	415 (3.3%)	414 (5.6%)	418 (4.9%)
1400 ms:	400 (3.3%)	427 (9.2%)	406 (2.2%)	406 (3.0%)	409 (4.4%)
Mean Lat. (ms)	416 (3.3%)	438 (9.0%)	414 (3.5%)	419 (4.3%)	
SD:	144	149	142	141	

The model¹² contained *Condition*, *Preparation Time*, and *Prompt* as fixed effects. There was a significant effect of *Condition* as well as *Preparation Time*. There was no interaction between any of the effects, nor was there an effect of *Prompt* on reaction times (see Table 14).

Table 14: Mixed Effects Model of Reaction Time data in Experiment 3 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
cond	1758347	439587	4	25.42	67.2962	<.0001*
beep	430536	215268	2	3145.28	32.9553	<.0001*
prompt	53387	13347	4	2594.39	2.0432	0.08578
cond*beep	19546	3258	6	3145.49	0.4987	0.80975
cond*prompt	58241	4853	12	2593.79	0.743	0.70992
beep*prompt	71242	8905	8	3150.88	1.3633	0.20755
cond*beep*prompt	180955	7540	24	3150.17	1.1543	0.27388

In the pairwise comparison, mean latencies for the phrasal condition differed significantly from compounds ($t=3.90^*$), bisyllabic words with initial stress ($t=4.19^*$) and monosyllabic words ($t=3.02^*$). Compounds were no different in mean naming latencies than both the words with initial stress ($t=-0.05$) or

¹² $rt \sim 0 + \text{cond}*\text{beep}*\text{prompt} + (0 + \text{cond}|\text{sub}) + (1|\text{item})$, d, contrasts=list(beep=contr.sum(3))

monosyllabic words ($t=-0.97$). Furthermore, reaction times for the bisyllabic and monosyllabic words did not differ significantly from one another ($t=1.11$). The shortest preparation time (800 ms) resulted in significantly longer naming latencies than those for both the 1200 ms beep ($t=5.61^*$) and the 1400 ms beep ($t=7.98^*$).

No effect was found for any of the combinations of frequency in these analyses. Models containing interactions between the left compound and phrase constituent frequencies and condition did not converge.

Table 15: Analysis of Frequency*Condition Interactions in Experiment 3 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
simple	663	331	2	52.002	0.0496	0.9517
Left	<i>model containing interaction did not converge</i>					
Right	<i>model containing interaction did not converge</i>					
Sum	4522	1507	3	51.09	0.2257	0.8781

Error rates were analysed using a generalized linear model (GLM)¹³ with a binomial distribution (link= logit). The analysis revealed an effect of preparation time on error rate (the shortest preparation time resulted in significantly more errors, all $ps < .04^*$); however errors were not significant for Condition.

Table 16: Generalized Linear Analysis of Errors in Experiment 3 (link=logit, distribution= binomial).

	Estimate	Std. Error	z value	Pr(> z)
Condition	-0.0080829	0.065233	-0.124	0.90
PrepTime	-0.0010702	0.0003353	-3.191	0.001*

5.1.4.3 Discussion

In this task, speakers consistently took longer to produce utterances containing adjective-noun phrases than any other condition in the task. Naming latencies for the compound condition were much shorter, similar to those for both morphologically-

¹³ `glm(error ~ condition + preparation time+prompt, family = binomial)`

simple word conditions. However, while this experiment shows that L2 speakers are continuing to make a distinction between the compound and phrasal conditions, it does not necessarily present solid evidence about the auxiliary *are*, namely whether it is reducing and attaching to the neighboring word. Even if the auxiliaries are reducing and attaching, we cannot gauge which direction (left or right) they fall: the total number of prosodic units predicting the naming latencies remains the same regardless of which unit the auxiliary is attaching to:

Table 17: Auxiliary behaviour in connected speech.

	Compounds	PW	Phrases	PW
Left attachment:	(lampshades are) _ω (nice) _ω	2	(low) _ω (shades are) _ω (nice) _ω	3
Right attachment:	(lampshades) _ω (are nice) _ω	2	(low) _ω (shades) _ω (are nice) _ω	3
No attachment:	(lampshades) _ω (are) _ω (nice) _ω	3	(low) _ω (shades) _ω (are) _ω (nice) _ω	4

Recall that the online task in Experiment 2 generated reaction time information about the **size of the first prosodic unit**: in our findings, for compounds, this was the entire unit; for phrases, only the first phonological unit (the adjective). If the auxiliary *are* is indeed reducing and attaching leftwards, then this should be evident from the reaction times for the adjective-noun phrases: that is, the latencies for adjective-noun phrases will only reflect encoding of the adjective (e.g. *low* in *low shades're*).

6 Experiment 4: Online production of compounds, phrases and clitics

6.1 Method

6.1.1 Materials, design and participants

Experiment 4 presents the target words of Experiment 3 in online task conditions. There were 6 blocks of 12 trials (72 items total including 12 filler words) with optional breaks between the blocks. As in Experiment 2, the experimental stimuli were only presented once in this task. Twenty-five participants between the ages of 18 and 22 took part in Experiment 4. Participants were native speakers of Standard Bengali, all recruited from Shri Shikshayatan School in Kolkata, India,

in March 2015. These speakers all scored very high on the proficiency tasks; most were comfortably within Band 8 or 9.

6.2 Results and discussion

6.2.1 Results

Errors resulted in a loss of 5.9% of data (8.5% of compounds, 5.0% of phrases, 6.2% of initially-stressed words, and 8.1% of finally-stressed words). All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a further loss of 22 items (a further 1.7% of the data), resulting in $N=1317$. The resulting mean RTs and percentage error rates (including stress errors) for each condition and preparation time are shown in Table 18.

Table 18: Naming Latencies (in ms) and percentage subject error rates (in parentheses) are shown for the four conditions of Experiment 4, on average ($N=1317$).

	(1) Compounds	(2) Phrases	(3) Bisyllabic	(4) Monosyllabic
PWds:	1	2	1	1
LexWs:	2	2	1	1
Syllables	2	2	2	1
Mean Lat. (ms)	301 (8.5%)	241 (5.0%)	300 (6.2%)	268 (8.1%)
SD:	102	100	100	108

Table 19: Mixed Effects Model of Reaction Time data in Experiment 4 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	1233395	308349	4	44.72	109.389	<.0001*
Prompt	127507	31877	4	1252.11	11.309	<.0001*
Cond*Prompt	30240	2520	12	1252.32	0.894	0.5527

The model¹⁴ contained the fixed effect factors *Condition* and *Prompt*. Although reaction times differed significantly according to both Prompt and Condition, there was no interaction between the two variables.

¹⁴ reaction time~ 0+condition*prompt + (0+condition | sub) + (1|item).

In a pairwise comparison, mean latencies for the phrasal-plus-clitic condition differed significantly from compound-plus-clitic condition ($t = -5.36^*$), the bisyllabic-plus-clitic condition ($t = 6.29^*$) and monosyllable-plus-clitic condition ($t = 3.09^*$). Compounds were no different in mean naming latencies than the words with initial stress ($t = 1.25$). Bisyllabic words with clitics elicited significantly different naming latencies than the monosyllabic-plus-clitic condition ($t = -3.53^*$).

No effect was found for any of the combinations of frequency in these analyses.

Table 20: Analysis of Frequency*Condition Interactions in Experiment 4 (Type III with Satterthwaite approximation for degrees of freedom).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
simple	41678	3473	12	1178.34	1.2283	0.2576
left	4513	4513	1	603.12	1.6285	0.2024
right	1741	1741	1	308.01	0.5699	0.450865
sum	<i>model containing interaction did not converge</i>					

Error rates were analyzed using a generalized linear model (GLM)¹⁵ with a binomial distribution (link= logit). The analysis revealed no effect of Condition on error rate ($\chi^2 = 2.21$, $p = .51$).

Table 21: Generalized Linear Analysis of Errors in Experiment 4 (link=logit, distribution= binomial).

	Estimate	Std. Error	z value	Pr(> z)
condition	-0.127	0.1998	-0.636	0.525

6.2.2 Discussion

Again, reaction times for this experiment reflect the planning of only the first prosodic unit of the utterance. The naming latencies of the compounds and phrases were, again, significantly different in this task, supporting our claim that these two structures were planned differently. However, compounds were only similar in naming latencies to the bisyllabic words; the mean naming latency for the monosyllabic word condition was significantly faster. These findings differ to those in

¹⁵ glm(error ~ condition+prompt, family = binomial)

Experiment 2, where the mean latency for the monosyllabic word condition was similar to that of the adjective-noun phrase.

Auxiliary reduction (contraction) is widespread with English auxiliary verb forms such as *is* or *are*, particularly when following a plural subject (cf. McElhinny 1993); however, there is evidence that L2 speakers exhibit less verb inflection in English (Meisel 1997; Lardiere 1998). Here, we found that the planning unit is again only a single prosodic unit: for the phrasal condition, it is the adjective (e.g. *low*). For the monosyllabic word condition, we see that the auxiliary *are* is attaching and reducing to the monosyllabic word, e.g. *lamps're*, resulting in a slightly-larger prosodic unit (and hence longer reaction times). For the compounds, the planning unit is the compound + the auxiliary (*lampshades're*), eliciting similar reaction times to the bisyllabic word condition (*lockers're*).

7 Conclusion

The aim of these experiments was to elicit psycholinguistic data about the generation of prosodic frames in non-native speakers of English. We began by examining how fluency affects the planning of compounds, phrases and monomorphemic words in Experiments 1 and 2. Our hypotheses for both experiments hinged upon the speakers' fluency levels: we predicted, on the basis of existing data (cf. Wynne et al. 2018; Wheeldon and Lahiri 1997, 2002), that compounds and phrases would elicit different naming latencies based on their distinctive prosodic shapes. Following this, we also theorized that non-native speakers, despite exhibiting certain markers of non-native speech (e.g. longer overall reaction times and/or more stress errors), would treat compounds and phrases differently in terms of prosodic size and shape.

In Experiment 1, we found that, when given time to plan their utterances, highly-fluent L2 English speakers treated noun-noun compounds as single prosodic units while utterances containing adjective-noun phrases elicited significantly-longer naming latencies. Although we saw a larger number of stress errors (recall that stress in Bengali is overwhelmingly word-initial) for the phrasal condition (e.g. *dark* 'time, *green* 'nut) than the compound (e.g. 'daytime, 'groundnut) condition, reaction times revealed a significant difference between these structures. Models of non-native language production assume that certain aspects of L2 phonological encoding cannot be shared: for example, languages differ in terms of rules of syllabification and stress assignment. Following this, we saw that, while the Bengali speakers were able to recognize the orthographic representation and activate the associated phonology of the English word, their

familiarity with the target word did not extend to stress placement. In most cases, the subjects produced the correct word, only with an incorrect stress pattern (e.g. 'gazelle for ga'zelle). Therefore, the encoding process simply defaults to the stored prosodic shape for bisyllabic words.

Our next step was to test what happened to responses when planning time was taken away in L2 English speakers. There is evidence (cf. Wheeldon and Lahiri 1997) that tasks involving restricted planning time generate different results from those that allow delayed response. We wanted to see how this would affect phonological encoding of compounds and phrases in L2 speech production. In Experiment 2, we found that reaction times reflected planning of only the first prosodic unit of the utterance. For compounds, this was the entire compound word (e.g. *bathtub*); for phrases, it was the mono-syllabic adjective (e.g. only the *bad* in *bad tub*).

Initially we had concerns that the long response latencies for the phrasal condition in Experiment 1 was due to the 'unnaturalness' of the phrases (e.g. *bad tub* or *nice mare*). However, this concern was nullified by the results of Experiment 2: despite the unnaturalness of the phrases, response times were **faster** for the phrasal condition than the compound condition. And while the adjectives used in the phrasal condition did indeed have a high mean word frequency, responses to these targets were not significantly faster than the monosyllabic word condition (which had a low mean word frequency).

Experiments 3 and 4 consisted of a more complex task: here, speakers were required to answer one of five prompts (e.g. *what are good?*) with the target word or phrase (e.g. *bookshops are good* or *big shops are good*). These experiments were designed to invoke reduction and encliticization of the auxiliary verb *are*, in order to investigate if compounds and phrases maintained certain criteria associated with phonological words: namely the reduction and attachment of function words to single prosodic units (e.g. [ˈdɪŋkðə], *drink the*). In Experiment 3, we found that the distinction between compounds and phrases was maintained, at least in the reaction time measurements. However, reaction times here only reflected the total number of prosodic units being planned. Following this, we could not determine whether the auxiliary was in fact reducing.

The real test came in the form of the online task: because speakers were only able to plan the first prosodic unit of the utterance. If the speakers were truly reducing the auxiliary, then this effect would be translated into the naming latencies for each condition. Phonetic analyses of non-native speech patterns have revealed that L2 speakers of English on average produce less syllable reduction (Anderson-Hsieh and Venkatagiri 1994), less reduction and deletion of vowels (Gut 2003), and less function word contraction (Aoyama and Guoin 2007). Therefore, we questioned whether the Bengali speakers would exhibit any reduction. We

found that results for the online task were supportive of both reduction **and** leftwards attachment in the Bengali speakers: the monosyllabic word-plus-clitic (*grapes are*) condition generated significantly longer reaction times than the adjective noun phrase-plus-clitic condition (*green yards are*).

This indicated that speakers were planning the monosyllabic word-plus-clitic condition as a single cliticized prosodic unit (e.g. *grapes're*). The naming latencies for the phrases, on the other hand, reflected a smaller prosodic unit: the monosyllabic adjective (e.g. *green* in *green yards're*). The compound-plus-clitic condition (e.g. *grandstands're*) took significantly longer than both the phrasal-plus-clitic and monosyllabic word-plus-clitic conditions, generating similar reaction times to those for the bisyllabic word-plus-clitic condition (e.g. *griddles're*). Therefore, in both experiments, it appeared that the native Bengali speakers had little trouble producing regular auxiliary deletion in English.

Our findings lend support to the claim that it is in fact the prosodic structure (not the lexical or morphosyntactic structure) of the utterance that is dictating the arrangement of prosodic frames during phonological encoding and that L2 speakers, if able to access the prosodic structure of the language, differentiate between compounds and phrases in English. Over the course of these experiments, we found evidence in all four tasks that the planning unit for both complex and simple words retained a prosodic shape in non-native speakers of English. Furthermore, we found that native Bengali speakers, despite difficulties in assigning correct stress (cf. the final-stress words in Experiment 1 and the phrases in Experiment 3), were able to distinguish between compounds and phrases across all four experiments. Specifically, they were able to plan the prosodic shapes of both complex structures despite errors in stress assignments. Furthermore, these findings support models of phonological encoding in which post-lexical material is prepared during separate subprocesses and suggest that the learnability of complex structures such as compounds in English involves the distinct preparation of prosodic and segmental components.

References

- Ahmadian, Moussa & Samaneh Hosseini. 2012. A study of the relationship between Iranian EFL learners' multiple intelligences and their performance on writing. *Mediterranean Journal of Social Sciences* 3 (1). 111–126.
- Anderson-Hsieh, Janet & Horabail Venkatagiri. 1994. Syllable duration and pausing in the speech of chinese ESL speakers. *TESOL Quarterly* 28 (4). 807–812.

- Aoyama, Katsura & Susan G. Guion. 2007. Prosody in second language acquisition. In Ock-Schwen Bohn & Murray J. Munro (eds.), *Language experience in second language speech learning 17: In honor of James Emil Flege*, 281–298. Amsterdam: John Benjamins B.V.
- Archibald, John. 1993. The learnability of English metrical parameters by adult Spanish speakers. *IRAL: International Review of Applied Linguistics in Language Teaching* 31 (2). 129–143.
- Archibald, John. 1997. The acquisition of second language phrasal stress: A pilot study. In S.J. Hannahs & Martha Young-Scholten. *Language Acquisition and Language Disorders*, 263–290. Amsterdam: John Benjamins Publishing Company.
- Archibald, John. 1998. Metrical parameters and lexical dependency: Acquiring L2 stress. In Suzanne Flynn, Gita Martohardjono & Wayne O'Neil (eds.), *The Generative Study of Second Language Acquisition*, 279–301. Mahwah, NJ: Lawrence Erlbaum Associates.
- Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59 (4). 390–412.
- Bauer, Laurie. 2009. Typology of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 343–356. Oxford: Oxford University Press.
- Bien, Heidrun, Willem J.M. Levelt & R. Harald Baayen. 2005. Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America* 102 (49). 17876–17881.
- Boersma, Paul & David Weenink. 2013. Praat software. Amsterdam: University of Amsterdam.
- Chatterji, Suniti Kumar. 1926. *The Origin and Development of the Bengali Language, Volume II*. Calcutta: D. Mehra, Rupa & Co.
- Costa, Albert & Mikel Santesteban. 2004. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50 (4). 491–511.
- De Bot, Kees. 1992. A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics* 13 (1). 1–24.
- De Bot, Kees. 2004. The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism* 1 (1). 17–32.
- Dickerson, Wayne B. 1989. *Stress in the Speech Stream: The Rhythm of Spoken English*. Urbana: Univ of Illinois Press.
- Ecke, Peter. 2004. Words on the tip of the tongue: A study of lexical retrieval failures in Spanish-English bilinguals. *Southwest Journal of Linguistics*, 23(2), 33–63.
- Fox, Anthony. 2000. *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. Oxford: Oxford University Press.
- Giegerich, Heinz J. 2009. The English compound stress myth. *Word Structure* 2 (1). 1–17.
- Giegerich, Heinz J. 2015. *Lexical Structures: Compounding and the Modules of Grammar*. Edinburgh: Edinburgh University Press.
- Green, David W. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1 (2). 67–81.
- Grosjean, François. 2004. Studying bilinguals: Methodological and conceptual issues. In Tej K. Bhatia & William C. Ritchie (eds.), *The Handbook of Bilingualism*, 32–63.
- Gut, Ulrike. 2003. Prosody in second language speech production: The role of the native language. In Gert Henrici, Frank G. Königs & Ekkehard Zöfgen (eds.), *Fremdsprachen Lehren und Lernen: Mündliche Produktion in der Fremdsprache* 32, 133–152. Tübingen: Gunter Narr Verlag.

- Hanulová, Jana, Douglas J. Davidson & Peter Indefrey. 2011. Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes* 26 (7). 902–934.
- Hatch, Evelyn Marcussen. 1983. *Psycholinguistics: A Second Language Perspective*. Rowley, MA: Newbury House Publishers.
- Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 9(1). 47–96.
- Humphreys, Glyn W., Lindsay J. Evett & Philip T. Quinlan. 1990. Orthographic processing in visual word identification. *Cognitive Psychology* 22 (4). 517–560.
- Kroll, Judith F. & Natasha Tokowicz. 2005. Models of bilingual representation and processing. In Judith F. Kroll & Annette M.B. De Groot (eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*, 531–553. Oxford: Oxford University Press.
- Lahiri, Aditi & Frans Plank. 2010. Phonological phrasing in Germanic: the judgement of history, confirmed through experiment. *Transactions of the Philological Society* 108 (3). 370–398.
- Lardiere, Donna. 1998. Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research* 14 (4). 359–375.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: M.I.T. Press.
- Levelt, Willem J., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–75.
- Li, Ping, Sara Sepanski & Xiaowei Zhao. 2006. Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods* 38 (2). 202–210.
- Li, Ping, Fan Zhang, Erlfang Tsai & Brendan Puls. 2014. Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition* 17 (3). 673–680.
- Liberman, Mark & Richard Sproat. 1992. The stress and structure of modified noun phrases in English. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical Matters*, 131–181. Center for the Study of Language and Information: Leland Stanford Junior University.
- Lorenz, Antje & Pienie Zwitserlood. 2016. Semantically transparent and opaque compounds in German noun-phrase production: Evidence for morphemes in speaking. *Frontiers in Psychology* 7. Article 1943.
- Lüttmann, Heidi, Pienie Zwitserlood, Andrea Böhl & Jens Bölte. 2011. Evidence for morphological composition at the form level in speech production. *Journal of Cognitive Psychology*, 23 (7). 818–836.
- Macnamara, John. 1967. The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior* 6 (5). 729–736.
- Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach*. Munich: Beck.
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen & Douglas Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94. 305–315.
- McElhinny, Bonnie S. 1993. Copula and auxiliary contraction in the speech of White Americans. *American Speech*, 68 (4), 371–399.
- Meisel, Jürgen M. 1997. The acquisition of the syntax of negation in French and German: contrasting first and second language development. *Second Language Research* 13. 227–263.

- Paivio, Allan & Alain Desrochers. 1980. A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/ Revue Canadienne de Psychologie* 34 (4). 388–399.
- Potter, Mary C., Kwok-Fai So, Barbara Von Eckardt & Laurie B. Feldman. 1984. Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior* 23(1). 23–38.
- Poulisse, Nanda & Theo Bongaerts. 1994. First language use in second language production. *Applied Linguistics* 15 (1). 36–57.
- Poulisse, Nanda. 1997. Language production in bilinguals. In Annette M.B. de Groot & Judith F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, 201–224.
- Reetz, Henning. 2010. “voicekeyhilarly.pl”. [Perl Computer program].
- Roberts, Patricia M., Linda J. Garcia, Alain Desrochers, and Denise Hernandez. 2002. English performance of proficient bilingual adults on the Boston naming test. *Aphasiology* 16 (4–6). 635–645.
- Roelofs, Ardi. 1997. The WEAVER model of word-form encoding in speech production. *Cognition* 64 (3). 249–284.
- Roelofs, Ardi. 2003. Shared phonological encoding processes and representations of languages in bilingual speakers. *Language and Cognitive Processes* 18 (2). 175–204.
- Roelofs, Ardi & Kim Verhoeft. 2006. Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition* 9 (2). 167–176.
- Sholl, Alexandra, Aruna Sankaranarayanan & Judith F. Kroll. 1995. Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science* 6 (1). 45–49.
- Ueyama, Motoko. 2000. *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese*. Los Angeles, CA: University of California Los Angeles PhD thesis.
- Wheeldon, Linda & Aditi Lahiri. 1997. Prosodic units in speech production. *Journal of Memory and Language* 37 (3). 356–381.
- Wheeldon, Linda & Aditi Lahiri. 2002. The minimal unit of phonological encoding: prosodic or lexical word. *Cognition* 85 (2). B31–B41.
- Wynne, Hilary S. Z., Linda Wheeldon & Aditi Lahiri. (2018). Compounds, phrases and clitics in connected speech. *Journal of Memory and Language* 98. 45–58.
- Wynne, Hilary S. Z., Linda Wheeldon & Aditi Lahiri. (2019). What makes a compound a compound? Production of spaced compounds in connected speech. In review.

Appendix: Target stimuli

Experiment 1

compounds	adj-noun phrases	simple initial stress	simple final stress
daybreak	dark time	denim	default
daylight	dark break	debit	deceit
daytime	dark light	dagger	decree
groundwork	green nut	gravel	gazelle
groundnut	green hog	griddle	gazette
groundhog	green work	gravy	guitar
nightcap	nice mare	nibble	neglect
nightmare	nice gown	nickel	noblesse
nightgown	nice cap	nitrate	notate
lockjaw	long smith	locker	lament
locksmith	long nut	locust	latrine
locknut	long jaw	lodger	lapel
bathrobe	bad room	banter	baboon
bathroom	bad tub	ballad	balloon
bathtub	bad robe	basil	bamboo

Experiment 2

compounds	adj-noun phrases	bisyllabic	monosyllabic
daybreak	dark time	denim	dame
daylight	dark break	debit	date
daytime	dark light	dagger	daze
groundwork	green nut	gravel	grease
groundnut	green hog	griddle	greed
groundhog	green work	gravy	greet
nightcap	nice mare	nibble	nine
nightmare	nice gown	nickel	Nile
nightgown	nice cap	nitrate	ninth
lockjaw	long smith	locker	lob
locksmith	long nut	locust	log
locknut	long jaw	lodger	loft
bathrobe	bad room	banter	bag
bathroom	bad tub	ballad	band
bathtub	bad robe	basil	bat

Experiments 3 and 4

compounds	phrases	bisyllabic	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

