Variable prosodic phrasing in a theory of incremental production planning

Aron Hirsch*, Michael Wagner**

Abstract

Although speakers can modulate their prosodic phrasing to convey the intended interpretation of a structurally ambiguous sentence, experimental results conflict as to how reliably speakers actually *do* produce disambiguating prosodic cues. Discrepancies between studies have been attributed to differences in the studies' respective tasks. In this work, we demonstrate that some ambiguities are consistently disambiguated and others not *within a single experiment*. We argue that the findings can be explained by a particular model of how syntax and prosody interact with incremental production planning. Building on earlier work in the planning literature, we propose a model in which the size of production planning windows is flexible, each window comprises exactly one syntactic constituent, and windows correspond to intonational phrases. The observed variability in prosodic phrasing is attributed to variability in the size of planning windows, which interacts with syntactic constituency in just the right way to account for the results.

Keywords: prosody, syntax, production planning, ambiguities, attachment

1. Introduction

It has been known at least since Lehiste (1973) that speakers can modulate their intonation to indicate the intended interpretation of an ambiguous sentence. The focus in this paper is on structural ambiguities like (1), where a prepositional phrase (PP) can attach at different syntactic positions. In (1-a), the PP *with the flower* attaches as a modifier in the NP headed by *frog* and, in (1-b), *with the flower* is a modifier in the VP headed by *tap*.

*Principal corresponding author

**Corresponding author

Email addresses: aronh@mit.edu (Aron Hirsch), chael@mcgill.ca (Michael Wagner)

(1) Tap the frog **with the flower**.

a. Tap the frog that has a flower. *NP modifier reading*

b. Tap the frog by using the flower. *VP modifier reading*

(2) a. tap [$_{NP}$ [$_{NP}$ the frog [$_{PP}$ with the flower]]] *NP modifier structure*

b. [$_{VP}$ [$_{VP}$ tap the frog] [$_{PP}$ with the flower]] *VP modifier structure*

To convey the NP modifier reading, a speaker can render (1) with a relatively stronger boundary after *tap* than after *frog*. To convey the VP modifier reading, a speaker can adopt the converse prosody. Throughout the paper, intonational phrase boundaries are indicated with the pipe symbol ('||').

(3) a. Tap || the frog with the flower. *(favors NP modifier reading)*

b. Tap the frog || with the flower. *(favors VP modifier reading)*

Although speakers *can* use prosody to resolve PP-attachment ambiguities, the question remains: *do* speakers reliably produce disambiguating prosodic cues?

This question is important for our understanding of how syntax and prosody relate. According to one view, prosodic phrasing is read off the syntactic structure (Chomsky and Halle, 1968; Cooper and Paccia-Cooper, 1980; Grosjean and Collins, 1979) (though not necessarily in an isomorphic way: Gee and Grosjean, 1983; Selkirk, 1984; Truckenbrodt, 1995). If prosody is a "translation" of syntax, a given structure should automatically be associated with a particular prosody and disambiguation should be reliable.

Previous experimental work on PP-attachment ambiguities has turned up conflicting results about how reliably speakers produce disambiguating prosodic cues. In one set of studies, speakers disambiguate only in the absence of other contextual cues indicating the intended reading and, perhaps even then, only when they are consciously aware that the sentence is ambiguous (Allbritton et al., 1996; Snedeker and Trueswell, 2003). For instance, with the example in (1), Snedeker and Trueswell find that speakers opt for the disambiguating intonational patterns in (3) only under these circumstances; otherwise, speakers typically render (1) with no prosodic boundaries. Another set of studies, however, reports that speakers reliably produce disambiguating cues independent of context and awareness (Kraljic and Brennan, 2005; Schafer et al., 2000).

Taken together, previous results support weakening the relationship between syntax and prosody: although a given syntactic structure may map to a particular prosody, that mapping is not automatic. A possible conception is that speakers tailor their utterance to the needs of their interlocutor, so prosodically cue syntax only when required to facilitate processing by the listener ("audience design", cf. discussion in Kraljic and Brennan, 2005). When pressure to aid the listener is not in force, speakers default to a flat prosody. In this spirit, Kraljic and Brennan (2005) attribute the divergent results between studies to differences in the studies' respective tasks. Certain tasks, for instance, are more communicative than others and, if prosody is for the listener, these may favor disambiguation more strongly.

The contribution of the present study is to show that there is more driving variability in prosodic disambiguation than just task-based differences. While all of the studies cited above investigated PP-attachment ambiguities, their ambiguities involved quite different syntactic configurations. Our hypothesis is that divergent results are, at least in some cases, due to inherent syntactic differences between the ambiguities studied. Some PP-attachment ambiguities are consistently disambiguated and others not, irrespective of task-based factors.

To test this, we focus on two previous studies with conflicting results: Snedeker and Trueswell (2003, ST) and Kraljic and Brennan (2005, KB). We create a set of stimuli that combine aspects of their respective stimuli and, in this way, study their ambiguities together in a single experiment. Our critical result is that the conflicting results of both studies replicate within one study: an ambiguity similar to KB's is consistently disambiguated, while one similar to ST's is not. This rules out an explanation purely in terms of task demands.

Although our results show that syntax plays an important role in conditioning disambiguation, we argue that our full data set is best explained in a model where syntax does not itself map to prosody (either automatically or variably). Rather, we propose that the placement of prosodic boundaries is sensitive to incremental production planning, which is constrained by both syntax and performance-based factors. Our model combines insights from theoretical syntax and semantics with results from the literature on incremental production planning.

The paper is organized as follows. In §2, we review ST and KB and, following KB, discuss how a task-based explanation could account for their divergent results. In §3-4, we introduce our stimuli combining ST's and KB's ambiguities, and present our results replicating the conflict between ST and KB within a single experiment. In §5, we spell out our proposal and, in §6, extend our approach to account for further phenomena. §7 concludes.

2. How reliably does prosody differentiate attachment ambiguites?

In this section, we review the two studies upon which we base our experimental materials: ST and KB, which are representative studies finding that disambiguation is unreliable (ST) and reliable (KB), respectively.

2.1. Snedeker & Trueswell (2003)

ST studied sentences like (1), repeated below, in a series of production experiments involving pairs of participants.

(1) Tap the frog with the flower.

a. Tap the frog that has a flower. *NP modifier reading*

b. Tap the frog by using the flower. *VP modifier reading*

One participant ('the speaker') said aloud a sentence like (1) as an instruction to the other participant ('the listener'). The listener then used a set of objects to act out the instruction. In all experiments, the objects provided to the listener could be used to act out either interpretation of the test sentences. If listeners guessed the intended interpretation with above chance accuracy, the speaker's productions must have included disambiguating prosody.

We focus on two of ST's experiments. In Exp. 1, the speaker, like the listener, was provided with a set of objects which could be used to act out either interpretation of the test sentences. The experimenter demonstrated the intended action to the speaker before the speaker produced the instruction. Whether the intended reading was an an NP modifier reading or a VP modifier reading was varied within each speaker-listener dyad in a Latin Square design. In Exp. 2, the objects provided to the speaker were only compatible with one reading, and the target reading was varied between dyads, rather than within dyads. A given speaker, therefore, was conveying the same reading in all trials.

In Exp. 1, disambiguating prosody like that in (3) was observed. Speakers rendered the NP modifier reading as *Tap* || *the frog with the flower*, and the VP modifier reading as *Tap the frog* || *with the flower*. Listeners picked up on the disambiguating cues in the speakers' utterances and correctly acted out the intended action in approximately two thirds of trials. In Exp. 2, however, speakers did not prosodically differentiate the two readings, as both readings were rendered with no boundaries at all. Correspondingly, listeners performed at chance.

Thus, ST report that disambiguation is unreliable: while the two readings of (1) can be differentiated with the help of prosodic boundaries, speakers do not consistently employ those boundaries. ST identified two factors potentially at play in conditioning disambiguation. First, taking the objects provided to the speaker as the context for their utterance, it appears that disambiguating cues arise when the sentence is ambiguous in the context (Exp. 1), but not when the context disambiguates (Exp. 2). Second, disambiguating cues may arise only when the speaker is consciously aware that the sentence is ambiguous. ST assessed whether or not speakers were consciously aware of the ambiguity using a post-experiment questionnaire, which showed that speakers in Exp. 2 were often unaware.

2.2. Kraljic & Brennan (2005)

KB studied the sentence in (4), which can be interpreted as in (4-a) or (4-b). In the structure for (4-a), the PP *in the basket* attaches as a modifier in the NP headed by *dog* and *on the star* is the goal argument of the verb *put*. In the structure for (4-b), *on the star* attaches as a modifier in the NP headed by *basket* and *in the basket on the star* is the goal of *put*.

(4) Put the dog in the basket on the star.
   a. Put the dog that is in the basket onto the star.
   b. Put the dog into the basket that is on the star.

(5) a. [put [$_{NP}$ the dog in the basket] [$_{PP}$ on the star]] *(structure for (4-a))*
   b. [put [$_{NP}$ the dog] [$_{PP}$ in the basket on the star]] *(structure for (4-b))*

KB's task again involved pairs of interlocoturs. One participant directed the other to move objects in a physical display using a sentence like (4). KB, unlike ST, allowed the listener to ask for clarification or otherwise provide feedback to the speaker. The configuration of objects in the display was varied so as to manipulate whether both readings of (4) could be performed, or only one reading. Additional experiments probed for how aware speakers were of the ambiguity. In contrast to ST, KB report reliable disambiguation: speakers consistently produced prosodic cues to differentiate (4-a) and (4-b), independent of context and awareness. (4-a) was rendered as (6-a), while (4-b) rendered as (6-b).

(6) a. Put the dog in the basket || on the star. *(rendering of (4-a))*
   b. Put the dog || in the basket on the star. *(rendering of (4-b))*

2.3. What is behind the paradox?

The results of ST and KB seem paradoxical. Whereas ST found that speakers disambiguate only when the context is ambiguous and they are aware of the ambiguity, KB found that speakers disambiguate irrespective of these factors.

KB suggested several possible explanations for the conflicting data, all of which are based on differences in the studies' respective tasks and design. First, since KB allowed their participants to communicate with one another beyond just the speaker uttering the test sentence, the task in KB was more interactive than the task in ST. If speakers design their utterance for the benefit of the listener, they may be more likely to disambiguate in a more interactive task. Second, the between-dyad manipulation in ST's second experiment resulted in a given speaker producing the same structure in all trials. As a result, the ambiguity was less relevant in the task, which may have led speakers to adopt an uninformative prosody. Third: KB's stimuli were longer than ST's, making it more likely that a boundary is necessary to break up the utterance into several chunks to aid processing (see also Snedeker and Yuan, 2008).

Our aim is to assess a possibility that KB did not consider: that the discrepancy between ST and KB traces to an inherent syntactic difference between the ambiguities studied. To investigate this idea, we report an experiment designed to replicate the discrepancy between ST and KB in a single paradigm using a single stimulus set. If we find that certain ambiguities are disambiguated while others are not when the task-related factors considered by KB are held constant, we can conclude that syntactic differences play a role in conditioning the reliability of disambiguation.

3. Experiment 1: Replicating the puzzle with a single stimulus set

The findings of ST and KB were obtained with sets of stimuli involving two ambiguities that are configurationally quite different from one another. We create a stimulus set which involves both types of ambiguities, and test whether the conflict between ST and KB replicates with task-based factors controlled.

3.1. Combining the two ambiguities

To bring out the syntactic difference between the ambiguities in ST and KB, consider their test sentences together, focusing on the relationship between the bolded PP and the constituents that precede or follow it.

(7) Tap the frog **with the flower**. *(ST)*

a. tap [$_{NP}$ [$_{NP}$ the frog [$_{PP}$ with the flower]]]
b. [$_{VP}$ [$_{VP}$ tap the frog] [$_{PP}$ with the flower]]

(8) Put the dog **in the basket** on the star. *(KB)*

a. [put [$_{NP}$ the dog in the basket] [$_{PP}$ on the star]]
b. [put [$_{NP}$ the dog] [$_{PP}$ in the basket on the star]]

Descriptively, *with the flower* in (7) either forms a constituent with the NP to its left, or it attaches itself as a VP-modifier. *In the basket* in (8) can also form a constituent with the NP to its left, but the other option is for it to form a constituent with the PP to its right and have that larger constituent attach within the VP as the goal of *put*. It is possible that 'left branching' and VP-modifier structures, as in (7), are not consistently differentiated from each other, while 'left branching' and 'right branching' structures, as in (8), are.

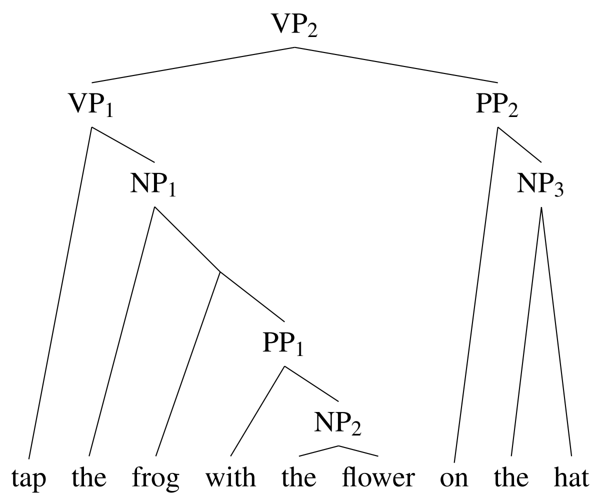
We consider the sentence in (9), which is built from ST's sentence, but with a second PP added, making it more similar to KB's. The important difference from both previous studies is that our sentence is three-way ambiguous, rather than just two-way ambiguous. The three interpretations are paraphrased in (9-a)-(9-c).[1]

(9) Tap the frog **[**$_{PP1}$ **with the flower]** [$_{PP2}$ on the hat].
  a. Tap the frog that has a flower and tap it on the hat.
  b. Tap the frog by using a flower and tap it on the hat.
  c. Tap the frog by using the flower that is lying on the hat.

The reading in (9-a) has the structure in (10), where PP_1 is a modifier in the NP headed by *frog*, and PP_2 is a modifier in the VP. PP_1 expresses which frog is to be tapped, and PP_2 expresses the goal of the tapping (i.e. the place where the frog is to be tapped). In this structure, *with the flower* forms a constituent with *the frog* to its left, so we will refer to it as the **'Left'** structure.

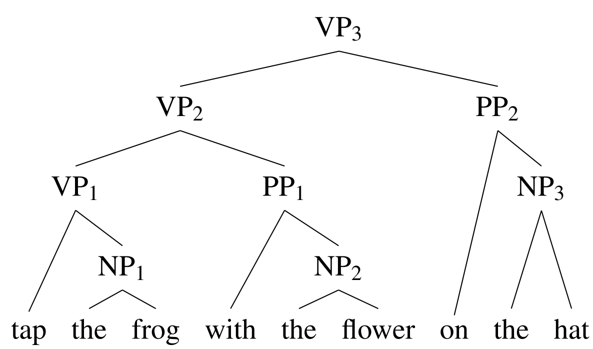
[1]There are in fact more readings available, which we did not investigate. In one reading, *both* PPs are interpreted as NP modifiers of *frog*, i.e. 'tap the frog that has a flower and is on a hat'. This reading is similar to the List reading in that both modifiers attach at the same level, but is different in that it only involves nominal modifiers.

(10) **Structure for Left reading**



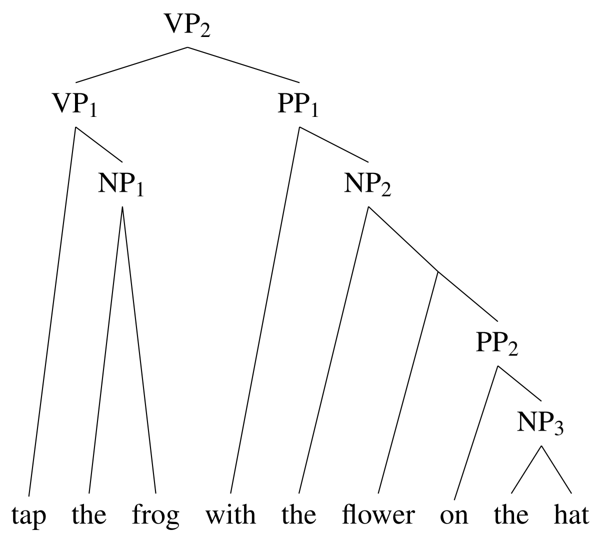
In (9-b), PP_1 and PP_2 are both VP-modifiers, as in (11). PP_1 expresses the instrument to be used to tap the frog, and PP_2 is again a VP-modifier expressing the goal of *tap*. In this structure, the PPs are each separate VP-modifiers, so we refer to this as the **'List'** structure.

(11) **Structure for List reading**



Finally, in (9-c), PP_2 is a modifier in the NP headed by *flower*, which is contained within PP_1. This larger PP_1 is itself a modifier in the VP. The structure is given in (12). PP_1 is an instrument, and PP_2 expresses which flower is to be used for the tapping. Since *with the flower* forms a constituent with *on the hat* to its right, this is the **'Right'** structure.

(12) **Structure for Right reading**



The comparison between the Left and List structures is parallel to ST's ambiguity. Consider the role of *with the flower*, bolded in (9). *With the flower* modifies *frog* in the Left structure, and attaches on its own as a VP-modifier in the List structure — just like in the two structures for ST's sentence in (7). The only difference from ST's sentence is that *on the hat* is also present in our stimulus. The role of *on the hat* is, however, controlled between the Left and List structures, as *on the hat* attaches on its own as a VP-modifier in both.

The comparison between Left and Right structures parallels KB's ambiguity. *With the flower* forms a constituent with *the frog* to its Left in the former structure, and forms a constituent with *on the hat* to its right in the latter structure. This is similar to *in the basket* in KB's sentence in (8) forming a constituent with *the dog* to its left in one structure, and *on the star* to its right in the other structure.

If speakers prosodically differentiate all three readings of our test sentence, we expect them to employ the phrasings in (13). The Left reading is marked by a relatively strong 'late' boundary before PP_2; the List reading is conveyed with a 'flat' prosody; and the Right reading is marked by a relatively strong 'early' boundary before PP_1.

(13) a. Tap the frog with the flower || on the hat. *(Left)*
b. Tap tap the frog (||) with the flower (||) on the hat. *(List)*
c. Tap the frog || with the flower on the hat. *(Right)*

3.2. Paradigm

We employed a production paradigm. Stimuli included 9 test sentences following the schema 'Tap the *x* with the *y* on the *z*', with different words filled in for *x*, *y*, and *z* in different items. We presented test sentences to participants in a written context biasing one of the three readings. Experimental conditions differentiated by which reading was contextually biased. The contexts from the Left, List, and Right conditions for the example sentence in (9) are given in (14).[2]

(14) a. **Context biasing <u>Left</u> reading**
John is in the forest. He sees a frog who is holding a flower. The frog is wearing a hat. You want John to reach over and use his finger to tap the frog's hat. This is what you say to him: (9).

b. **Context biasing <u>List</u> reading**
John is in the forest. He sees a frog who is wearing a hat. There is a flower nearby. You want John to take the flower and use it to tap the frog's hat. This is what you say to him: (9).

c. **Context biasing <u>Right</u> reading**
John is in the forest. He sees a frog. The frog's hat is on the ground. There is a flower on top of it. You want John to take the flower and use it to tap the frog. This is what you say to him: (9).

Participants were instructed to read the context and test sentence silently, and then to record themselves saying the sentence aloud, as if in a casual conversation. Recordings were made in a sound attenuated booth at McGill University.

3.3. Awareness manipulation and design

Apart from syntactic structure, we also manipulated how aware participants were of the ambiguity. There were two experimental groups, the high awareness group and the low awareness group. Two factors varied between groups. First, in the high awareness group, participants were told prior to the experiment that the test sentences would be ambiguous and were shown an example sentence, along with an explanation of the three ways that sentence could be interpreted. In the low awareness group, participants were not informed about the ambiguity.[3]

[2] All experimental materials are reproduced in Appendix A.
[3] The instructions for the two groups are reproduced in Appendix B.

Second, following ST, the intended reading of the test sentences was a within-subject variable in the high awareness group (Latin Square), and a between-subject variable in the low awareness group. That is, in the low awareness group, all contexts shown to a given participant were from the same condition. We refer to the high awareness group as the 'Within' group and the low awareness group as the 'Between' group in subsequent discussion.

We measured awareness in the Between group using a post-experiment questionnaire, adapted from ST's. The questionnaire included three questions, two of which are important here. The first question (Q1) asked participants their impression about what they thought the experiment was investigating. The second question (Q2) directly asked participants whether they had been aware of ambiguity in the test sentences. Participants were instructed to answer Q1 (on the front side of the page) before looking at Q2 (on the reverse side).[4] We classified participants on a three-level scale of awareness. Participants who showed clear signs of being aware of the ambiguity in their response to Q1 and responded positively to Q2 were coded as *aware*. Participants who showed no signs of awareness in their response to Q1 and responded negatively to Q2 were coded as *unaware*. All other participants were coded as *possibly aware*.[5]

We ran a total of 30 participants in the Within group, and a total of 60 participants in the Between group, since we expected a large effect in the Within group, and more subtle effects in the Between group, given the between-subject design. All utterances were screened for whether participants had produced the correct words without major disfluencies by an RA who was not aware which group or condition a particular utterance was from. We only included participants for whom at least 5 out of 9 utterances were usable, leaving 55 participants in the Within group and 26 in the Between group (although including all data did not lead to different results).

3.4. Results

One way to acoustically assess the prosodic phrasing of an utterance is to look at the degree of pre-boundary lengthening. We report the duration of the word immediately preceding a boundary as a cue to the strength of that boundary.

---

[4]The awareness questionnaire is reproduced in Appendix C.

[5]Data from the Possibly Aware group will not be discussed here for reasons of space. Results in this group were generally more similar to the aware group, but, as would be expected, the pattern was not as clear. We also note that, obviously, our questionnaire does not tell us at which point in time participants became aware if they did.

Referring to the test sentence in (9), we take the duration of *frog* as a cue to the boundary preceding PP_1, and the duration of *flower* as a cue to the boundary preceding PP_2. The degree of pre-boundary lengthening positively correlates with the strength of the boundary.

Figure 1 illustrates the duration of *frog* relative to *flower*. Data from participants in the Within group and the Between group are presented separately.

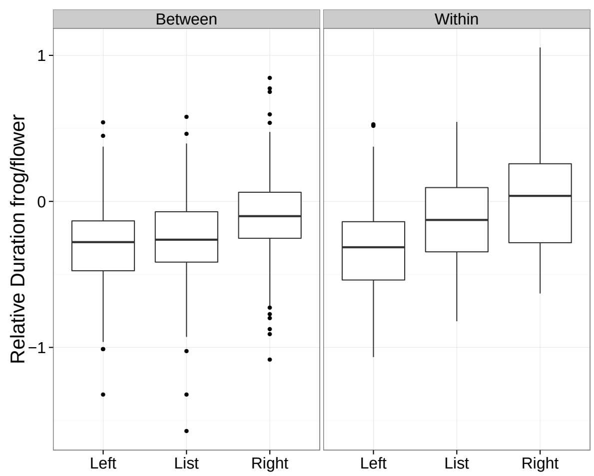


Figure 1: Log difference of duration of words of interest corresponding to *frog* and *flower*.

In the Within group, participants differentiate all three readings, and the pattern is consistent with the expected phrasings in (13), repeated:

(13) **Prosodic phrasing in the Within group**

a. Tap the frog with the flower || on the hat. *(Left)*
b. Tap tap the frog (||) with the flower (||) on the hat. *(List)*
c. Tap the frog || with the flower on the hat. *(Right)*

In Figure 1, with the left structure, *frog* has decreased duration relative to *flower*, as compared to the other conditions. This indicates that the early boundary is comparatively weak and the late boundary comparatively strong, as in (13-a). With the Right structure, *frog* shows its greatest duration relative to *flower*, indicating the converse prosody, as in (13-c). The relative duration with the List structure is intermediate, as expected from (13-b).

Critically, the result in the Between group is different. The distributions for the Left and List structures are largely overlapping, indicating that those two structures are not well differentiated from one another. Although the Right structure is somewhat less well differentiated from the other two readings than in the Within group, Right is still clearly differentiated. Recall that the ambiguity between Left and List parallels ST's ambiguity, and the ambiguity between Left and Right parallels KB's. The results for the Between group replicate the conflicting results of both studies: Left and List are not differentiated, while Left and Right are.

We fit a mixed effects linear regression model predicting relative duration from syntactic structure (Left, List, Right), group (Within, Between), and their interaction. The model included random slopes for structure and group in the items random effect, and a simple random effect for participants. We coded syntactic structure using treatment coding, so that the model compares both the List and Right structures to the Left structure, which served as the baseline. The output of the model is summarized in Table 1. [6]

	RelDuration
(Intercept)	-0.30 (0.07)***
DesignBetween.vs.Within	0.02 (0.05)
BracketingList	0.10 (0.03)***
BracketingRight	0.25 (0.04)***
DesignBetween.vs.Within:BracketingList	-0.15 (0.06)**
DesignBetween.vs.Within:BracketingRight	-0.12 (0.06)*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Mixed Effects Regression Model for Relative Duration (difference in log duration) in Experiment 1

The comparison between the Left and List structures significantly interacts with group, confirming that the differences apparent in the plots reflect a significant effect: Left and List are less prosodically distinct in the Between group than in the Within group. Note that the interaction for the difference between the Left and the Right structures is also significant. As noted, it appears from the plot that

[6]The reported model is the one with the maximal random effects structure we could fit such that the model converged (Barr et al., 2013). The full model was as follows: lmer(rdur1~Design*Bracketing+(1|participant)+(Design+Bracketing|item),data=subsRel). Due to the between-subject design, there are no slopes in participant random effects. We could also not include the interaction of Design and Bracketing in the by-item random effect, since the model would not converge if we did. Even when we used Helmert coding of of the factor Bracketing (which involves two orthogonal contrasts), including the interaction of Bracketing and Group lead to non-convergence. The model table reports estimates of the coefficients and standard errors. The p-values were estimated using the Wald test.

the Left and Right are better differentiated in the Within group than in the Between group — but, they do still appear to be differentiated in the Between group.

Focusing on the neutralization of the distinction between Left and List structures, our data set allows us to explore two further questions. First, is participants' awareness of the ambiguity the factor that conditions whether Left and List are disambiguated, or is the between-subject design in the Between group sufficient for neutralization? And, second, when participants fail to disambiguate Left and List, what phrasing do they employ: is neutralization to a list-like flat prosody, or a left-like late boundary phrasing?

To assess the effect of awareness, we consider just the Between group, and take into account the awareness questionnaire.[7] Despite not receiving explicit instruction about the ambiguity and only encountering one reading across trials, many participants in the Between group did still notice the ambiguity. Out of the 55 participants in this group, there were 22 that were clearly aware and 13 that were clearly unaware. We look for prosodic differences between these Aware and Unaware sub-groups. The relative duration of *frog* and *flower* for these sub-groups is provided in Figure 2.

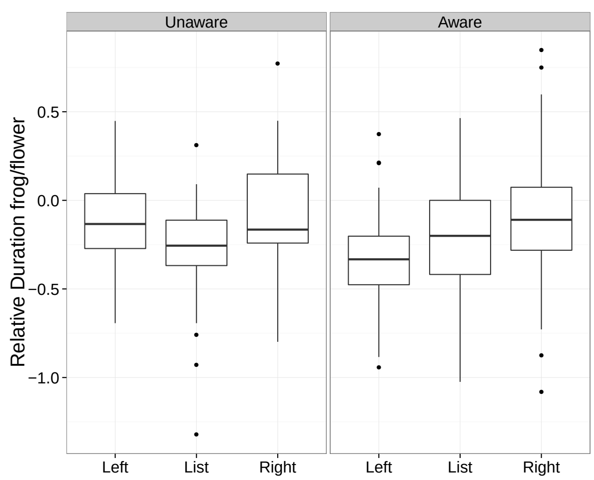


Figure 2: Log difference of duration of words of interest corresponding to *frog* and *flower* for the between-subject group, plotted by awareness according to post-experiment questionnaire.

[7]Recall that participants in the Within group were not administered the questionnaire since they received explicit instruction about the ambiguity, so were clearly aware.

Awareness does seem to play a key role in conditioning disambiguation, as the Aware and Unaware groups in Fig 2 show qualitatively different patterns. The Aware group shows a pattern very similar to the Within group in Figure 1, with all three readings differentiated. For the Unaware group, the distributions for the Left and List structures largely overlap, showing that these two structures are not well differentiated. In fact, the numerical difference in relative duration between Left and List goes in the opposite direction from the expected one.

As above, we fit a model predicting relative duration from syntactic structure (Left, List, Right) and awareness (Aware, Unaware) with random effects for items and participants. The output of the model, summarized in Table 2, confirms that the comparison between Left and List structures significantly interacts with awareness. There was no significant interaction with awareness in the comparison between Left and Right, although there is an empirical trend in the plot that the Aware group differentiated the two somewhat better.

	RelDuration
(Intercept)	-0.24 (0.07)**
AwarenessUnaware.vs.Aware	0.21 (0.08)**
BracketingList	-0.01 (0.05)
BracketingRight	0.14 (0.06)*
AwarenessUnaware.vs.Aware:BracketingList	-0.27 (0.11)*
AwarenessUnaware.vs.Aware:BracketingRight	-0.19 (0.11)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Mixed Effects Regression Model by Awareness for Between-Group in Experiment 1

We now consider how the Left and List structures are realized when they are not disambiguated. To do this, we plot the duration of *frog* and *flower* individually, as in Figure 3, which also includes the duration of *tap* and *hat*. As in the original plot, data from all Between participants are aggregated and presented separately from Within participants.

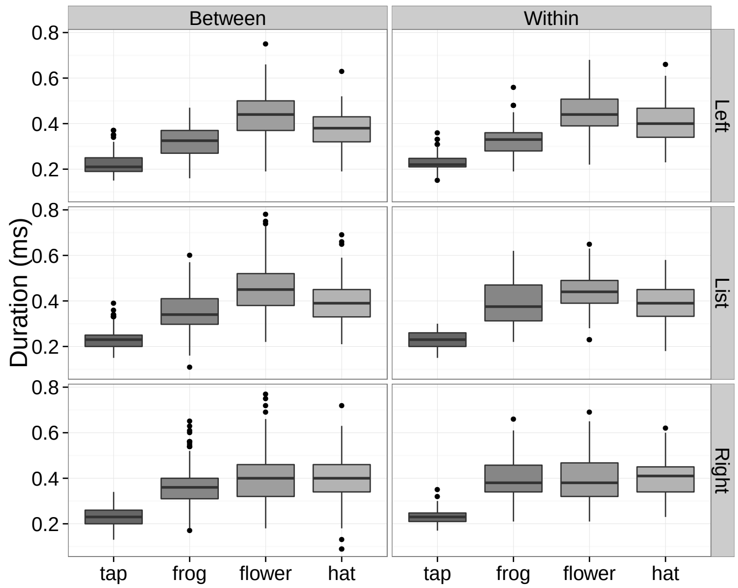


Figure 3: Duration of words of interest

In the Within group, the Left reading clearly is rendered with a strong late boundary, as *flower* is of greater duration than *frog*, while the List reading is rendered with a flatter prosody, as *flower* and *frog* are closer in duration than in the Left condition. This is the same result observed discerned earlier from the relative duration plot in Figure 1. Critically, we observe that, in the Between group, the Left and List structures are both rendered with a strong late boundary, like the Left structure in the Within group. Neutralization is not to a flat prosody, but to an articulated prosody otherwise characteristic of the Left reading.

Note that the duration results for the Right structure in Figure 3 are difficult to interpret. In the Within group, it appears that *frog* and *flower* are quite close in duration, which also seems to be the case in the Between group. In the next section, we present further results which clarify the pattern for the Right structure.

3.5. Section summary

The relative duration plots make clear that when participants are aware of the ambiguity, they prosodically differentiate all three readings. The Left structure is rendered with reduced duration of *frog* and increased duration of *flower*; with the the List structure, the duration of the two words is intermediate; and with the Right structure, the duration of *frog* is increased and that of *flower* decreased compared to the other two structures. The Between group replicates the puzzle raised by ST and KB. The ambiguity parallel to ST's (Left vs. List) is not disambiguated, while the ambiguity parallel to KB's (Left vs. Right) is disambiguated (even if these structures appear somewhat less well differentiated in the Between group).

The results for the Between group (in particular, the unaware speakers in the Between group) have uncovered a new dimensions to the data not apparent in the earlier studies. Compare again our stimuli to ST's:

(15) a. Tap the frog with the flower. *ST*
b. Tap the frog with the flower on the hat. *Exp. 1*

ST found that the two readings of (15-a) were not distinguished by unaware speakers because these speakers did not employ any boundaries at all. Our data reveal that failure to disambiguate is not always due to a flat prosody. Unaware speakers in our Left and List conditions did not render (15-b) with any boundary internal to *tap the frog with the flower*, consistent with ST's data — but, they did employ an articulated prosody, as they produced a late boundary preceding *on the hat*. The List structure, like the Left structure, tended to be rendered: *Tap the frog with the flower || on the hat*.

Because the conflicting results of ST and KB obtain even when tasked-based factors are controlled, our syntactic hypothesis receives support. The fact that the distinction between Left and List neutralizes to an articulated late boundary phrasing offers a window into *how* syntax conditions the reliability of prosodic disambiguation. Before discussing this, however, we report additional results which corroborate the conclusions so far, and add a further new generalization having to do with the realization of the Right structure.

4. Experiment 2: Establishing the prosodic phrasing

While the duration results are suggestive, they are not fully conclusive. First, there are acoustic cues for prosodic phrasing other than duration and, once those are taken into account, it could be that all three readings are differentiated after all, even by unaware speakers. Second, as noted, it is difficult to discern the precise phrasing for the Right structure from the absolute duration plot in Figure 3. Finally, duration data do not tell us about the likelihood of different phrasings, or whether the flat phrasings we observed involved two boundaries of equal strength or no boundaries. To further explore the data, we conduct a perception experiment, where production data from Exp. 1 are annotated for prosodic phrasing by naïve participants. The results from naïve participants are further extended with a perceptual annotation by expert annotators.

4.1. Naïve listeners

A subset of sound files from the production experiment were presented to naïve participants, who selected which of the three prosodic phrasings in (16) gave the most adequate description of the pronunciation of the utterance. Participants received instructions as to how to interpret the parentheses in (16), and found the task simple and intuitive.

(16)	a.	(a b) c	*Late boundary*
	b.	(a b c)	*Flat phrasing*
	c.	a (b c)	*Early boundary*

Separate perception experiments were run with sound files stemming from the Within and Between groups. From the Between group, we selected 243 sound files, including 81 from each of the Left, List, and Right conditions. We ran 36 naïve participants in the perception experiment. Each participant evaluated 81 sound files, resulting in each of the 243 sound files being annotated 12 times. The reported results exclude one participant who self-reported as not being a native speaker of English. Since the results in the perceptual experiment for the productions of the Between group were very similar for each subset of 81 sound files, we decided to run a smaller perception experiment for productions from the Within group. We selected a subset of 90 files (30 from each condition), and presented them to 16 participants.

Figure 4 plots the results of both perception experiments. It shows the proportion (and absolute number) of trials in which participants perceived the utterance in the sound file to be phrased with an early boundary, *(a b) c*, a flat phrasing, *(a b c)*, and a late boundary, *a (b c)*. Data are plotted according to the group and original condition in which the sound file was recorded. The Aware and Unaware sub-groups of the Between group are presented separately.

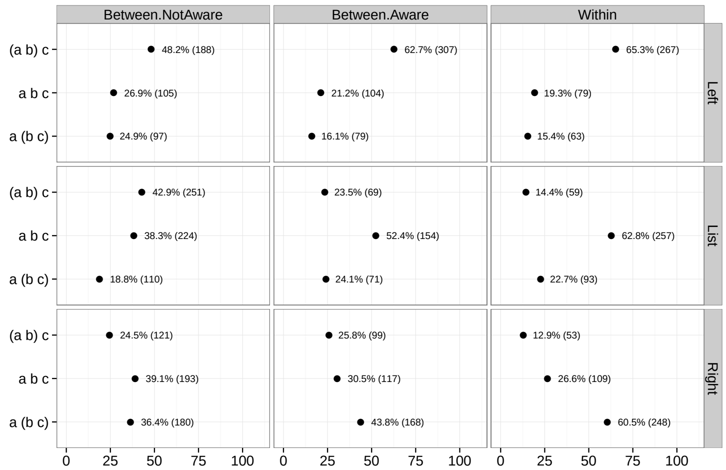


Figure 4: Perception of late boundary by naive participants

Perceptual results confirm that participants in the Within group are relatively reliable in conveying a disambiguating prosodic phrasing: in approximately two thirds of data, a late boundary is perceived with sound files from the Left condition, a flat prosody with sound files from the List condition, and an early boundary with sound files from the Right condition. The pattern is qualitatively similar for aware participants in the Between group.

When comparing this pattern to the results from the unaware participants in the Between group, a main interest given the acoustic results regards the realization of the Left and List structures — in particular, the rate at which these structures are realized with a late boundary. We observe that the List condition shows almost as high a rate of late boundary phrasings as the Left condition in the Unaware group (42.9% vs. 48.2%). This contrasts with the Within group and the Aware group, where the Left and List conditions sharply differed (23.5% vs. 62.7%; 14.4% vs. 65.3%). Thus, the List and Left structures were not very well differentiated by unaware speakers, consistent with out conclusion from the acoustic data.

We evaluated whether the difference in likelihood of a late boundary phrasing in the Left vs. List conditions interacted with awareness using a logistic regression, with the choice of the late boundary phrasing as the dependent variable. The model included random effects for items and participants, with random slopes for structure and awareness for the items random effect. The model output is summarized in the left-hand column in Table 3 (the column labeled 'Naïve').

	Naive	Expert
(Intercept)	-0.71 (0.12)***	-1.14 (0.23)***
BracketingRight.vs.Other	-1.06 (0.13)***	-2.16 (0.46)***
BracketingLeft.vs.List	1.52 (0.16)***	1.75 (0.43)***
GroupBetween.vs.Within	0.42 (0.16)**	0.17 (0.39)
GroupUnaware.vs.Aware	0.08 (0.11)	-0.47 (0.47)
BracketingRight.vs.Other:GroupBetween.vs.Within	0.48 (0.20)*	-0.54 (0.68)
BracketingLeft.vs.List:GroupBetween.vs.Within	-1.56 (0.22)***	-1.45 (0.73)*
BracketingRight.vs.Other:GroupUnaware.vs.Aware	-0.24 (0.20)	-0.30 (1.00)
BracketingLeft.vs.List:GroupUnaware.vs.Aware	-1.55 (0.22)***	-2.97 (1.12)**

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Logistic Mixed Effects Regression Models for Perception of the Rate of Late Boundary Phrasing, Based on the Naive and Expert Perceptual Annotations

We observe two critical interactions: the comparison between Left and List interacts (i) with the comparison between Within and Between groups, and (ii) with the comparison between Aware and Unaware sub-groups of the Between group. A late boundary rendering is statistically more likely in the Between group than the Within group, and more likely with unaware participants than aware participants in the Between group.

The comparison between the Right structure and the other structures also interacted with Within vs. Between group. This is expected given the plot in Figure 4: the Right structure is much more likely to be realized with an early boundary in the Within group. Importantly, though, the perceptual categorization of the Right structure shows a very different signature from the Left and List structures even in the Between group: the predominant phrasing used in these latter structures, the late boundary phrasing, is not dominant with the Right structure.

While the perceptual data show that the Right reading is conveyed with an early boundary by aware speakers, the pattern for unaware speakers is still unclear, as participants are close to chance in which phrasing they perceive for sound files from the Right condition in the unaware group. This could suggest that the Right reading was less clearly articulated prosodically. Alternatively, this could indicate that speakers did employ all three phrasings with similar frequency. To better understand the pattern, we also had trained RAs annotate the data.

4.2. Expert Annotation

Two trained annotators listened to all sound files from the production experiment and, for each, coded where prosodic boundaries were produced. Their coding included a category not involved in the naïve annotation: rather than simply coding for an early boundary, a flat prosody, or a late boundary, the category of flat prosody was subdivided into the categories of two boundaries and no boundaries.

(17) **Coding schema for expert annotation**

a. Tap the frog || with the flower || on the hat. Two boundaries
b. Tap the frog || with the flower on the hat. Early boundary
c. Tap the frog with the flower || on the hat. Late boundary
d. Tap the frog with the flower on the hat. No boundaries

The two annotations were qualitatively very similar. We statistically assessed inter-annotator agreement by computing Cohen's Kappa for whether or not the two annotators perceived a late boundary, and found that agreement was substantial (Cohen's K > 0.7, cf. Landis and Koch 1977). We report results from one of the two annotations. To select which annotator to report, we fit a linear regression model with the response variable the duration of the pre-boundary word (*frog*, *flower*) summed with the duration of the subsequent pause; annotator (annotator 1 vs. annotator 2) was included as a predictor. We report the annotation that explained a greater amount of the variability in the acoustic results ($R^2 = 0.3$ for the reported annotation vs. $R^2 = 0.21$ for the second annotation). The data according to that annotation are summarized in Figure 5.

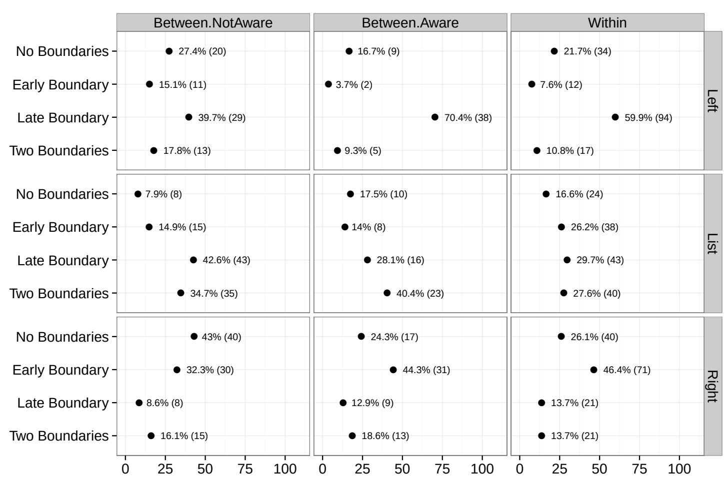


Figure 5: Perception of late boundary expert annotators

The results confirm our previous interpretation of the data, and clarify the pattern for the Right structure. We address each structure in turn.

For the Left structure, the late boundary rendering is most common in all groups, convergent with all data we have seen.

For the List structure, a flat prosody — either two boundaries or no boundaries — is most common in the Within group (two boundaries and no boundaries together, 58%) and the Aware sub-group of the Between group (44%).[8] In the Unaware group, our critical result is confirmed: the List reading is commonly realize with a late boundary, like the Left reading. As summarized in Table 3 (in the column 'Expert'; see preceding subsection), we evaluate significance in the same way as with the naïve annotations, and find that the difference in rate of late boundary renderings between the Left vs. List conditions interacts with Between vs. Within group and with awareness in the Between group.

For the Right structure, the early boundary phrasing is most common in the Within group and the Aware sub-group of the Between group. The pattern in the Unaware group clarifies the earlier results: although early boundary phrasing is still quite common, the predominant phrasing is flat, most commonly involving no boundaries. In light of this finding, the variability observed in the naïve annotation was likely due to speakers realizing an unarticulated prosody, which naïve participants found confusable with other phrasings. Importantly, the phrasings employed with the Right structure (early boundary for aware participants, and early boundary or no boundaries for unaware participants) are not dominant with either the Left structure or the List structure.

By coding separately for two boundaries and no boundaries, the expert annotation allows us to detect a new result about the nature of flat phrasings: the flat phrasing for the List structure involved two boundaries more frequently than no boundaries, with no boundary renderings particularly rare in the Unaware group. This contrasts with the Right structure, for which, as noted, flat phrasing more frequently involved no boundaries, especially in the Unaware group.

4.3. Section summary

The perceptual data corroborated our conclusions based on the acoustic data, and clarified the pattern for the Right structure. Taking all the data together, our

[8]Early and late boundary renderings are also observed for the List reading. We defer discussion of this issue until Section 5, after we have presented our conception of how syntax and prosody interact.

results show that aware speakers tend to render the three readings as in (18), and unaware speakers tend render the three readings as in (19).

(18) **Phrasing by aware speakers**
  a. Tap the frog with the flower || on the hat. *(Left)*
  b. Tap tap the frog || with the flower (||) on the hat. *(List)*
  c. Tap the frog || with the flower on the hat. *(Right)*

(19) **Phrasing by unaware speakers**
  a. Tap the frog with the flower || on the hat. *(Left)*
  b. Tap tap the frog with the flower || on the hat. *(List)*
  c. Tap the frog with the flower on the hat. *(Right)*

In addition to replicating the conflict between ST and KB within a single experiment, we have established two new generalizations, stated in (20). Note that the frequent occurrence of no boundaries with the Right structure, flagged in (20-ii), is noteworthy, as the dominant phrasings for the Left and List structures always included *some* boundary.

(20) (i) Unaware speakers preferentially realize List structures with a late boundary, resulting in a neutralization with the Left structure.

  (ii) Unaware speakers preferentially realize Right structures with a flat prosody involving no boundaries (compared to Left and List structures, which are rarely produced without any boundary).

We must now address the question: *why* are certain syntactic ambiguities consistently disambiguated and others not? Our two generalizations provide a window into the relationship between syntax and prosody.

5. Syntax, prosody, and incremental production planning

We began the paper by entertaining two views of how syntax and prosody relate. In the first view, syntax automatically maps to prosodic phrasing. In the second view, syntax can map to prosodic phrasing, but does so variably. KB suggested that prosodic cues surface for audience design in particular tasks or when stimuli are above a certain length; otherwise, speakers default to a flat prosody, as in ST. Our data belie both views.

Convergent with ST, our results rule out the possibility that prosody is an automatic translation of syntax by showing that more prosodic variability is observed than an automatic mapping can explain. Under our syntactic assumptions, an automatic mapping would predict that the Left, List, and Right readings should always be prosodified as in (21), contrary to fact.

(21) a. Tap the frog with the flower || on the hat. *(Left)*
b. Tap the frog || with the flower || on the hat. *(List)*
c. Tap the frog || with the flower on the hat. *(Right)*

Our results also contradict the variable view in several ways. First, if the occurrence of disambiguating prosody were fully a function of awareness of the ambiguity or its relevance for the task at hand, or other task-dependent factors, all three readings should be disambiguated, or none of them should be, depending on these factors. In fact, Left and List structures are not disambiguated, while Left and Right structures are disambiguated, with these factors controlled. Second, when disambiguation did not occur, we did not observe a flat default prosody: unaware speakers rendered each of Left and List with a late boundary. At the same time, unaware speakers did render the Right reading with no boundaries, despite the stimuli being as long as KB's.

To achieve a descriptively adequate account of our data, syntax must not automatically map to prosody — but it must constrain prosodic variability so that the Left and List structures pattern differently from the Right structure. In the following, we propose an incremental model of speech production in which syntax places limits of possible prosodic phrasings, and task-dependent factors including awareness also exert an influence on the prosodic outcome.

5.1. Why incremental production planning is relevant

To show why we think production planning is relevant, let us compare again the dominant prosodic realizations by aware speakers, repeated in (22), and unaware speakers, repeated in (23).

(22) **Prosodic phrasing by aware speakers**
a. Tap the frog with the flower || on the hat. *(Left)*
b. Tap the frog || with the flower || on the hat. *(List)*
c. Tap the frog || with the flower on the hat. *(Right)*

(23) **Prosodic phrasing by unaware speakers**
a. Tap the frog with the flower || on the hat. *(Left)*
b. Tap the frog with the flower || on the hat. *(List)*

c. Tap the frog with the flower on the hat. *(Right)*

Our first generalization — that the List structure is realized with a late boundary in (23-b) — and our second generalization — that the Right structure is realized with no boundaries in (23-c) — both involve an early boundary present in the Aware group being omitted in the Unaware group. (23-b) differs from (22-b), and (23-c) differs from (22-c), in the omission of an early boundary. Late boundaries, on the other hand, are not omitted. The late boundary in (22-a) and the late boundary in (22-b) are retained in (23-a) and (23-b).

In this way, our two generalizations converge on a left-to-right (or 'early' vs. 'late') asymmetry: early boundaries, but not late boundaries, may be omitted. We relate this left-to-right asymmetry in prosodic phrasing to the incremental nature of production planning.

5.2. Prosody and incremental production planning

The literature on production planning has shown that speakers often do not plan their entire utterance in one go. Rather, they chunk the utterance into smaller planning units (what we call "planning windows"), and plan incrementally: they plan one unit and then continue on to plan the next unit, successively from left-to-right, until the entire utterance is planned. For ideas and empirical results supportive of incremental planning, refer to Levelt (1989), Ferreira (1993), Ferreira and Dell (2000), Roelofs (1998), Sternberg et al. (1978, 1980), Watson and Gibson (2004b), among others.

To link prosody to incremental production planning, we propose that the edges of planning windows are marked by prosodic boundaries, per the following:

(24) **Prosodic Planning Window Hypothesis (PPW)**
Intonational phrases correspond to planning windows.

A key consequence of the PPW has to do with the locus of prosodic variability: if prosodic phrase boundaries mark the edges of planning windows, prosodic variability must reflect variability in how the utterance is planned. We take it that the size of planning windows is flexible (cf. Cooper and Paccia-Cooper, 1980, 5). Planning windows may be bigger or smaller, depending on various factors, including audience design and other task-based considerations of the sort KB discussed, as well as speech rate and working memory resources.

In a series of papers on production and motor planning, Sternberg et al. (1978, 1980) offer insights into the size of planning windows. These studies show that the amount of time a speaker takes to initiate their utterance correlates (i) with the number of syllables in the first word of the utterance, while the number of syllables in subsequent words has no effect; and (ii) with the number of upcoming words. The first result suggests that fine phonological and phonetic detail is planned over very small domain, approximately comprising one prosodic word (Levelt, 1989). The second result suggests that some aspects of planning occur over a larger unit, which comprises multiple lexical items.

Our interest is in the higher-level planning unit, and our proposal is that the exact amount of material planned in a given window at this level exhibits variability. This idea finds precedents in, for example Watson and Gibson (2004b) and Wheeldon (2012).

5.2.1. Syntax, Prosody and production planning

We have proposed that the size of planning windows is flexible, and conditioned by audience design and performance-based factors. The role for syntax, we suggest, is to constrain the space of possibilities for how an utterance can be partitioned into planning windows.

While planning proceeds incrementally left-to-right, this does not mean that hierarchical structure plays no role. Lashley (1951) provided an early suggestion that production planning must be organized according to the hierarchical structure of the sentence (see also Lee et al., 2013). Current production planning models converge on positing at least two planning stages for higher level structure: the functional level, at which the intended message is mapped to a basic argument structure, and the positional level, where a particular syntactic structure that conveys the intended message is committed to and planned (Bock and Levelt, 1994). In this model, often called the 'consensus model', Sternberg's higher-level planning unit would likely reflect assembly at the positional level.

Although the size of planning windows is flexible, we propose that the make-up of planning windows is constrained such that the material planned in a given window must constitute a syntactic constituent.

(25) **Syntactic Planning Window Hypothesis (SPW)**
Planning windows correspond to syntactic constituents.

In the remainder of this subsection, we show that the SPW need not be a primitive stipulation at all, but rather is a consequence of our model, once the processing advantage of incremental planning is made explicit and considered in light of theoretical work in syntax and semantics.

To set up discussion, consider an arithmetic formula:

(26) $(((3+4) * 5) - 8) *3)$

This formula is straightforward to evaluate incrementally. *3+4* is evaluated first, and the interim result of *7* is computed. Then, only that result is retained when **5* is considered, and another interim result of *35* is computed. Evaluation proceeds in this way to compute a final result of *81*. Since the intermediate result for each operation is computed at each step, there is minimal memory load: one can just keep in mind the intermediate result and forget the material that led to it. If (26) were not evaluated incrementally, the entire sequence of operations would have to be held in memory at once, making the result difficult or impossible to compute.

Incremental planning of an utterance should have a similar advantage. In each planning window, a meaning is computed and carried over to the next planning window, with the steps taken to compute that meaning forgotten, minimizing memory demands. With this result in hand, we draw upon current approaches to semantic composition in the theoretical literature to derive the SPW.

Underlying all current semantic theories is the Principle of Compositionality: the idea that the meaning of a sentence is a function of the meanings of the words that make up that sentence and how those words are put together syntactically. The semantic composition thus "piggy-backs" on syntactic structure. (27-a) illustrates:

(27) John tapped Bill.

In the structure for (27-a) in (28), *tapped Bill* is a VP constituent and *John* then forms a TP constituent with that VP. Semantically, *tap* is modeled as a predicate which requires two individual arguments to be saturated. In the notation of Heim & Kratzer (1998), this two-place predicate can be written as (29-a), where the λ terms introduce individual arguments. The NPs *John* and *Bill* denote individuals, the individual John (29-b), and the individual Bill (29-c), respectively.

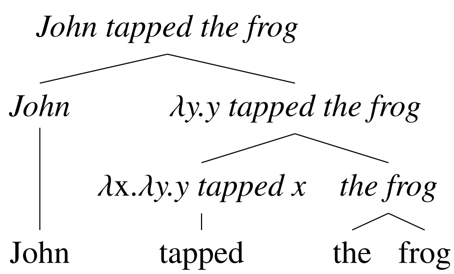
(28) [$_{TP}$ John [$_{VP}$ tapped Bill]]

(29) a. $[\![\text{tap}]\!] = \lambda x . \lambda y$. y tapped x
b. $[\![\text{John}]\!]$ = John
c. $[\![\text{the frog}]\!]$ = the frog

Under this analysis of an SVO sentence, semantic composition proceeds by applying the function in (29-a) to its arguments, following the syntax in (28). *Tapped* composes first with the object *Bill*, which saturates one argument slot in *tapped*.

The denotation of the VP is thus a one-place predicate. *John* composes with that one-place predicate to yield the final truth-conditions for the sentence as the denotation of the TP. The composition is sketched in (30).

(30)



The critical observation is that every syntactic constituent — the terminals *tap*, *John*, and *Bill*; the VP; and the TP — has its own denotation, and non-constituents do not have any denotation. *John tapped*, for instance, is not interpretable, given the composition in (30). Since the goal of a planning window, in our view, is to compute an intermediate meaning to ease memory demands and only constituents have meanings, it follows that the material planned in a given window must correspond to a syntactic constituent. Hence, the SPW is a consequence of our model, and does not have to be stipulated as a primitive assumption.[9]

With this result in hand, consider the sentence in (31), which is parallel to (27), except the subject is quite complex, containing a relative clause.

(31) [$_{TP}$ The man who John saw in New York [$_{VP}$ tapped Bill]]

As before, *tapped* and *Bill* compose to yield a denotation for the VP, and that meaning composes with the subject to yield the final meaning for the sentence. Here, however, determining the meaning of the subject itself involves quite a few steps of semantic composition. Example (31) is much like the arithmetic formula

[9] Based on our assumptions about planning, it should be possible to partially compute the subject, or to compute the entire structure in one go, but it should not be possible to computer *John tapped* in one planning window, since it doesn't form a syntactic unit. Importantly, the apparent non-constituent *John tapped* is actually a possible planning window after all, at least under certain circumstances (Steedman, 2000)—contrary to what we would expect based on what we said so far. We will return to this observation in Section 6, showing how such alternative phrasings actually reveal alternative syntactic constituent structures, with corresponding alternative composition paths.

in (26). If the entire utterance were planned in one go, all of the operations would have to be held in memory at once. Memory demands can be alleviated if (31) is planned in two windows. The first window could comprise the subject, with the second then comprising the VP, as in (32).

(32) [The man John saw in New York]$_{-1}$ [tapped Bill]$_{-2}$

In the first planning window, the internal structure of the subject is built and the steps of semantic computation required to determine the denotation of the subject are executed. Once this is done, the internal structure of the subject and the steps required to interpret it can be forgotten: only the highest NP node and its final denotation are retained. In the second planning window, the remaining structure is built and interpreted. Semantically, the denotation of the VP is computed and composed with the retained denotation of the subject NP. Compatible with the idea that speakers will decide to split up a structure into multiple planning windows with greater internal complexity of the individual constituents is the observation that constituent size correlates with the likelihood and strength of adjacent boundaries (Krivokapić, 2007; Watson and Gibson, 2004b).

Ferreira (2000) has proposed a model of production planning based on tree adjoining grammar which bears some similarities to our own, though our model is different in important ways. In this model, the speaker plans an "elementary tree" (essentially, the main predicate of the sentence with its various argument slots, similar to our semantic representation of the verb in (29-c)) and one dependent prior to utterance initiation, with the remaining argument slots filled in at later steps. Our model differs from Ferreira's in that we do not assume that the main predicate (or the 'elementary tree' containing it in the terminology of tree adjoining grammar) is necessarily planned before utterance initiation. The amount of syntactic material planned in the first window varies, and the first window may well be too small to include the main predicate. Schriefers et al. (1998) provide supportive evidence that the main predicate is not always planned in the initial window (cf. Lee et al., 2013, for discussion). Ferreira and Dell's model also differs from our own in that a planning windows does not necessarily constitute a syntactic constituent.

5.2.2. Summary of model

The preceding subsections have introduced the two critical tenets of our approach: the link between production planning and prosody captured in the PPW, and the link between production planning and syntactic constituency captured in the SPW. We see the size of intonational phrases as being due to a speaker's choice of how much material to plan within a single planning window, and hold that these choices are constrained by syntactic constituency structure.

(33) **Prosodic Planning Window Hypothesis (PPW)**
Intonational phrases correspond to planning windows.

(34) **Syntactic Planning Window Hypothesis (SPW)**
Planning windows correspond to syntactic constituents.

Taking the PPW and SPW together, our proposal is an advancement of a number of previous ideas in the literature. Because planning in a given window involves computing the meaning of the syntactic constituent making up that window, one antecedent is Selkirk's (1984, 291) *Sense Unit Condition*, which is based on the idea that prosodic phrase boundaries delimit domains for semantic interpretation. An even closer precedent is Schafer's (1997) *Interpretive Domain Hypothesis*, which closely relates to the Sense Unit Condition, and holds that "an intonational phrase boundary defines a point at which the processor performs any as yet outstanding semantic/pragmatic evaluation and integration of material within the intonational phrase."

A third antecedent proposal is the Left hand side/Right hand side Boundary hypothesis ('LRB') proposed in (Watson and Gibson, 2004b). Watson and Gibson (2004b) demonstrate that the likelihood of a prosodic boundary being placed at a particular position varies with the size of the syntactic constituent to the left of that position and with the size of the syntactic constituent to the right of that position. They suggest that "boundary placement is the result of planning processes in production" (p. 750), with boundaries inserted in order to give the speaker time to recover from or prepare for the computation of adjacent constituents, an idea very compatible with the view taken here. We will compare our approach with the LRB in more detail later at the very end of this section.

Finally, we note that, if the PPW and SPW are assumed, the placement of boundaries will always be compatible with the *Informative Boundary Hypothesis* proposed in Clifton et al. (2002). This hypothesis has the effect that placing optional prosodic boundaries which directly contradict syntactic structure are dispreferred or prohibited.

As discussed with Ferreira's model, one respect in which our approach distinguishes itself from most of its progenitors is in the claim that planning windows and, thus, intonational phrases must constitute syntactic constituents. Non-constituent intonational phrases are allowed with both the Sense Unit Condition and the LRB. There is, in fact, some apparent counter-evidence to our proposal. Data in Lindsley (1975) suggest that the subject and the verb in SVO sentences are planned before utterance initiation, but not the object. With the standard constituency (S)(VO), (SV) does not seem to be a licit initial planning window. An SVO sentences can also be prosodically rendered with a single boundary between the verb and object in some cases, running into the same problem. We return to this issue in Section 6, where we argue that certain very limited apparent non-constituent planning windows are straightforwardly reconciled with the SPW.

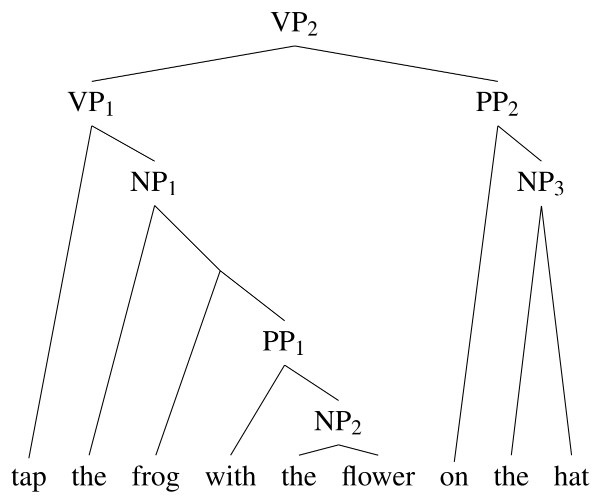
The ingredients are now in place to understand the relationship between prosody and syntax as manifest in our experimental results. Given the SPW, the way our test sentence may be partitioned into planning windows depends on its syntactic structure. We apply our model to each of the Left, List, and Right structures in turn and show how it can explain the observed prosodic phrasing with both aware and unaware speakers.[10]

5.2.3. Deriving the prosody of the Left structure

We begin our discussion with the Left structure, repeated in (35).

[10]Note that we are not arguing that chunking in production planning is the *only* source of prosodic phrasing. We believe that the syntactic difference between two structures can have prosodic reflexes even they are articulated in a single planning window. For example, Price et al. (1991) found that the ambiguity observed in sentences like *The student dropped off the table* between a phrasal verb taking a direct object as its argument and a verb taking a prepositional phrase as its argument is reliably encoded on prosody, and Norcliffe and Jaeger (2005) show that this differences is preserved when intonational phrasing is ruled out as a source of phrasing by placing the structure in the deaccented domain after the sentence focus. We will not discuss how such prosodic differences in the absence of intonational phrasing come about.

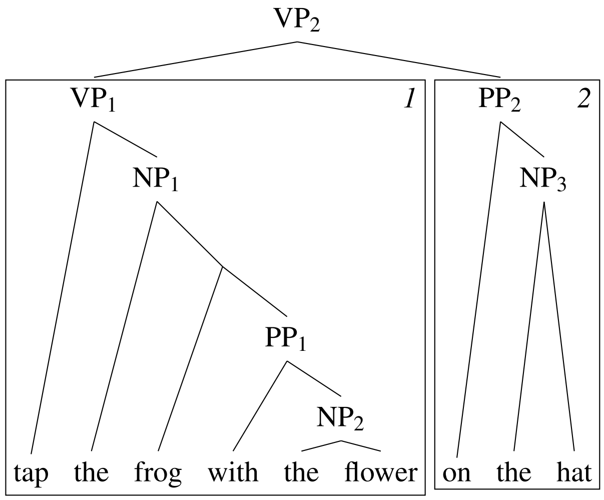
(35) **Left structure**



On the assumption that planning windows are syntactic constituents, there are at least two possibilities for how the speaker could partition the utterance into windows. The first is to simply plan the entire utterance in a single window: the one and only planning window could be VP_2. By the PPW, if the utterance is planned in one window, the entire utterance should be a single prosodic phrase.

The second possible partition involves a more articulated planning structure, where the utterance is divided into multiple planning windows. The first constituent which could be built in left-to-right planning is VP_1:

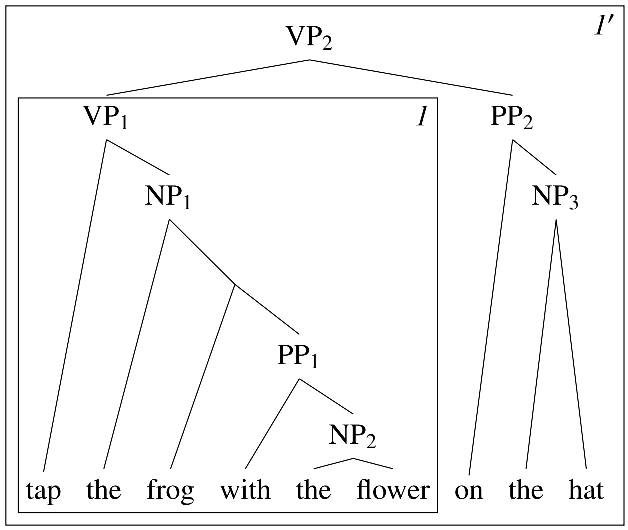
(36) **Left structure with two planning windows**



Planning in window 1 involves building the internal structure of VP_1 and computing its semantic value, i.e. the meaning of *tap the frog with the flower*. Once this is done, the internal structure of VP_1 can forgotten, relieving memory demand. All that is retained is the mother node (VP_1) and its semantic value. Then, planning in window 2 is commences: the internal structure of PP_2 is built up, and its semantic value, i.e. the meaning of *on the hat*, is computed. The meaning of PP_2 (from window 2) is then integrated with the meaning of VP_1 (from window 1). Note that if VP_1 is the first planning window, we take it that VP_2 is *not* an available node to target for the second planning window: VP_2 properly contains VP_1 and we assume that planning windows cannot overlap. Regarding prosodic phrasing, both planning windows in (36) should correspond to prosodic phrases and, accordingly, the predicted phrasing has a late boundary after *flower*, marking the end of of window 1.

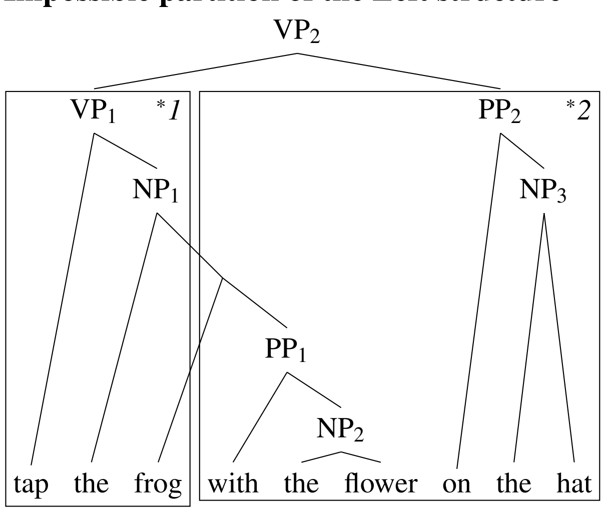
In this way, the Left structure can be computed either by processing the entire structure in one go, resulting in no boundaries (one window), or by first processing a smaller sub-constituent, resulting in a late boundary (two windows). We can visualize the optionality by plotting the two potential first planning windows in the same tree:

(37) **Left structure with a smaller (*1*) or bigger (*1′*) first planning window**



Critically, other prosodic renderings — e.g. a single early boundary after *frog*, or two boundaries after *frog* and *flower* — cannot be derived, as these would require planning windows that are non-constituents. This is illustrated in (38) (overleaf), which would correspond to a single early boundary after *frog*. Neither *tap the frog* nor *with the flower on the hat* are constituents in the Left structure.

(38) **Impossible partition of the Left structure**



The overall predictions of our proposal as to which prosodic renderings can and cannot be derived with the Left structure are summarized in (39):

(39) **Summary of prosodic predictions for the Left structure**

a. Tap the frog with the flower on the hat. *No boundaries; 1 window*
b. Tap the frog with the flower || on the hat. *Late boundary; 2 windows*
c. Tap the frog || with the flower on the hat. **Early boundary*
d. Tap the frog || with the flower || on the hat. **Two boundaries*

We now address how a given speaker chooses among the different derivable prosodic renditions, and how awareness affect their choice. We take it that different factors regulate the choice for aware and unaware participants. We discuss unaware participants first, and then aware participants.

Unaware participants: We assume that when speakers are not actively trying to disambiguate, considerations of ease of processing determine the choice between different planning options.

More specifically, two opposing pressures are at play. On the one hand, speakers will opt for a planning structure that avoids placing too high a demand on processing resources. Because planning an utterance in more windows requires less processing resources at each individual step than planning the same utterance in fewer windows, this pressure favors a parse with more windows. On the other hand, planning windows that are too simple will allocate too much time to each step, and not result in an efficient use of the existing processing resources. We take it that these pressures result in speakers planning as much material in a given window as possible, given limited processing resources.

We can understand experimental our result that participants in the unaware group most frequently render the Left structure with a late boundary if planning the utterance in one go would place too great a demand on processing resources. Our stimuli differ in this regards from the shorter stimuli in ST, where most unaware speakers produced no boundary.

Aware participants: We suggest that aware participants strategize so as to adopt a planning structure for each of the three readings which maximizes their prosodic differences. To do so, they tend to opt for more articulated planning structures which result in a greater number of boundaries. Even if it were not too demanding to plan the entire utterance in one go, planning in a single window is always a licit option under the SPW, so rendering the utterance with no boundaries is not informative about its internal syntax. Thus, aware speakers converge with unaware speakers in opting for the two window and produce a late boundary phrasing, converging with unaware speakers, but for different reasons.

Further issues: In addition to the phrasings discussed above, our proposal does predict certain other phrasings for the Left structure which are not observed, creating an apparent over-generation problem.

For instance, we predict (40) to be possible. This derives if the utterance is planned in three windows, the first being *tap*, the second *the frog with the flower*, and the third *on the hat*, all of which are constituents in the Left structure. Taking this to its extreme, we also predict (41) to be possible, where each terminal is planned in its own window, as all terminals are constituents.[11]

(40) Tap || the frog with the flower || on the hat.

(41) Tap || the || frog || with || the || flower || on || the || hat.

We believe that there are principled reasons for why these phrasings are not observed, as we explain with (40). For unaware speakers, the pressure to make efficient use of processing resources means that this partition is not likely to be chosen. *Tap* is sufficiently easy to process that there is no advantage to partitioning it in a separate window from *the frog with the flower*, so efficiency favors a bigger first window. For aware speakers, a boundary after *tap* is not strategic in cueing the Left structure. *Tap* is a constituent in all of the Left, List, and Right structures, so could be planned in its own window and followed by a boundary regardless of syntax. The partition in (41) is ruled out by similar reasoning.

In addition to apparent over-generation, there is also some apparent under-generation. Recall that our proposal predicts early boundary and two boundary phrasings to be underivable. Although this is mostly a desirable result since these phrasings are comparatively infrequent, they were occasionally observed according to the perceptual annotations. The question is: how could they arise at all?

We think that there are several possible interpretations of the data. One is annotator error: annotators might confuse one phrasing with another and miscategorize a late boundary or no boundary phrasing as an early or two boundary phrasing. Based on comparison of our two annotations, it appears that early boundary and no boundary phrasings are particularly confusable. A second source of error may be in the production experiment itself: if some participants do not fully absorb the context, they may have a different reading in mind from the one

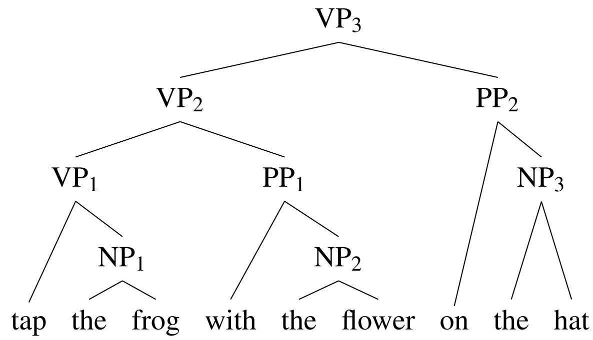
[11] Recall that Snedeker and Trueswell (2003) do report a boundary after *tap* in the NP-modifier reading of *Tap the frog with the flower* in their Exp. 1. With the shorter stimulus, that boundary is required to cue syntax: by putting a boundary after *tap*, the speaker shows that *the frog with the flower* can be planned together in the second window, diagnosing it as a constituent.

intended when they produce their utterance, and this could lead to unexpected phrasings in each condition. Finally, it could be that planning windows sometimes do comprise non-constituents after all, in cases that one could classify as disfluencies. More data is needed to explore which interpretation of the results is more accurate.

5.2.4. Deriving the prosody of the List structure

We now turn to the List structure, repeated in (42).

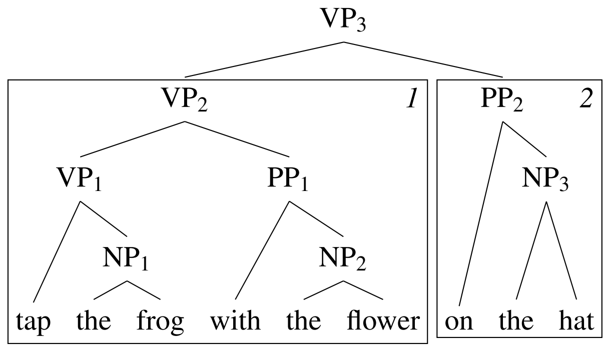
(42) **List structure**



There are at least three ways the utterance could be partitioned into planning windows, given the List structure. As with the Left structure, the first possibility is to plan the entire utterance in one shot, predicting the entire utterance to be a single intonational phrase.

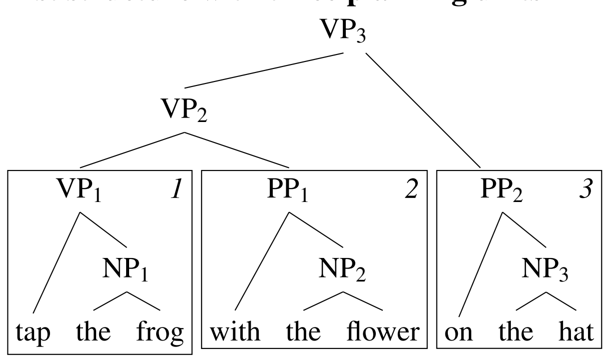
A second constituent which could constitute the initial planning window is VP_2, demarcated in (43). After the initial planning window, PP_2 remains to be planned and constitutes the second planning window, also demarcated. *Tap the frog with the flower* (VP_2) and *on the hat* (PP_2) are each intonational phrases as a result, yielding a late boundary rendering.

(43) **List structure with two planning windows**



The final possibility is that VP_1 constitutes the initial planning window, as in (44). After VP_1 is planned, the material that remains is *with the flower on the hat*, corresponding to PP_1 and PP_2. Because PP_1 and PP_2 do not together form a constituent, PP_1 and PP_2 must be planned in separate windows. PP_1 is planned in window 2 and then PP_2 is planned in window 3. *Tap the frog* (VP_1), *with the flower* (PP_1), and *on the hat* (PP_2) each constitute separate intonational phrases, yielding a two boundary rendering.

(44) **List structure with three planning units**



Depending on how the utterance is partitioned, no boundary (one window), late boundary (two windows), and two boundary (three windows) phrasings derive. An early boundary phrasing, however, cannot derive. A single boundary after *frog* would require *tap the frog* to be planned in one window, and PP_1 and PP_2 together in a second window. Because PP_1 and PP_2 do not form a constituent, this is not a licit partition. Possible and impossible phrasings are:

(45) **Summary of prosodic predictions for the List structure**

a. Tap the frog with the flower on the hat. *No boundaries; 1 window*
b. Tap the frog with the flower || on the hat. *Late boundary; 2*
c. Tap the frog || with the flower || on the hat. *Two boundaries; 3*
d. Tap the frog || with the flower on the hat. **Early boundary*

We can now account for the experimental results. A first observation from (45) is that the List structure is predicted to be compatible with a late boundary rendering, similar to the Left structure. Our proposal is thus consistent with the finding that the Left and List structures may not be prosodically distinct, and both be rendered with a late boundary. But, we still need to make precise how unaware and aware participants choose their partition.

Unaware participants: Recall our result for the Unaware group: according to the expert annotations, the List structure is most commonly realized with a late boundary, though two boundaries is also quite common; no boundaries is rare.

As seen with the Left structure, the utterance is sufficiently difficult to plan all at once that most participants opt to partition into multiple windows, accounting for the rarity of the rendering with no boundaries. Variability between two and three boundary renderings can be understood as variability in the processing resources dedicated to the task. Many participants budget sufficient resources that they can plan the utterance in just two windows, while some do not, in which case the more articulated three window partition obtains.

It will be helpful to rationalize this further by considering how planning proceeds incrementally, step-by-step. The speaker begins at the left corner of the tree, and builds the leftmost constituent, VP_1, *tap the frog*. They then arrive at a choice point: they can either (a) end the window, or (b) continue on to build more structure in the same window before interpreting. Depending on allocation of processing resources, VP_1 may or may not be complex enough to warrant wrapping up the planning window.

If the window is continued, structure is built up to the next node (VP_2). The speaker then again reaches a choice point. If they take option (a), they will end the initial window after *flower* and plan *on the hat* in a second window. If they take option (b), they will plan the entire utterance in a single window. Planning *tap the frog with the flower* seems to place a heavy enough demand on processing resources that participants usually opt to close the window, just as in the Left structure, and hence the two structures are frequently realized identically.

If the the speaker does end the first window after VP_1, they begin the second window by planning *with the flower*. Because *with the flower* does not form a constituent with the subsequent PP, the speaker must end the second window after *flower*, even if they have adequate processing resources to continue further. *On the hat* is then planned in the third window. Comparing the predictions for the Left structure in (39) and the List structure in (45), note that a three window partition like that in (45-c) is not possible with the Left structure, since *tap the frog* is not a constituent to the exclusion of *with the flower* in the Left structure (recall (38)). Accordingly, participants have no choice but to allocate sufficient resources to plan in (at most) two windows with the Left structure, and a two boundary rendering of the Left structure cannot obtain.

Aware participants: If participants strategize to adopt a planning structure which yields a prosody distinctive for the List structure, they will adopt the three window partition, and a two boundary rendering. Unlike the List structure, the Left structure and, as we will see, the Right structure are not compatible with a two boundary rendering.

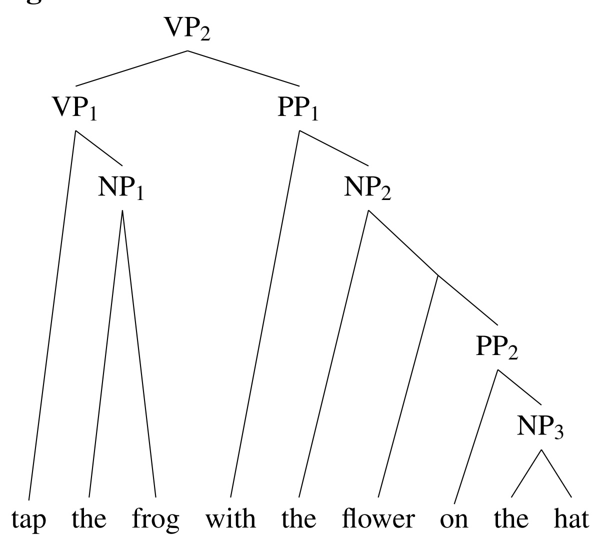
Overall, our model has achieved further success with the List structure: it accounts for the critical result that the late boundary phrasing is possible, provides a rationale for why unaware speakers might choose it over other options, and explains how a two boundary rendering also obtains — sometimes for unaware speakers and as the predominant phrasing for aware speakers.

5.2.5. Deriving the prosody of the Right structure

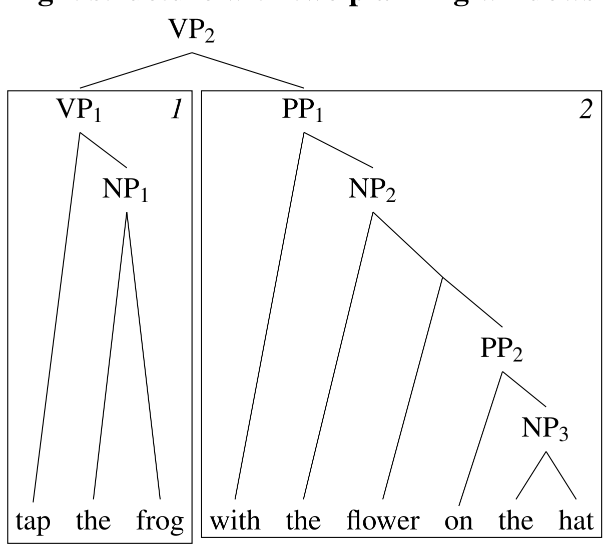
Finally, we consider Right structure, repeated in (46) (overleaf). In partitioning the utterance into planning windows, the first possibility, once again, is to take the initial planning window to be VP_2, and plan the entire utterance in one shot, predicting no prosodic boundaries.

The second possibility for an initial planning window is VP_1, shown in (47) (also overleaf) as window 1. This leaves PP_1, *with the flower on the hat*, to be planned in window 2. Regarding prosody, *tap the frog* (VP_1) and *with the flower on the hat* (PP_1) should be separate prosodic phrases, predicting an early boundary.

(46) **Right structure**

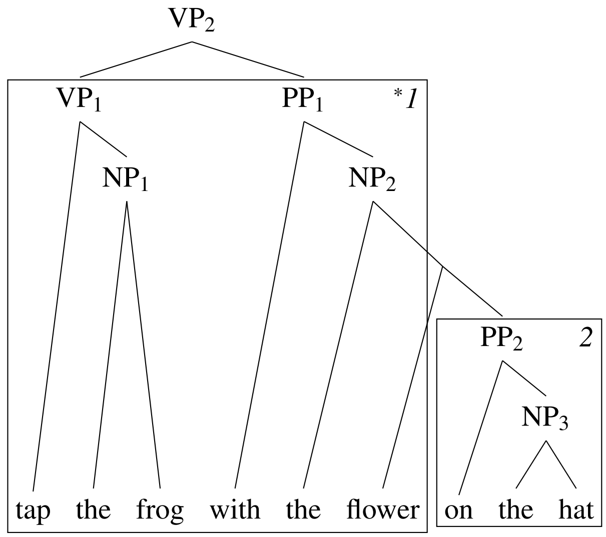


(47) **Right structure with two planning windows**



A rendition with a late boundary or two boundaries cannot be derived for the Right structure. To derive a late boundary, for instance, the partitioning into planning windows would have to be as in (48), where *tap the frog with the flower* is planned in the first window. This string does not constitute a constituent.

(48) **Impossible derivation of Right structure**



Overall, with the Right structure, no boundary and early boundary phrasings can be derived, but late boundary and two boundary phrasings cannot be.

(49) **Predictions for Right structure**

a. Tap the frog with the flower on the hat. *No boundaries; 1 window*
b. Tap the frog || with the flower on the hat. *Early boundary; 2*
c. Tap the frog with the flower || on the hat. **Late boundary*
d. Tap the frog || with the flower || on the hat. **Two boundaries*

Unaware participants: Recall that unaware participants frequently realized the Right structure with either no boundaries or an early boundary.

The occurrence of no boundary phrasing seems unexpected. With the Left and List structures, although a no boundary phrasing could be derived, it was rarely observed. Our idea to account for this was that our test sentence is too difficult to plan in one go: speakers must plan in multiple windows to alleviate memory demands. If so, it is surprising that participants in the Right condition would frequently plan their utterance in a single window (yielding no boundaries), when a partition into two windows is also possible.

The pattern demystifies when we consider how planning proceeds step-by-step. The Right structure, like in the Left structure, *tap the frog* is the leftmost maximal projection, VP_1. The speaker planning the Right structure begins building at the left corner of the tree and constructs VP_1. As seen with the Left structure, they then reach a choice point, where they can either (a) end the window after *frog*, or (b) decide to build one node up and include additional material in the initial window.

Suppose first that they do not allocate sufficient processing resources to continue the window and select (a). Then, PP_2 *with the flower on the hat* remains to be planned in the second window. This results in a two window partition and early boundary phrasing. As with the List structure, however, many participants do continue at the choice point. Such a speaker will opt for option (b) and continue planning after *frog* in the initial window. Because *with the flower on the hat* is a single constituent, the speaker then finds that there is no place to end the window until after *hat*, i.e. the end of the utterance. That is, after opting for (a) at the choice point, there is no further option but to plan the entire utterance in one window, resulting in a no boundary phrasing.

Although option (a) turns out to be a poor planning choice, anticipating this would require a global knowledge of the entire structure. If utterances are planned incrementally, speakers do not have the privilege to evaluate the utterance globally and decide on early sub-optimal planning windows to avert even more sub-optimal later planning windows.

Aware participants: Once again, the situation is straightforward in the aware group. If participants aim to keep the three readings prosodically distinct, they will opt for the more articulated two window partition, and an early boundary phrasing. The early boundary phrasing is a unique identifier of the Right structure.

5.3. Section summary

The experimental findings left us with two puzzles which pointed towards left-right asymmetries in how syntax and prosody correspond. Recall:

(20) (i) Unaware speakers preferentially realize List structures with a late boundary (as opposed to the expected realization with equal boundaries, or a realization with an early boundary)—resulting in a neutralization with the Left structure.

(ii) Unaware speakers preferentially realize Right structures without any boundaries (compared to Left and List structures, which are rarely produced without any boundary).

We have taken these generalizations to motivate a model where intonational phrases demarcate the edges of production planning windows (PPW), and syntax constrains planning such that planning windows must be constituents (SPW). For participants unaware of the ambiguity, the choice between phrasings is governed by a pressure to minimize processing cost, and an opposing pressure to plan the utterance in as few windows as possible. These pressures interact with one another and with the incremental nature of production planning to predict the generalizations. When aware of the ambiguity, participants strategize to maximize prosodic differences between alternative structures. To include more prosodic boundaries in their utterances, they tend to opt for more articulated planning structures, correctly predicting a late boundary with the Left structure, two boundaries with the List structure, and an early boundary with the Right structure.

The success of our model in accounting for the observed generalizations in the data set suggests that prosodic phrasing is not a purely grammatical phenomenon, but rather is a reflex of incremental planning and it interactions with syntax. Incremental planning can be motivated based on performance-based considerations such as alleviating working memory demands.

5.4. Comparison with the LRB

We conclude this section by point out some more specific differences between our approach and the LRB of Watson and Gibson (2004b). Recall the LRB: the likelihood of a boundary at a given position simply increases with the length of the constituent that immediately precedes (L='left') or follows (R='right') that position. Like our proposal, the LRB takes prosodic phrasing to be a tool to facilitate production planning.

One aspect of our results unexpected under the LRB is the asymmetry between Left and Right structures: if the size of both preceding and following constituents matters equally, why is a late boundary almost always observed with the Left structure, while an early boundary is compartively rare with the Right structure in the Unaware group? The constituent preceding the boundary in *Tap the frog with the flower* ‖ *on the hat* is of roughly equal complexity to the constituent following the boundary in *Tap the frog* ‖ *with the flower on the hat*.

This asymmetry could be accounted for with the LRB if we introduce an additional incrementally-inspired assumption: that the size of the constituent *preceding* a boundary has a bigger effect than the size of the constituent *following* the boundary. In fact, other theories have proposed similar asymmetries: Cooper and Paccia-Cooper (10 1980) considered the possibility that right or left edges might have a bigger effect for the placement of prosodic structure.[12]

[12] An idea adapted in prosodic phonology (Chen, 1987; Selkirk, 1986) with 'edge alignment'.

One place where our account sharply differs from the LRB is in its predictions for certain structures in which several constituents attach at a similar syntactic height. Consider the co-ordination in (50) on the given bracketing.

(50) (Sally and Bill) or Jolene or Sue.

The LRB predicts that a boundary after *Sally and Bill* should have increased likelihood given that it follows a complex constituent, while a boundary after *Jolene* should be unlikely, because the immediately adjacent constituents (the NPs *Jolene* and *Sue*) are each simple. The most likely phrasing should be:

(51) (Sally and Bill) || or Jolene or Sue.

Wagner (2005) tested how speakers produce structures like these, and found evidence from final lengthening that there tends to be a boundary both after *Sally and Bill* and after *Jolene*. In other words, the most likely phrasing is actually (52). The early boundary after *Sally and Bill* seems to commit a speaker to placing a late boundary after *Jolene*, given the structure show.

(52) (Sally and Bill) || or Jolene || or Sue.

Whereas such 'long-distance effects' (Wagner, 2005) (see also the 'distal' effects in Dilley and McAuley 2008) are unexpected if only the length of immediately adjacent constituents matters, our model correctly predicts the pattern. If *Sally and Bill* is planned in one window (deriving the boundary after *Bill*), the material remaining to be planned is *Jolene* and *Sue*. Because *Jolene* and *Sue* do not form a constituent, they must be planned in separate windows, resulting in each of them constituting a separate intonational phrase. This is similar to the phrasing pattern we observed above in our List structure, where placing an early boundary after *frog* committed speakers to also placing a late boundary after *flower*, since otherwise a non-constituent would have to be planned in one window.

Stated more generally, the LRB does not fully predict the following generalization, proposed by Taglicht (1998):

(53) Taglicht's Coordination Constraint
For any coordinating node X, if any two daughter nodes of X are separated by an I(ntonational) Ph(rase) boundary, all the daughter nodes of X must be separated by IPh boundaries.

This generalization is straightforwardly captured in the model presented here. As we saw, an early boundary in the List structure commits a speaker to also place a late boundary—the reason is that otherwise, the material in the second planning window would not form a constituent.[13]

6. Syntactic Variability

The idea that language production proceeds incrementally is uncontroversial and, as we discussed, the idea that intonational phrases might correspond to planning units not only helps account for out data, but also seems compatible with prior findings in the literature. One key ingredient of the model proposed here seems problematic, however: the idea that each planning window has to correspond to a syntactic constituent. Often, the content of an intonational phrase appears to be a non-constituent. In the following, we argue that such cases reveal a second factor that introduces variability in prosodic phrasing: the syntactic structure of utterances themselves can vary in systematically constrained ways.

6.1. Apparent non-constituents that form prosodic units

Let's consider an actual case. Steedman (1991) and Steedman (2000) discuss renditions of SVO sentences with an intonational boundary separating the direct object form the verb (our example is a variation of one in Steedman, 1991). Under usual assumptions, *Marcel proved*, the first intonational phrase, is not a constituent.

(54) Marcel proved || the hard theorem.

Similar possible phrasings that present apparent mismatches between syntactic constituency and prosodic phrasing have been observed by other authors:

(55) a. Everyone knows (||) that this is not true. (Taglicht, 1998)
b. She gave her friend (||) an interesting book. (Taglicht, 1998)

[13] A limitation of examples involving coordinate structures as in (50) is that the syntactic structure corresponding to the phrasing in (52) may be unlikely to be used by speakers in certain contexts, but is certainly possible in principle with the intended reading—all one needs is a context that motivates to group the last two constituents together. The data in this paper therefore provides a better argument for the long-distance effects observed in Wagner (2005). Further evidence for this type of effect at a distance is presented in (Hirsch and Wagner, 2015), based on parenthetical placement.

c. We know that this charge (||) is completely baseless. (Taglicht, 1998)

d. Tom washed and dried (||) the dishes. (McCawley, 1998)

e. Sesame street is presented by || the children's television network. (Shattuck-Hufnagel and Turk, 1996)

The situation is, however, less dire for our proposal than it appears. Referring back to (54), virtually every syntactic theory has a means to turn *Marcel proved* into a constituent. Steedman (1991, 2000), for example, assumes the framework of Categorial Grammar, in which a given meaning can often be built with different syntactic constituencies. An SVO sentence like *Marcel proved the hard theorem* can receive the constituent structure (Marcel proved) (the hard theorem) by virtue of a rule of composition that creates a predicate of the form *Marcel proved x*, and takes *the hard theorem* as the argument of that predicate.

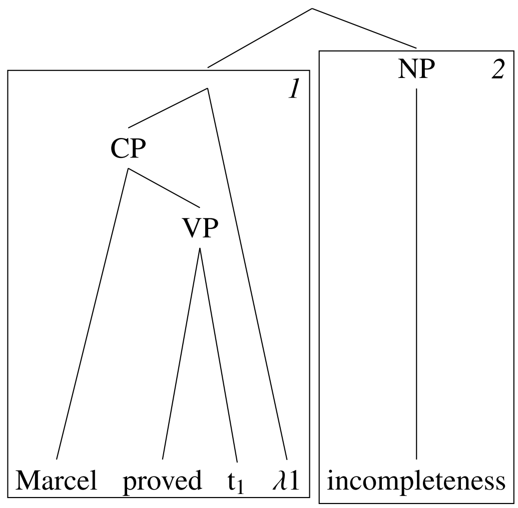
We will assume the corresponding mechanism in Generative Grammar: rightward movement. In the base structure, the constituency is (Marcel)(proved the hard theorem), but *the hard theorem* can move rightward to attach on the clausal spine, leaving a trace in its base position. The result is:

(56) $[_{TP}$ $[_{TP}$ Marcel proved $t_1]$ [the hard theorem]$_1]$

This movement dislocation has a semantic correlate: the trace is interpreted as a variable, and a lambda-operator that binds that variable enables the dislocated constituent to compose with the constituent it attaches to (cf. 'Predicate Abstraction' in Heim and Kratzer 1998).

Accordingly, after restructuring, *Marcel proved t* can be built and interpreted in one planning window, and then the output of that window can compose with *the hard theorem*, planned in a second window. The full partition:

(57) **Two planning units after restructuring**



The availability of rightward movement is independently detectable in examples like (58), where *the hard theorem* is displaced to the right of the sentential adverb *yesterday* (Ross 1967, Kayne 1989, i.a.).

(58) [[[Marcel proved t_1] yesterday] [the hard theorem]$_1$]

Although this instance of rightward movement (so-called "heavy NP shift"), is only available to constituents of sufficient complexity, rightward movement is more widely available. Right Node Raising in (59), for instance, involves conjunction of the apparent non-constituents *Trump assumed* and *Hillary refuted* and may be analyzed with rightward movement of *incompleteness* (e.g. Sabbagh 2007).

(59) a. Trump assumed and Hillary refuted incompleteness.
b. [[[Trump assumed t_1] and [Hillary refuted t_1]] [incompleteness]$_1$]

Mutatis mutandis, these constructions would provide evidence for forward composition or similar operations in Categorial Grammar.

Continuing to frame within Generative Grammar, it has been proposed that heavy NP shift has a processing advantage. Gibson (1998) suggests that a complex constituent is less demanding for working memory when that constituent comes later in the parse (see also Hawkins, 1994, 2004). In our view, rightward movement has a more general functional purpose: it can facilitate incremental production planning by creating new viable planning windows. In (54), *Marcel proved* and *the hard theorem* can be separated into two planning windows only with rightward movement of *the hard theorem*.

Many apparent counterexamples to our claim that each planning window is a constituent reconcile with our view once we take operations like rightward movement into account. In fact, non-standard syntactic constituency has been argued to be behind apparent prosodic mismatches in the earlier literature, e.g. Chomsky and Halle (1968) and Langendoen (1975) (non-standard bracketing derived by 'syntactic readjustment'); Steedman (1991, 2000) (derived by additional composition rules in a Categorial Grammar); Phillips (1996) (derived by left-to-right structure building), and Wagner (2005) (derived by allowing for alternative assemblies of list structures).

To bolster the argument that syntactic flexibility is behind cases of prosodic phrasing not otherwise predicted in our model, we should ask: do the constraints on this kind of prosodic variability mirror constraints on operations like rightward movement? Various authors have observed that there are limits to where optional prosodic boundaries may be placed:

(60) a. *But [$_{DP}$ almost || all of them] knew that. (Taglicht, 1998)
b. *[$_{DP}$ Danish || beer] is better. (Taglicht, 1998)
c. *[$_{DP}$ Three mathematicians || in ten] derive a lemma. (Pierrehumbert, 1980)
d. *[$_{DP}$ George and || Mary] give blood. (Shattuck-Hufnagel and Turk, 1996)
e. *Tom washed || and dried the dishes. (McCawley, 1998)

As Taglicht (1998) notes, there is a systematic pattern distinguishing the acceptable boundaries in (55) above from those in (60): whereas the good examples have an apparent non-constituent forming a prosodic domain on the left, the bad examples have an apparent non-constituent forming a prosodic domain on the right. In Taglicht's terminology, 'rightward annexation' is accepted, while 'leftward annexation' is not. Taglicht states the governing constraint as (62), which rightward, but not leftward annexation, respects.[14]

[14] Taglicht's constraint also correctly predicts that a head should not be separated from its complement unless later arguments are also separated by boundaries:

(i) *give || a book to John

In the present theory, this generalization, if true, can only be derived if *give a book* forms a con-

(61) Taglicht's Headed Structure Constraint
A headed node is ill-formed if it has a daughter ending in an I(ntonational) P(hrase) boundary followed by a daughter not ending in an IP boundary.

Rather than needing to stipulate a special constraint, however, the data are directly predicted if rightward movement is the mechanism responsible for non-constituent intonational phrases. Bing (1985, 43) reports on a 1978 talk by Mark Liberman, who observed similar asymmetries. Liberman noted that if optional prosodic boundaries reflect rightward movement, non-constituents should always be *initial* within a sentential domain: rightward movement displaces a constituent to the right, and potentially strands a non-constituent. The requisite constituent structure for the bad cases in (60) would be impossible to derive by rightward movement, since the material following the boundary is constitute a constituent.

Not only is the contrast between rightward and leftward annexation consistent with adding rightward movement to our model as a source of prosodic variability, but it is unclear how alternative models could account for it without appeal to rightward movement. For the LRB, while the licit examples in (55) do involve boundaries after long strings of words, these strings do not form constituents based on standard assumptions, so the *constituent* preceding the site of the boundary is not particularly long (and can, in fact, be quite short). Moreover, the LRB doesn't provide a reason for the asymmetry between left and right annexation. Once one adopts the assumptions of syntactic restructuring we make here, then the LRB makes correct predictions. However, we believe that once one allows for this syntactic restructuring, there is nothing that speaks against our more restrictive assumption that every planning window has to be a constituent.

We could take (55) to show something quite different: that syntax and prosodic phrasing simply do not match. The Sense Unit Condition we mentioned earlier was designed in this spirit (Selkirk 1984, p. 293, using parentheses to mark prosodic constituency):

(62) a. (Jane gave the book to Mary).
b. (Jane gave the book) (to Mary).
c. (Jane gave) (the book) (to Mary).
* (John gave) (the book to Mary).

stituent to the exclusion of *to John*. This is a potential problem that we will not discuss here. For experimental evidence that heads are unlikely to be separated from their complement by a boundary, see Watson et al. (2006).

Under standard syntactic assumptions, *Jane gave the book* and *Jane gave* do not form a constituent, and yet there are legitimate intonational patterns that treat these as phrases: (62-b) and (62-c). The Sense Unit Condition holds that prosodic phrasing is in fact *not* in tune with syntax, but rather reflects semantically coherent units: ‘intonational phrasing is syntactically free but semantically constrained’ (p. 290). The definition of a sense-unit is as follows:

(63) The Sense Unit Condition on Intonational Phrasing
Two constituents C_i, C_j form a sense unit if either (a) or (b) is true of the semantic interpretation of the sentence:
  a. C_i modifies C_j (a head)
  b. C_i is an argument of C_j (a head)

While we agree with the SUC that intonational phrases are a domain that has to be semantically interpretable, the assumption of semantic domains that mismatch syntactic constituency goes against the Principle of Compositionality. Our approach is semantically compositional, as discussed.[15]

6.2. More variation than we need?

We have proposed that prosodic variability directly reflects variability in the size of planning windows. However, we have now also identified a second source of variability: there may be multiple syntactic parses for the same overall meaning. In our model, syntactic variability affects prosody only indirectly, as different syntactic structures enable different partitions into planning windows. We must ask, though: do we really need both sources of variability to account for our data? In fact, we *do*: syntactic variability could not by itself predict our key generalizations. Let us play out what predictions would arise if syntax automatically mapped to prosody, with all prosodic variability linked to syntactic variability.

Recall our first generalization: that the List reading is conveyed with a late boundary by unaware speakers. The structure assumed for the List reading earlier in the paper was (64-a), where *with the flower* and *on the hat* each attach as VP-level modifiers. The corresponding phrasing in (64-b) has two boundaries.

(64) a. [$_{VP}$ [$_{VP}$ [$_{VP}$ tap the frog]] [with the flower]] [on the hat]]
  b. Tap the frog || with the flower || on the hat.

[15] Empirical problems with the sense unit condition were discussed in Taglicht (1998) and Watson and Gibson (2004a), but we will not explore these here.

Given the possibility of rightward movement, however, another structure should also be available: the structure in (65-a), where *on the hat* extraposes out of the VP to adjoin higher, for instance to TP. If the syntax-prosody mapping has a higher TP-level modifier preceded by a stronger boundary than a lower VP-level modifier, the phrasing would be (65-b), with a late boundary.

(65) a. $[_{TP}$ $[_{VP}$ $[_{VP}$ $[_{VP}$ tap the frog]] [with the flower]] t_1] [on the hat]$_1$]
b. Tap the frog with the flower || on the hat.

With syntactic variability considered, the direct view can derive the range of observed renderings for the List reading: without movement, a flat prosody is predicted (as in the Aware group); with movement of *on the hat*, a late boundary is predicted (as in the Unaware group). Yet, the account is incomplete. Although each phrasing *can* derive, there is no explanation for *when* each phrasing should obtain. It remains a mystery why the structure without movement would be favored by aware speakers — and, in particular, why the more complex structure with movement would be favored by unaware speakers. Although movement is an available process, movement cannot in itself explain our data.

Our second generalization is even more challenging for the view that the only source of variability is the syntax. Recall: the Right reading is conveyed with no boundaries in the unaware group. The structure assumed for the Right reading was (66-a), where the PP headed by *with* properly contains *on the hat*, predicting the early boundary phrasing in (66-b), as observed in the Aware group.

(66) a. $[_{VP}$ $[_{VP}$ tap the frog] [with the flower on the hat]]
b. Tap the frog || with the flower on the hat.

To derive a flat prosody, *on the hat* would have to extrapose out of its containing PP and adjoin to the VP, as in (68). This way, *with the flower* and *on the hat* are both attached as VP-level modifiers.

(67) $[_{VP}$ $[_{VP}$ $[_{VP}$ tap the frog] [with the flower t_1]] [on the hat]$_1$]

The structure in (67) is not sufficient in two ways. First, if all VP-modifiers are preceded by boundaries of equal strength, we might expect two boundaries from this structure; the most frequent flat rendition of the Right structure in fact involved no boundaries. The second problem is more severe: even if (67) could derive a no boundary phrasing, it can be independently shown that (67) is not an available structure.

It is possible to force extraposition of *on the hat* by inserting an adverbial preceding it, as in (68-a), where *please* is inserted.

(68) a. Tap the frog with the flower please on the hat.
b. [$_{VP}$ [$_{VP}$ [$_{VP}$ [tap the frog] [with the flower t_1]] please] [on the hat]$_1$]

Intuitively, (68-a) can convey the Left reading or the List reading — but not the Right reading. We verified this intuition in a rating experiment reported in Hirsch and Wagner (2015). Participants were presented with sentences like (69-a)-(69-c) in a context that motivated the Right reading.[16] The mean naturalness rating for (69-b) was significantly lower than that for (69-a) and (69-c), indicating that extraposition of *on the hat* out of its containing PP is marginal or unavailable.

(69) a. Tap the frog *please* with the flower on the hat.
b. Tap the frog with the flower *please* on the hat.
c. Tap the frog with the flower on the hat *please*.

Thus, the flat phrasing for the Right reading in the unaware group cannot be derived with reference to syntactic constituency alone, even with rightward movement and similar operations that allow for different types of syntactic structures taken into account.

7. Conclusion

Prosodic phrasing is variable. In this paper, we presented data revealing asymmetries in that variability that have not been observed previously. These asymmetries shed new light on a puzzle in the literature on prosodic disambiguation, namely why some studies find that prosody is a reliable cue to resolve attachment ambiguity, while others find that it is not. We have defended the hypothesis that at least some conflicting results trace to inherent syntactic differences between the ambiguities studied.

We argued that syntax does not itself map to prosodic phrasing, but plays a key indirect role in constraining prosodic variability. Following ideas in the production planning literature, we proposed that intonational phrases correspond to planning windows. Prosodic variability reflects flexibility in planning. By inte-

[16] In fact, the experiment comprised all readings, but we only summarize one aspect of the results here. The experiment involved 9 different sets of sentences of this kind, which were presented in a Latin Square design to 27 participants. The participants were instructed to rate on a 7-point Likert scale the naturalness of the sentence given the context.

grating this model with theoretical insights about the syntax/semantics interface, we were able to motivate a syntactic constraint on planning: planning windows must constitute syntactic constituents. The interaction of this constraint with reasonable assumptions about processing and performance considerations predicted the variability pattern in our experiment.

Our strong syntactic assumption ran into problems when considering other types of data, where apparent non-constituents form intonational phrases. Based on observations from the previous literature, we argued that these apparent mismatches with syntax are limited to certain configurations, where syntactic operations such as rightward movement are available to turn the apparent non-constituent into a constituent. The observed restrictions on such apparent mismatches are not explained by alternative models that are not similarly syntactically constrained.

By invoking string-vacuous rightward movement to account for apparent non-constituent intonational phrases, our proposal suggests that a reasonably high proportion of utterances that speakers produce involve syntactic restructuring. One might find it implausible that restructuring would be so ubiquitous. There are two responses to this worry. First, because the existence of rightward movement is established by overtly detectable displacement (e.g. with extraposition across sentential adverbs, and Right Node Raising), we know that rightward movement is an available mechanism in the grammar. There is, then, no reason to expect that it can't occur in cases that are string-vacuous. Without considering prosody, our syntactic tools might just not be precise enough to detect it.

As mentioned briefly in Section 6, we think that a more explanatory response seems plausible, however. We can ask: why would mechanisms that allow for extraposition exist in the first place? Our proposal offers an answer. Language generally lets us construct complex messages, many of which might be too complex to plan within a single window. The solution natural languages employs so that processing limitations do not unduly limit the messages we can express are precisely the mechanisms that derive extraposition. Rightward movement is what enables us to chunk complex material into multiple licit planning windows and thus make it more easily processable. The string-vacuous cases of restructuring might actually be the more typical cases.

Since most current theories of sentence processing in perception assume that similar mechanisms are at play as in speech production (MacDonald, 2013; Pickering and Garrod, 2013), these would merit discussion also in the context of our proposal: it would be fruitful to explore the relationship between production models and parsing models. In the NLP literature on parsing, there is a family of algorithms that convert syntactic structure to facilitate 'left-corner' incremental parses, essentially a way to remove left-branching and facilitate top-down processing (Johnson, 1998; Johnson and Roark, 2000). We think that it would be fruitful to try to look at flexibility in planning scope and at flexibility in syntactic structure from this point of view. Our assumptions about constraints on syntactic restructuring are more restrictive than 'left-corner' transforms often assumed in the NLP literature. Our results also raise interesting questions for recent semantic theories that allow for a strictly left-to-right interpretation of linguistic structure. For the example, the theory of 'continuations' (Barker, 2002; Shan and Barker, 2006) offers a powerful compositional framework to interpret linguistics expressions in an incremental way. It has recently been argued to provide an insightful way to look at the online processing of natural language in Bott and Sternefeld (2016). Here again, it would be interesting to explore to what extent the syntactic constraints on incrementality observed in our data could be accommodated or even rationalized within such approaches.

Acknowledgements

Both authors receive financial support from the Social Sciences and Humanities Research Council of Canada. Further acknowledgements to be added in the final version.

References

Allbritton, D., McKoon, G., and Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of experimental psychology: Learning, Memory, and Cognition*, 22(3):714–735.

Barker, C. (2002). Continuations and the nature of quantification. *Natural language semantics*, 10(3):211–242.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Bing, J. M. (1985). *Aspects of English prosody*. Garland.

Bock, J. and Levelt, W. (1994). Language production: Grammatical encoding, ma, handbook of psycholinguistics, 1994, 945-984. In Gernsbacher, editor, *Handbook of psycholinguistics*, pages 945–984. Academic Press, San Diego.

Bott, O. and Sternefeld, W. (2016). An event semantics with continuations for incremental interpretation. MS. Eberhard Karls Universität Tübingen.

Chen, M. Y. (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook*, 4:109–49.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper & Row, New York.

Clifton, C. J., Carlson, K., and Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45(2):87–114.

Cooper, W. E. and Paccia-Cooper, J. (1980). *Syntax and Speech*. Harvard University Press, Cambridge, Mass.

Dilley, L. and McAuley, J. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59(3):294–311.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100:233–253.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In Wheeldon, L. R., editor, *Aspects of language production*, pages 291–330.

Ferreira, V. and Dell, G. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4):296–340.

Gee, J. and Grosjean, F. (1983). Performance structures: A psycholinguistic appraisal. *Cognitive Psychology*, 15:411–458.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68:1–76.

Grosjean, F. and Collins, M. (1979). Breathing, pausing, and reading. *Phonetica*, 36:98–114.

Hawkins, J. (1994). *A performance theory of order and constituency*. Cambridge Studies in Linguistics. Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press Oxford.

Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell, Oxford.

Hirsch, A. and Wagner, M. (2015). Rightward movement affects prosodic phrasing. In Ozyildiz, D. and Bui, T., editors, *Proceedings of the 45th Meeting of the North-East Linguistic Society*.

Johnson, M. (1998). Finite-state approximation of constraint-based grammars using left-corner grammar transforms. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 619–623. Association for Computational Linguistics.

Johnson, M. and Roark, B. (2000). Compact non-left-recursive grammars using the selective left-corner transform and factoring. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*, pages 355–361. Association for Computational Linguistics.

Kraljic, T. and Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50:194–231.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of phonetics*, 35(2):162–179.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Langendoen, D. T. (1975). Finite-state parsing of phrase-structure languages and the status of readjustment rules in grammar. *Linguistic Inquiry*, 6(4):533–554.

Lashley, K. S. (1951). The problem of serial order in behavior. In A., J. L., editor, *Cerebral mechanisms in behavior*, pages 112–136. Wiley, New York.

Lee, E.-K., Brown-Schmidt, S., and Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, 129(3):544–562.

Lehiste, I. (1973). Phonetic disambigation of syntactic ambiguity. *Glossa*, 7:107–122.

Levelt, W. (1989). *Speaking. From Intention to Articulation*. MIT Press.

Lindsley, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, 7(1):1–19.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in psychology*, 4.

McCawley, J. D. (1998). *The Syntactic Phenomena of English*. Chicago University Press, Chicago, 2nd edition.

Norcliffe, E. and Jaeger, T. (2005). Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain. In *Proceedings of Interspeech 2005*.

Phillips, C. (1996). *Order and Structure*. PhD thesis, MIT.

Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04):329–347.

Pierrehumbert, J. (1980). *The phonology and phonetics of English Intonation*. PhD thesis, MIT.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90(6):2956–2970.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4):904.

Schafer, A. (1997). *Prosodic parsing: The Role of prosody in sentence comprehension*. PhD thesis, University of Massachussetts, Amherst, Ma.

Schafer, A. J., Speer, S. R., Warren, P., and White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29(2):169–182.

Schriefers, H., Teruel, E., and Meinshausen, R.-M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language*, 39(4):609–632.

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

Selkirk, E. O. (1984). *Phonology and Syntax. The relation between sound and structure*. MIT Press, Cambridge, MA.

Shan, C.-c. and Barker, C. (2006). Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy*, 29(1):91–134.

Shattuck-Hufnagel, S. and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2):193–247.

Snedeker, J. and Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48:103–130.

Snedeker, J. and Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of memory and language*, 58(2):574–608.

Steedman, M. (1991). Structure and intonation. *Language*, 67(2):260–296.

Steedman, M. (2000). Information structure and the phonology-syntax interface. *Linguistic Inquiry*, 34:649–689.

Sternberg, S., Monsell, S., Knoll, R., and Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information processing in motor control and learning*, pages 117–152.

Sternberg, S., Wright, C., Knoll, R., and Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. *Perception and production of fluent speech*, pages 507–534.

Taglicht, J. (1998). Constraints on intonational phrasing in English. *Journal of Linguistics*, 34:181–211.

Truckenbrodt, H. (1995). *Phonological phrases: Their relation to syntax, focus, and prominence*. PhD thesis, MIT, Cambridge, Mass.

Wagner, M. (2005). *Prosody and Recursion*. PhD thesis, MIT.

Watson, D. G., Breen, M., and Gibson, E. (2006). The Role of Syntactic Obligatoriness in the Production of Intonational Boundaries. *Learning, Memory*, 32(5):1045–1056.

Watson, D. G. and Gibson, E. (2004a). Making sense of the sense unit condition. *Linguistic Inquiry*, 35:508–517.

Watson, D. G. and Gibson, E. (2004b). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*.

Wheeldon, L. (2012). Producing spoken sentences: The scope of incremental planning. In Fuchs, S., Weirich, M., Pape, D., and Perrier, P., editors, *Speech production and perception: Vol. 1. Speech planning and dynamics*, pages 97–118. Peter Lang, Frankfurt, Germany.

Appendix A. Experimental Materials

Item 1

- Left: John is in the forest. He sees a frog who is holding a flower. The frog is wearing a hat. You want John to reach over and use his finger to tap the frog's hat. This is what you say to him: *Tap the frog with the flower on the hat.*
- List: John is in the forest. He sees a frog who is wearing a hat. There is a flower nearby. You want John to take the flower and use it to tap the frog's hat. This is what you say to him: *Tap the frog with the flower on the hat.*
- Right: John is in the forest. He sees a frog. The frog's hat is on the ground. There is a flower on top of it. You want John to take the flower and use it to tap the frog. This is what you say to him: *Tap the frog with the flower on the hat.*

Item 2

- Left: Sally is in the stable. She sees a horse who has a carrot in its mouth. The horse is wearing a saddle. You want Sally to reach over and use her finger to tap the horse's saddle. This is what you say to her: *Tap the horse with the carrot on the saddle.*
- List: Sally is in the stable. She sees a horse who is wearing a saddle. There is a carrot on the ground. You want Sally to take the carrot and use it to tap the horse's saddle. This is what you say to her: *Tap the horse with the carrot on the saddle.*
- Right: Sally is in the stable. She sees a horse. The horse's saddle is on the ground. There is a carrot on top of it. You want Sally to take the carrot and use it to the tap the horse. This is what you say to her: *Tap the horse with the carrot on the saddle.*

Item 3

- Left: John is at the mall. He sees a girl who has a feather in her hair. The girl is wearing gloves. You want John to reach over and use his finger to tap the girl's glove. This is what you say to him: *Tap the girl with the feather on the glove.*

- List: John is at the mall. He sees a girl who is wearing gloves. There is a feather on the counter. You want John to take the feather and use it to tap the girl's glove. This is what you say to him: *Tap the girl with the feather on the glove.*

- Right: John is at the mall. He sees a girl. The girl's gloves are lying on the counter. There is a feather on top of one of them. You want John take the feather and use it to tap the girl. This is what you say to him: *Tap the girl with the feather on the glove.*

Item 4

- Left: Sally is on a ship. She sees a sailor who is holding a compass. The sailor is wearing a coat. You want Sally to reach over and use her finger to tap the sailor's coat. This is what you say to her: *Tap the sailor with the compass on the coat.*

- List: Sally is on a ship. She sees a sailor who is wearing a coat. There is a compass on the table. You want Sally to take the compass and use it to tap the sailor's coat. This is what you say to her: *Tap the sailor with the compass on the coat.*

- Right: Sally is on a ship. She sees a sailor. The sailor's coat is on the table. There is a compass on top of it. You want Sally to take the compass and use it to tap the sailor. This is what you say to her: *Tap the sailor with the compass on the coat.*

Item 5

- Left: Sally is in the park. She sees a hiker who is holding a bottle. The hiker is wearing a pack. You want Sally to reach over and use her finger to tap the hiker's pack. This is what you say to her: *Tap the hiker with the bottle on the pack.*

- List: Sally is in the park. She sees a hiker who is wearing a pack. There is a bottle on the ground. You want Sally to take the bottle and use it to tap the hiker's pack. This is what you say to her: *Tap the hiker with the bottle on the pack.*

- Right: Sally is in the park. She sees a hiker. The hiker's pack is on the ground. There is a bottle on top of it. You want Sally to take the bottle and use it to tap the hiker. This is what you say to her: *Tap the hiker with the bottle on the pack.*

Item 6

- Left: John is at the beach. He sees a dog who is holding a shell in its mouth. The dog is wearing a sweater. You want John to reach over and use his finger to tap the dog's sweater. This is what you say to him: *Tap the dog with the shell on the sweater.*

- List: John is at the beach. He sees a dog who is wearing a sweater. There is a shell in the sand. You want John to take the shell and use it to tap the dog's sweater. This is what you say to him: *Tap the dog with the shell on the sweater.*

- Right: John is at the beach. He sees a dog. The dog's sweater is on the ground. There is a shell on top of it. You want John to take the shell and use it to tap the dog. This is what you say to him: *Tap the dog with the shell on the sweater.*

Item 7

- Left: Sally is at school. She sees a boy who is wearing glasses. The boy has a necktie on. You want Sally to reach over and use her finger to tap the boy's tie. This is what you say to her: *Tap the boy with the glasses on the tie.*

- List: Sally is at school. She sees a boy who has a necktie on. There is a pair of glasses on the table. You want Sally to take the glasses and use them to tap the boy's tie. This is what you say to her: *Tap the boy with the glasses on the tie.*

- Right: Sally is at school. She sees a boy. The boy's necktie is on the desk. There is a pair of glasses on top of it. You want Sally to take the glasses and use them to tap the boy. This is what you say to her: *Tap the boy with the glasses on the tie.*

Item 8

- Left: John is at the circus. He sees a clown who is holding a balloon. The clown is wearing big red shoes. You want John to reach over and use his finger to tap the clown's shoe. This is what you say to him: *Tap the clown with the balloon on the shoe.*

- List: John is at the circus. He sees a clown who is wearing big red shoes. There is a balloon on the ground. You want John to take the balloon and use it to tap the clown's shoe. This is what you say to him: *Tap the clown with the balloon on the shoe.*

- Right: John is at the circus. He sees a clown. The clown's big red shoes are on the ground. There is a balloon sitting on top of one of them. You want John to take the balloon and use it to tap the clown. This is what you say to him: *Tap the clown with the balloon on the shoe.*

Item 9

- Left: Sally is in the jungle. She sees a monkey who is holding a pen. The monkey is wearing a jacket. You want Sally to reach over and use her finger to tap the monkey's jacket. This is what you say to her: *Tap the monkey with the pen on the jacket.*

- List: Sally is in the jungle. She sees a monkey who is wearing a jacket. There is a pen on the ground. You want Sally to take the pen and use it to tap the monkey's jacket. This is what you say to her: *Tap the monkey with the pen on the jacket.*

- Right: Sally is in the jungle. She sees a monkey. The monkey's jacket is on the ground. There is a pen on top of it. You want Sally to take the pen and use it to tap the monkey. This is what you say to her: *Tap the monkey with the pen on the jacket.*

Appendix B. Instructions for the two groups

High Awareness/Within group: Welcome! Thank you for participating! On each of the following slides, you will see a short passage that describes a particular situation and an action to be performed by an individual in that situation. At the end of the passage, there will be a sentence instructing the individual to perform the specified action. You will be asked to read this instruction aloud. The instruction will be introduced by the phrase: "This is what you say to him/her:". When you see a new slide, please read all text silently to yourself. The sentences vary in their bracketing, and the story will make it clear which meaning is intended. A sentence might group words together in one of three ways: (tap the gnome with the umbrella)(on the hat) meaning "tap the gnome that has an umbrella, and tap him on the hat"; or (tap the gnome)(with the umbrella)(on the hat) meaning "use the umbrella to tap the gnome on the hat"; it could mean (tap the gnome)(with the umbrella on the hat) meaning "using the umbrella that is on the hat, tap the gnome". Please read the sentence silently first. When you're read, press a key to record and say it as if you were talking to another person. It is important that you say it in a natural way, just as if you were in a conversation. Note that you should not read the entire passage aloud, just the final instruction. Once you have finished, press the space key again to continue. If you have any questions, please ask the experimenter now! Press the space key to begin!

Low Awareness/Between group: Welcome! Thank you for participating! On each of the following slides, you will see a short passage that describes a particular situation and an action to be performed by an individual in that situation. At the end of the passage, there will be a sentence instructing the individual to perform the specified action. You will be asked to read this instruction aloud. The instruction will be introduced by the phrase: "This is what you say to him/her:". When you see a new slide, please read all text silently to yourself. You may read it as many times as you wish. You should be comfortable with its content before proceeding. Be sure to read the instruction following the passage to yourself enough times that you feel able to say it aloud fluently. Once you are ready, press the space key and read the instruction aloud. Please read the sentence as if you were talking to another person. It is important that you say it in a natural way, just as if you were in a conversation. Note that you should not read the entire passage aloud, just the final instruction. Once you have finished, press the space key again to continue. If you have any questions, please ask the experimenter now! Press the space key to begin!

Appendix C. Awareness questionnaire for the Between group

Q1: What do you think the experiment was about? All of the sentences you read in this experiment could be interpreted in different ways. Take, for example, "Tap the frog with the flower on the hat." This could be understood in three different ways. You could be asking someone: (1) to tap a frog who is holding a flower; (2) to use some flower to tap a frog; or (3) to take a flower that is on top of a hat and use that flower to tap a frog.

Q2: Were you aware during the experiment that the sentences could mean more than one thing? those explained above?

Q3: If so, were the different interpretations you had thought of the same as those explained above?