Jackson, S.R., Townsend, D.J., and Bever, T.G. (2005) Separating similar effects of conjunction and intonation in the resolution of lexical ambiguity. ***Studia Linguistica*** **59(2/3), pp.259-278**

SEPARATING SIMILAR EFFECTS OF CONJUNCTION AND INTONATION IN THE RESOLUTION OF LEXICAL AMBIGUITY*

Scott Jackson, David Townsend & Thomas Bever

Abstract. Recent work has shown that prosodic information has a number of important effects on human sentence processing, including effects on the semantic integration of clauses (Schafer 1997). In an effort to learn more about the nature of these prosodic effects, we ran an experiment that manipulated not only prosody, but also the type of initial subordinating conjunction, which we predicted might have a similar effect on processing. Our results confirmed previous work on the effects of conjunction type (Townsend & Bever 1978; Townsend 1983), and they replicated the basic pattern of the prosodic effect from Schafer (1997). However, a more in-depth analysis suggests (i) that prosody and conjunction type are playing different, interacting roles in this "makes sense" decision task, and (ii) that the prosodic effect may be more directly dependent on the gradient phonetic feature of phrase-final lengthening, rather than on an abstract phonological prosodic category.

1. Introduction

Prosodic features, including pitch, loudness, and duration, are ubiquitous in spoken language, but we still have much to learn about how they are processed in the minds of speakers and listeners, and how they affect other aspects of language processing. One way to learn more about the nature of prosodic effects on processing is to study how they interact with other non-prosodic factors. In this paper, we examine a prosodic effect that on the surface resembles an effect produced by the choice of conjunction in a subordinate clause. Based on the findings of past research (Schafer 1997; Townsend & Bever 1978) it initially occurred to us that both prosody and the type of clause-initial subordinating conjunction are cues for the semantic integration of a clause. Do these factors produce the same underlying effect, or are they somehow independent? Based on the work of Schafer (1997) and Townsend & Bever (1978), we carried out an experiment in an effort to answer this question. Our results suggest that the effects are underlyingly different, and that we might need to revise the categorical aspect of Schafer's hypothesis of prosodic boundaries and semantic integration.

Schafer (1997) reported an effect of prosody, whereby a certain prosodic boundary encourages the early resolution of lexical ambiguity.

*We would like to thank Allison King for helping to run subjects and for assisting with the phonetic measurements. Thanks also to the many helpful comments on this manuscript provided by an anonymous reviewer. All errors that remain are solely the fault of the authors. Correspondence should be addressed to the first author.

Studia Linguistica 59(2/3) 2005, pp. 259–278.
Published by Blackwell Publishing Ltd., 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

She assumed a prosodic theory following Pierrehumbert (1980) (see also Nespor & Vogel 1986; Selkirk 1984; etc.) and instantiated in the Tones and Break Indices system (ToBI; Silverman et al. 1992; Beckman & Ayers-Elam 1997; etc.). Because our experiment is based on Schafer's (1997) methodology, we will use this system's terminology throughout. In this system, prosodic constituents are arranged in hierarchical layers and are represented by abstract labels consisting of pitch accents and boundary tones. Every utterance (at least in English) must consist of at least one Intonational Phrase (IPh), which must end with a high or low boundary tone (H% or L%). Every IPh must contain at least one Phonological Phrase (PPh), which is a constituent that requires both a pitch accent (H* or L*), and a final phrase accent (H- or L-).

In her experiments, Schafer (1997) was interested in testing her Interpretive Domain Hypothesis:

(1) "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."

That is, an IPh boundary (H% or L%) forces early semantic integration, but a PPh boundary (H- or L-) neither encourages nor discourages early semantic integration. In this paper, we will use the term "semantic integration" to refer to the mental process of integrating the meanings of words into the meaning of the rest of the clausal or pragmatic context. For semantically polysemous words, this would presumably involve integrating only one of the meanings of the word – most likely the meaning that best fits the context. If the context is neutral between possible meanings, ambiguity arises. Semantic integration could be delayed until a point of disambiguation, or the processor could go ahead and select a single meaning based on some other factor, such as frequency, plausibility, or likeliness of a meaning. Schafer (1997) constructed a methodology exploiting this kind of lexical ambiguity to test when semantic integration occurs during processing.

Schafer had subjects listen to sentences and judge whether or not the sentence "made sense" by pressing one of two levers. Reaction times were measured from the end of the sentence stimuli. For her stimuli, she constructed sentences similar to the following:[1]

(2) a. Because the glasses are ugly, Stacy won't wear them.
b. Because the glasses are ugly, Stacy won't drink from them.

Each sentence is composed of two clauses and begins with a subordinate clause. There is an ambiguous word (e.g. *glasses*) in the first clause, and no disambiguating lexical content arrives until some point in the second

[1] This is one of our experimental materials, but it very closely resembles one of Schafer's sentences, as do most of our materials.

clause (e.g. *wear* or *drink*). If the processor is prompted to integrate the meaning of *glasses* before the disambiguating context is heard, there is a chance that it will select a meaning that will not integrate well with the following main clause (e.g. 'spectacles' in sentence (2b)). This should result in longer reaction times for judgments in the conditions where the wrong meaning is selected. This longer reaction time presumably represents some kind of recovery or repair process, in which the processor must change its initial meaning assignment in order to judge the sentence sensical.

In order to somewhat control the meaning the processor would select in this scenario, lexical items were chosen that had a bias for one meaning over another. Schafer (1997) established these biases by a written pretest, using different subjects. Subjects were given only the initial clause and were asked to write a completion for the sentence. Words were chosen for the makes-sense experiment that had a strong bias towards one of the possible meanings of that word. For example, in (2) above, the second clause of (2a) fits with the preferred meaning of *glasses* as 'spectacles,' while (2b) fits with the less preferred meaning of 'beverage containers.'

This design allows us (and Schafer 1997) to infer that if (enough) semantic integration takes place early on, the 'spectacles' meaning is the most likely initial choice for the processor for both (2a) and (2b). We may or may not expect that the reaction times to the sentences with preferred continuations (2a) would be faster in general than those with dispreferred continuations like (2b), but whatever the preferred-dispreferred difference is, it should be increased if significant integration occurs between the ambiguous word (e.g. *glasses*) and the disambiguating context. If we can assume this, then any factor that encourages early semantic integration should increase the reaction time advantage of the preferred meaning (2a) over the dispreferred meaning of (2b).

Schafer's (1997) Interpretive Domain Hypothesis makes the prediction that prosody is one such factor that can encourage early semantic integration. In one of her conditions the items contained an IPh boundary between clauses, and in the other condition they contained only a PPh boundary. Her results showed that when there was an IPh boundary between clauses, reaction times were significantly slower than those in the PPh condition for the sentences of type (2b), but not slower for sentences like (2a). These results suggest that a strong prosodic boundary (i.e. an IPh boundary) is a cue to the processor to integrate meaning early, or even immediately.

In order to test how this prosodic factor might interact with a non-prosodic factor, we needed a non-prosodic factor that is also a cue for early semantic integration. The type of subordinating conjunction seemed to be a good candidate for such a factor, based on the clausal processing work of Townsend & Bever (e.g. 1978, 2001; Townsend et al. 2000). Results first reported in Townsend & Bever (1978) showed that when listeners heard initial subordinate clauses beginning with a strongly causal conjunction, *if*, they performed faster on a task that asked them to evaluate whether a visually-presented paraphrase was related in meaning to what they had heard, compared to when the clauses started with an adversative conjunction, *though*. When the task was changed to identifying simply whether or not a visually-presented word had occurred at all in the preceding auditory stimulus, subjects' response times were more dependent on the location of the target word when it was in an initial *though* clause than when it was in an initial *if* clause. Townsend (1983) showed that when listeners heard a fragment of an initial subordinate clause and then saw a word, their time to read the word was more strongly influenced by syntactic consistency between the spoken fragment and the printed word when the fragment was introduced by *if* rather than *though*. Townsend (1983) reported similar patterns with other connectives, showing that the results for *if* extended to other causal conjunctions like *because*, and the pattern for *though* extended to other adversative conjunctions like *although*.

One way to interpret these results is that initial *because* (and other causal conjunctions) encourages early semantic integration and less attention to literal form (i.e. recalling exactly which words were used), while initial *although* encourages later semantic integration in favor of retaining access to literal form. Giving a full explanation of why these conjunctions are expected to have these effects is not within the scope of this article, but the explanation given by Townsend, Bever, and colleagues (see especially Townsend 1983) has to do with effective processing strategies given the causal relations between clauses. For our purposes here, the results show that subjects are more sensitive to meaning in clauses with *because* than in clauses with *although*, and one possible view is that this means *because* prompts earlier semantic integration than *although*. In other words, the choice of conjunction alters how early semantic integration takes place. Because Schafer's materials all began with subordinate clauses,[2] it was a relatively simple matter for us to use her methodology to test the interaction of effects that conjunction type and intonational boundary have on semantic integration. This experiment and the results are reported below.

2. Experiment

2.1. *Methods*

We used virtually the same methodology as the experiment from Schafer (1997) as described above. Like Schafer, we manipulated the factors of

[2] Schafer's materials were well balanced with various conjunctions, so it would appear that the choice of conjunction was not a confound for her results.

intonation (IPh boundary or PPh boundary) and disambiguated meaning (preferred or dispreferred). In addition, we manipulated the initial conjunction. All experimental items began with either *because* or *although*. An example of the full paradigm is given in (3), and (3a-h) are arranged in their respective conditions in Table 1:

(3) a. Although the glasses are ugly $_{\mathbf{PPh)}}$, Stacey will wear them.
b. Although the glasses are ugly $_{\mathbf{PPh)}}$, Stacey will drink from them.
c. Although the glasses are ugly $_{\mathbf{IPh)}}$, Stacey will wear them.
d. Although the glasses are ugly $_{\mathbf{IPh)}}$, Stacey will drink from them.
e. Because the glasses are ugly $_{\mathbf{PPh)}}$, Stacey won't wear them.
f. Because the glasses are ugly $_{\mathbf{PPh)}}$, Stacey won't drink from them.
g. Because the glasses are ugly $_{\mathbf{IPh)}}$, Stacey won't wear them.
h. Because the glasses are ugly $_{\mathbf{IPh)}}$, Stacey won't drink from them.

Table 1. The paradigm sentences arranged by condition

***Although* sentences**	PPh boundary	IPh boundary
Preferred meaning	3a	3c
Dispreferred meaning	3b	3d
***Because* sentences**	PPh boundary	IPh boundary
Preferred meaning	3e	3g
Dispreferred meaning	3f	3h

16 items were constructed, each with the full paradigm. A male ToBI-trained native English speaker recorded several examples of each sentence in a sound-attenuated booth. From these recordings, pieces were digitally cut and spliced, so that within an item paradigm, common materials between conditions were exactly the same acoustic tokens. For example, the same recording of *although* was used for (3a-d) and the same recording of *the glasses are ugly*$_{\mathbf{IPh)}}$ was used for (3c), (3d), (3g), and (3h). The purpose of this was to ensure the exact same prosody across the same conditions within an item. We checked the resulting items to make sure that they still sounded natural and that there were no audible acoustic artifacts of the splicing technique.

Phonetic measurements were taken in order to confirm the prosodic judgments of PPh vs. IPh boundaries. We chose two different phonetic measures for the boundary, one based on pitch movement and the other on phrase-final lengthening, two major cues to prosodic boundaries in English (Pierrehumbert 1980). The tonal pattern for all of the IPh conditions was a H* L-H% contour. We measured the fundamental frequency (F_0) at the peaks of the H* pitch accent and the H% boundary tone and at the lowest point of the L- tone. We then calculated the difference between the low F_0 and the average of the two peaks. The idea was to get a measure for how strong the overall pitch movement was, and not just the final rise.[3] For phrase-final lengthening, we simply measured the length of the phrase-final syllable. The distribution of length measurements overlapped more between PPh and IPh conditions than the pitch measurements did, given that there are other factors for syllable duration such as metrical position and segmental length. Nevertheless, this turned out to be a very useful measure, both for distinguishing the PPh vs. IPh categories, as well as in later correlation analyses. As expected, these phonetic measurements showed very significant differences between the two prosodic boundary categories (both $p < 0.0001$). These measurements confirmed that the prosody was correct and significantly different for the different prosodic conditions, and they were used in later analyses, described below. The means by category are given in Table 2.

Table 2. Means of phonetic measurements by prosodic condition (Std Error)

	PPh condition	**IPh condition**
$(F_{0H^*} + F_{0H\%})/2 - F_{0L-}$	5 Hz (0.24)	44 Hz (0.31)
Phrase-final syllable length	276 ms (3.35)	385 ms (3.57)

64 filler items were created with various syntactic and prosodic structures. Eight lists were constructed with a Latin-square design, so that each list contained one condition from each item, with a total of two items in each condition. The 16 experimental items per list were semi-randomly distributed throughout the filler items, and these positions and ordering were fixed and common to each list.

68 undergraduates at the University of Arizona participated as subjects, in exchange for psychology course credit. All reported speaking English as their first language and having normal hearing. Subjects listened to one experimental list over headphones in a sound-attenuated booth. They were instructed to listen to the entire sentence in each trial, and then click the left mouse button if the sentence made sense or the right mouse button if the sentence did not make sense. 40 of the filler sentences were either ungrammatical or were nonsensical in some way, in order to give subjects an even distribution of "yes" and "no" responses.

[3] We did also calculate the final rise from the L- to the H% and included it in our analyses, but this was never a better effect predictor than the contour difference, and the contour difference was slightly better correlated with the IPh vs. PPh distinction than just the final rise ($r^2 = 0.90$ vs. 0.85).

SuperLab© experimental software was used to present the stimuli and record the responses and reaction times.

2.2. *Data coding and exclusions*

Responses to experimental items were coded as errors when the subjects responded "no" for the makes-sense judgment, since all experimental sentences were sensical. Error rates were analyzed, but errors were excluded from the reaction time analyses (total 22.4%) because it was assumed that on error trials, subjects were not successfully completing the process of semantic integration that we needed them to do in order for us to evaluate the reaction times. The error rate for nonsense trials (filler trials for which "no" was the correct answer for the makes-sense judgment) was much lower (7.9%), but this was comparable to the error rate for just the preferred-meaning experimental trials (8.6%). This suggests that the nonsense trials, which were included only to balance the kind of stimuli presented to the subjects, were of comparable difficulty as the preferred-meaning experimental trials, which is as expected.

Responses faster than 200 milliseconds (ms) and slower than 2.5 times the overall standard deviation (5031 ms) were also excluded (4.3%), because it is likely that subjects were doing something very different in these trials. After excluding these and the errors, if any subject had fewer than 10 (out of 16) data points remaining, the entire subject was excluded, on the basis that these subjects might not be cooperating with the task. This excluded 14 subjects. The total data excluded by the above methods was 37.3% of the total data from experimental items. The remaining data was analyzed and the results are presented below.

3. Results

One of the assumptions underlying this experiment was that the preferred meaning conditions actually were preferred by the subjects. We used the same lexical items as Schafer (1997) in similar contexts, and we carried out a written pre-test similar to Schafer's in order to confirm the biases that she found. Additionally, after the experiment was completed, we analyzed the error rates. Errors were much more frequent in the dispreferred meaning conditions (35% overall error rate) than in the preferred meaning conditions (10% overall error rate). A chi-squared test confirmed that this was significant ($\chi^2 = 94.572$, $p > 0.0001$). This confirms our assumption that the preferred and dispreferred meanings were overall correctly categorized. Chi-squared tests revealed no significant effect of intonation or conjunction on error rates ($\chi^2 = 2.483$, $p = 0.478$ and $\chi^2 = 1.425$, $p = 0.700$, respectively). An omnibus ANOVA performed on the reaction times of the errors also showed a significant effect of meaning ($F(1, 198) = 5.55$, $p < 0.02$), but no other effects or interactions involving intonation or conjunction reached significance (all $p < 0.25$ or worse).

More interestingly, the reaction times for errors were reversed from what we will see later for the non-errors. Namely, reaction times for the preferred meanings (mean RT of 2311 ms) were slower than the reaction times for the dispreferred meanings (mean RT of 1805 ms). As noted above, these differences were statistically significant. A possible explanation for this is that there were many more errors on trials that resolved to a dispreferred meaning because subjects were simply responding too quickly, before their processing systems had time to "find" the alternate, dispreferred meaning. This view expects that the bulk of errors on dispreferred meanings will be on the faster end of the RT scale, which explains the RT pattern for errors given above. No further patterns were noted in the error data, and all following analyses use only the data from the trials with correct responses.

We replicated the basic pattern of Schafer (1997), namely a significant interaction of intonation and meaning. Our results showed this interaction significantly by Items ($F_2(1, 13) = 6.14$, $p < 0.028$) and by Subjects ($F_1(1, 22) = 3.34$, $p < 0.082$).[4] Our results appear to confirm those of Schafer (1997) in that the presence of an IPh boundary between clauses gave rise to a larger difference between the preferred and the dispreferred meaning conditions. The overall means averaged across conjunction type are given in Table 3. The preferred-dispreferred difference was on average 323 ms greater for the IPh conditions than for the PPh conditions.

Table 3. Mean reaction times by intonation and meaning, in milliseconds (Std Error)

	PPh boundary	**IPh boundary**	
Preferred meaning	1500 (66.2)	1387 (66.2)	
Dispreferred meaning	1782 (78.5)	1992 (74.8)	**Effect of IPh**
Difference (pref – dispref)	282	605	323

Finding a clear effect of conjunction proved to be more problematic. The hypothesis that *because* encourages early semantic integration predicts that the preferred-dispreferred difference should also be greater for *because* conditions than for *although* conditions. Simply given the means (Table 4), this appears to be the case, but the interaction of

[4] The by Subject test is significant because we are testing a one-tailed directional hypothesis.

intonation by conjunction was not statistically significant by Subjects[5] (F_1 (1, 52) = 1.77, $p < 0.189$) or by Items (F_2 (1, 14) = 1.41, $p < 0.255$).

Table 4. Mean reaction times by conjunction and meaning, in milliseconds (Std Error)

	although	*because*	
Preferred meaning	1455 (66.6)	1432 (65.9)	
Dispreferred meaning	1807 (77.4)	1974 (75.9)	**Effect of *because***
Differences (pref − dispref)	352	542	190

Numerically, this effect seems smaller than the intonational effect as well. Where the preferred-dispreferred difference was 323 ms greater for IPh vs. PPh conditions, it was only 190 ms greater for *because* vs. *although*, though it is in the predicted direction.

This suggestive but inconclusive result for the effect of conjunction led us to carry out several additional analyses, in order to understand just what role the conjunction was playing in this experiment. First, we carried out planned contrast analyses on the effect of meaning for each combination of intonation and conjunction. Table 5 gives a summary of the mean reaction times for each condition, the differences between the preferred meaning reaction time and the dispreferred meaning reaction time, and an indication of which of these differences was statistically significant. Figure 1 shows these results in graphical form. The F-statistics for this data are given in the following text.

Table 5. Summary of reaction times by condition, in milliseconds (Std Error)

	***although*, PPh**	***although*, IPh**	***because*, PPh**	***because*, IPh**
Pref. meaning	1558 (94.4)	1353 (93.9)	1444 (93.0)	1419 (93.4)
Dispref. meaning	1747 (112.2)	1861 (106.8)	1816 (109.8)	2117 (104.7)
Difference	189	508	372	698
Significant?	**no**	**yes**	**yes**	**yes**

[5] Because there was a low number of subjects with data in all eight conditions, we took advantage of planned comparisons whenever possible. For example, in this analysis we were not concerned with intonation, so we only analyzed the data with the factors of conjunction and meaning. This allowed us to use 53 subjects as opposed to only 23, because we only needed half as many conditions to be filled.

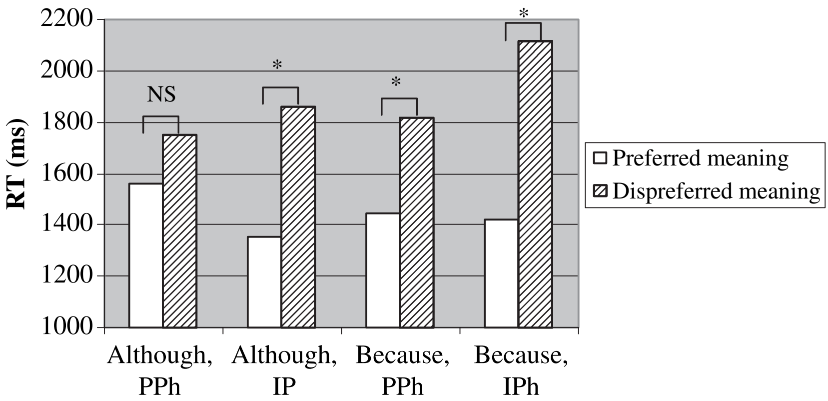

Figure 1. Graph representation of Table 5

In the conditions with *although* and a PPh boundary, there was no effect of meaning by Subjects or by Items (F_1 (1, 42) = 1.89, $p < 0.177$; F_2 (1, 13) = 2.25, $p < 0.158$), although the difference was in the same direction as the other conditions. In all other conditions, there was a significant effect of meaning both by Subjects and by Items: for *although* IPh conditions (F_1 (1, 43) = 14.96, $p < 0.001$; F_2 (1, 13) = 18.63, $p < 0.001$), for *because* PPh conditions (F_1 (1, 45) = 5.96, $p < 0.019$; F_2 (1, 13) = 7.78, $p < 0.015$), and for *because* IPh conditions (F_1 (1, 45) = 16.48, $p < 0.001$; F_2 (1, 13) = 21.49, $p < 0.001$). These results suggest that either the presence of an IPh boundary or the presence of *because* was sufficient to produce a significantly slower reaction time for the dispreferred meaning condition, because the only condition that did not show a significant preferred-dispreferred difference was the *although*/PPh condition. This pattern supports the hypothesis that *because* encourages early semantic integration, even though the effect may have been too small for the conjunction by meaning interaction to reach significance.

This still does not give us a picture of how the intonation and conjunction may be interacting. The data from Tables 3 and 4 suggest that there is an effect of intonation on the preferred-dispreferred advantage. That is, the advantage is increased by about 323 ms in the IPh conditions over the PPh conditions. Similarly, the conditions with *because* show a 190 ms greater preferred-dispreferred advantage. If we arrange the preferred-dispreferred differences from Table 5 (the third row in the chart) by intonation and conjunction, shown below in Table 6, we can see that the effects on these differences are consistent across the other factor. That is, the effect of intonation is around 323 ms for conditions with *although* as well as *because*, and the effect of conjunction is around 190 ms for conditions with PPh boundaries as well as IPh boundaries.

Table 6. Reaction time differences between preferred and dispreferred meaning conditions, by intonation and conjunction (ms)

	PPh	**IPh**	Difference
Although	189	508	319
Because	372	698	326
Difference	183	190	

This consistency of effects suggests further that the effect of conjunction is real, though it was too small to be detected by the interaction ANOVA test. It also suggests that if this pattern is legitimate, that the effects of conjunction and intonation are purely additive.

However, there is still the possibility that this apparently additive pattern is more complex. If we return to the data in Table 5 and Figure 1, we can see that although the IPh boundary increases the gap between the preferred and dispreferred meaning conditions across the board, it might be doing so in ways that are lost in the generalization of Table 6. For *although*, most of the increase of the gap is created by speeding the responses to the preferred meaning rather than by slowing the responses to the dispreferred meaning (205 ms vs. 114 ms, respectively). For *because*, this pattern is reversed, such that the IPh barely speeded the responses to preferred meanings if at all, while dispreferred meanings were much slower (25 ms vs. 301 ms, respectively). In order to test whether this pattern was substantive, we ran a series of *post hoc* analyses.

So far, all of our analyses have used the categorical PPh vs. IPh distinction in describing the effects of intonation. In the following analyses, we decided to replace this categorical measure of intonation with the gradient phonetic measures we took to indicate the relative strength of the prosodic boundary. We wanted to explore the possibility that finer acoustic cues could provide us with better predictors of the observed effects than the gross abstract characterization of IPh vs. PPh. The phonetic values were neither within item nor within subject factors, so we used an omnibus ANOVA. Interestingly, the different phonetic correlates of the phonological boundaries produced different results in the analyses. The two different methods we used to measure the degree of pitch contour (see section 2.1) showed weaker interactions between meaning and intonation (rise: $F(1, 674) = 2.60$, $p < 0.11$ and fall + rise: $F(1, 674) = 3.40$, $p < 0.06$) than an omnibus test using the categorical IPh/PPh measure ($F(1, 674) = 5.05$, $p < 0.03$), but the measure of final syllable length appeared to be just as sensitive as the categorical measure ($F(1, 674) = 4.73$, $p < 0.03$). Using the pitch measurements, no other effect reached significance. However, when we used the measure of phrase-final syllable duration, the interaction of conjunction and intonation became significant ($F(1, 674) = 4.30$, $p < 0.039$). This effect did not reach significance using the categorical measure of intonation ($F(1, 674) = 1.70$, $p < 0.193$).

Therefore, if we trust the pattern of results produced in the analysis using clause-final syllable length, prosody interacts with both meaning and conjunction in this task. That is, the longer the clause-final syllable, the greater the reaction time advantage of preferred meaning over dispreferred meaning, as well as the advantage of *although* over *because*. This gives statistical support for the pattern that we noticed above. An IPh boundary appears to have a consistent effect whereby the RT advantage of the preferred meaning over the dispreferred meaning is increased. However, as we noted above, it also appears as if this advantage is produced differently in the different conjunction conditions. In sentences with *although* (the "slow integration" conjunction), the preferred meaning is made faster, and for *because* (the "fast integration" conjunction), the dispreferred meaning is made slower.

We further tested this specific interpretation with correlation analyses, correlating final syllable duration with reaction time. In conditions with *although*, phrase-final syllable length was negatively correlated with reaction time (the predicted direction) in the preferred meaning conditions ($p < 0.012$, $r = -0.181$), but uncorrelated in the dispreferred meaning conditions ($p < 0.629$, $r = -0.041$). With *because*, syllable length was uncorrelated with reaction time in the preferred meaning conditions ($p < 0.508$, $r = 0.048$) but marginally correlated in a positive direction (the predicted direction) for the dispreferred meaning ($p < 0.102$, $r = 0.135$).[6] This backs up the interpretation of the interactions presented above, and the pattern is summarized in Table 7.

Table 7. Summary of correlation results: Effect of greater final syllable lengthening

	Although	***Because***
Preferred meaning	Faster	No effect
Dispreferred meaning	No effect	Slower

Before we continue, it may be helpful to quickly review the main findings so far. We have replicated the basic pattern shown by Schafer (1997), that stronger intonational boundaries produce greater reaction time differences between sentences that disambiguate for a preferred

[6] Again, because we are making a directional and thus a one-tailed hypothesis, $p < 0.10$ is significant here, because the result is in the predicted direction.

meaning vs. those that disambiguate for a dispreferred meaning. This is predicted by any theory that claims that the time course of semantic integration can be affected by intonation.

Furthermore, we found results that confirm the findings of Townsend & Bever (1978 and following). The causal conjunction *because* also produces a greater preferred-dispreferred reaction time advantage than the adversative conjunction *although*. This, too, is predicted as long as our theory allows initial conjunction type to also modulate semantic integration.

However, while at first blush these effects seem to be simply additive (see Table 6), our deeper probing has suggested something more complex. By examining the different directions of the effects of intonation and conjunction, and by doing some *post hoc* correlations using finer phonetic measurements than IPh/PPh, our results suggest that the effect of intonation may be somehow modulated by the type of conjunction. Namely, it appears that intonation primarily facilitates preferred meanings in the context of *although*, but hinders dispreferred meanings in the context of *because*. Finally, it is important to note that our analyses have also suggested that final syllable length is a better predictor for the effects under investigation than other pitch-based phonetic properties of prosodic boundaries. In the final section, we will propose an explanation of how this interaction might work and come to some conclusions about the significance of these findings.

4. Discussion

Our initial question asked whether the effects of intonation and conjunction are independent, or underlyingly part of the same process. The results that show a significant interaction of intonation and conjunction suggest that these effects are somehow different or independent. Now we must try to understand how they are different, but also how they seem to be related. Before we propose a full story, we need to look more closely at the properties of these two effects.

So far we have somewhat simplified the role of the conjunctions by focusing on the fact that *because* encourages early semantic integration. The other side of the coin is the relative access to literal form, i.e. the actual lexical items. The theory presented by Bever & Townsend (1979; Townsend & Bever 2001) claims that there is a trading relation between processing meaning (i.e. semantic integration) and retaining literal form. *Because* encourages the processing of meaning but allows less access to literal form, while *although* allows more access to literal form at the cost of less emphasis on meaning. Another way to look at this is that there are two basic processes at work – retaining access to literal form and processing meaning – and the conjunction affects how processing resources are divided among these.

In the analysis that follows, we will hypothesize that allocating more resources to semantic integration means that any semantic processing that occurs is able to finish more rapidly. In this view, *although* allows semantic processing, but that processing is not completed as quickly as it is with *because*, for the reason that the two conjunctions allocate processing resources differently. Including literal form in the discussion is important to our analysis as well, because we assume that access to literal form plays a major role in recovering from a semantic misanalysis. For example, if the processor assumes that the meaning of *glasses* is 'spectacles,' but must recover from this when it hears "…Stacy will wear them", it will be able to find an alternate meaning and recover faster if it can easily recall the word *glasses* rather than simply the meaning of 'spectacles.'

Now let us re-examine the effect of prosody. Why should prosody have this kind of effect, and more specifically, why should duration be the important phonetic cue? One possibility, in the spirit of Schafer (1997), is that because of the typical alignment of large prosodic boundaries with large syntactic constituents, the processor recognizes prosodic boundaries as good places to integrate meaning. The boundary encourages a kind of wrapping up of processing because it is likely that the sentence or at least a major clause ends at such a boundary.

Another possibility is that the boundaries facilitate processing simply by giving the processor more time. This second idea is more closely related to the production-centered theory of Watson & Gibson (e.g. 2002, this volume). They hypothesize that the speaker makes prosodic breaks at points in the sentence when more processing time is needed, predicting that the duration of these breaks is the key factor. Because our results have picked out final lengthening as the important cue for the effect on semantic processing, we will adopt a view along these lines. Our hypothesis, *contra* Schafer (1997), is that prosody does *not* provide a categorical instruction for the processor to wrap up any remaining semantic integration, but rather it gives the processor more time to complete any current processing.[7]

We are now in a position to propose a basic model of how these effects are interacting in this experiment. There are two critical representations, literal form and integrated meaning. There are two processes, one that allows access to literal form, and one that does semantic integration and selection of one meaning from a polysemous word. Recall that in our task, this amounts to selection of the preferred meaning. In the theory we are formulating, *because* allows faster semantic integration (by allowing

[7] It is important to note that this claim is not unique to final lengthening, but to any pause as well. All pauses between clauses were digitally spliced out of our materials, which might explain why final lengthening alone was such a good cue. We would expect any interclausal pause to simply add to the effect of final lengthening.

more processing resources to be devoted to it), but only a weak link to the literal form. *Although* grants easier access to the literal form, but semantic integration takes longer. Prosody, specifically the duration of the phrase-final syllable, represents extra time available for any semantic integration, because this is essentially "dead" time in which no new lexical information is coming in to the processor.

This model makes the following predictions. For clauses with *although*,[8] fewer resources are allocated to semantic processes, meaning that it will take longer for semantic integration to be carried out. The processor therefore does not have enough time to integrate the preferred meaning in the conditions with weak prosodic boundaries (i.e. short final syllables), but it does have enough time with the strong prosodic boundaries. This correctly predicts that the preferred meaning should be facilitated by prosody with *although*, because early integration of the correct meaning should save time later on. The dispreferred meaning is not at a much greater disadvantage, though, because the relatively easy access to surface form provides helpful information to any repair strategy that activates when the wrong meaning is selected. This access to surface form is continually provided by *although*, and should not change significantly with prosody. This correctly predicts that prosody should not give the dispreferred meaning a disadvantage for *although*.

With *because*, more resources are allocated to semantic processing, meaning that the process of semantic integration needs less time, so it will tend to get far enough in both short and long prosodic boundaries. This correctly predicts that with *because*, prosody will not give the preferred meaning any further advantage. However, since *because* has only a weak link at best to literal form, if the wrong meaning (i.e. a preferred meaning in a context that later disambiguates for a dispreferred one) is selected by integration, "repair" strategies are hindered by the difficulty in recovering the literal form. This correctly predicts that with *because*, the extra time provided by strong prosodic boundaries will make it harder to later access literal form (because deeper integration further "buries" the literal form), thus giving the dispreferred meaning a disadvantage.

This view contrasts in several important ways from the view of Schafer (1997). First of all, semantic integration is an ongoing process in this model, not a process that only occurs at distinct points like clausal or prosodic boundaries. Secondly, intonation is not a "control" cue that prompts integration. It simply provides a resource, extra time, for a process that is already ongoing. Similarly, conjunctions like *because* and *although* act not as direct cues to "start" or "wait" to process meaning, but rather they modulate the mental resources dedicated to semantic

[8] For accuracy, it should be pointed out that the effects of *because* vs. *although* are also dependent on whether or not they are in sentence-initial or sentence-final clauses. The interested reader should see Townsend & Bever (1978 and following) for more details.

processing, thereby adjusting the relative speed of ongoing integration. This is fully consistent with the view of *because*- and *although*-type effects in Townsend & Bever (1978), and Townsend (1983).

This model not only makes the correct predictions for the results discussed above, but it provides an explanation for why prosody has the observed effects, and why duration at the prosodic boundary is the crucial factor for that effect. However, there are a few points that should be brought up as potential weaknesses of this model.

First and foremost, there are a few major concerns that we need to acknowledge.[9] Our initial claims about the properties of *because* and *although* are based on Townsend & Bever's (1978 and following) work which shows general patterns for a range of different conjunctions. It may be possible that since we used only these two exact conjunctions in our experimental items, the effects observed might be due to their idiosyncratic lexical properties (including frequency, plausibility, etc.), and not due to the general effects on semantic processing observed throughout the larger classes of similar conjunctions. If this concern is taken seriously, the claim would have to be that *because* is more frequent or plausible (or something else) with preferred meanings than with dispreferred meanings, and the opposite would have to be true of *although*. It may be possible to conceive a plausible independent reason for why this might be true. For example, since *although* functions to deny an anticipated event, it might turn out to be more frequent with dispreferred or unexpected meanings of words. With *because*, one could reasonably expect that it would occur more frequently with the more frequent (preferred) meanings, almost by definition. This would appear to offer an independent explanation for why *because* and *although* have the basic effects shown in our results.

There are a few ways to respond to this concern. First of all, this is an empirical question that needs further testing. Simply using a wider range of conjunctions than just *because* and *although* will not really suffice, because a similar rationale as the frequency-based one given above could work just as well for any of the conjunctions we would expect to produce the effects under investigation, such as *though* and *if*. We would need to first establish whether *although* is in fact more frequent with dispreferred meanings and vice versa for *because*. If it is not the case, then this concern simply evaporates. However, even if it is the case, it would suggest a different possible explanation than what we have given here, but it would not actually provide evidence against our story. The pattern of different frequencies with different meaning preferences is still quite consistent with our theory. That is, if we are right, then *because* has a better chance of creating problems (misanalysis, etc.) when a dispreferred meaning is being used. If speakers are tacitly aware of this increased burden on the

[9] Our thanks to an anonymous reviewer for raising several of these issues.

listener, then there might be an impact on the frequency of these situations. While these questions are interesting and important for the further validation or rejection of our theory, they are unfortunately out of the scope of the present paper, and so we must leave these issues aside for the moment for further research.

However, even if the reasons behind the conjunction effect are not what we now believe, this does not take away from the results that suggest that Schafer's (1997) Interpretive Domain Hypothesis may be an oversimplification of a complex, dynamic process. Indeed, if something like the lexical frequency of a clause-initial conjunction can modulate the processing effects of a clause-final intonational boundary, the situation is much more complex than even we suspect.

Another major concern is that we have based too much interpretation on a myriad of *post hoc* analyses. If we simply managed to get the facts wrong by overextending our statistics, then the story must change completely. It may in fact be a simpler story in which the effects of intonation and conjunction are simply additive, in which case we should have stopped with Table 6. However, even if this is the case, there are still issues with Schafer's (1997) story that need to be resolved. For example, if an IPh boundary categorically "defines a point at which the processor performs any as yet outstanding semantic/pragmatic evaluation and integration", then it is not clear how the effects of an IPh boundary and *because* could be cumulative, if it is such an all-or-nothing kind of process. A possible way out for Schafer's (1997) view would be that even though the process itself is all-or-nothing in each individual trial, the probability of it occurring at an IPh (or even the probability of an IPh being perceived) is not 100%, and so the cumulative effect of *because* reflects a cumulative probability of the effect, not a cumulative effect size.

If this turns out to be closer to the truth, then a further question arises: what other factors could contribute to the probability of semantic integration, in addition to the prosodic and conjunction-based effects we see here? What mechanism if any coordinates these various factors, and what would be the motivation for such a system? In any case, future research should provide a more robust confirmation or rejection of our more in-depth analyses. Until then, we will assume that all of the results reported here underlie real effects and are not artifacts of familywise error, because in our view, the results here taken together provide a cohesive, explanatory picture of a complex interaction.

Returning to our model, there are two stipulated properties of the model that can be tested. One is that in sentences with *although*, access to literal form should not be affected by how much semantic integration takes place. This allowed us to predict that prosody should not make a difference for the dispreferred meaning with *although*. The other stipulation is the converse, that with *because*, access to literal form *should* be affected by how much semantic integration takes place. An explanation of such a pattern would not necessarily be problematic for our story. For example, it could simply be a matter of degree. Our results so far have suggested that some integration must take place in the PPh/*because* condition, perhaps not as much as in the IPh/*although* condition. If we accept that access to literal form "fades" some time after integration, it might be that with *although*, it does not have time to fade completely, while it would have time to fade with *because*, since *because* involves more integration overall. This explanation works, but it currently remains a stipulation, and the facts on this issue are an open empirical question.

A further issue with the model that deserves more discussion is the claim that duration is the only important aspect of prosody for effects on semantic integration. This idea is based on our results, but as we have pointed out, these were *post hoc* analyses that exploited the natural variability within prosodic categories, not controlled experimental manipulations. It may not be possible to create natural-sounding stimuli with completely independent values for final syllable duration and amount of pitch contour, but it should certainly be possible to create materials with more cross-variation. It could also be the case that our pitch measurements were not an accurate reflection of the information that the auditory processor extracts from pitch contours. It is possible that a more sophisticated phonetic measure of pitch contour could yield better results for pitch as a cue to these effects, which would seriously undermine our purely timing-based account.

Another aspect of our claim about duration is that it is a gradient acoustic phenomenon that cares nothing about the phonological status of a boundary. For example, our model predicts that a PPh boundary with a small pause should show the same effects as an IPh boundary with no pause but an equal amount of final lengthening. A theoretically possible alternative is that duration is simply the best cue for the perception of a categorical boundary in general. If this were true, however, then based on these results, we would expect that pitch would not be as good a cue for the perception of prosodic boundaries in English in general. Our claim here is not about the general status of prosodic boundaries, but just about this specific prosodic effect on semantic processing. We are not claiming that prosody cannot be categorical, but rather that in the effects observed here, the process of semantic integration is blind to those categories and cares only about how much time it has to complete its processing.

Prosody has been found to have a wide variety of effects on processing, including the semantic effects examined here, effects on syntactic disambiguation (e.g. Price et al. 1991; Kjelgaard & Speer 1999), and effects on information structure and reference resolution (e.g. Terken & Nooteboom 1987; Dahan, Tanenhaus, & Chambers 2002). It might be the case that all of these effects are produced by accessing the same prosodic representations, whether that is abstract categories like PPh and IPh or gradient phonetic properties like final lengthening. It could also be the case that different processes access different representations or aspects of prosody. The effect we have investigated here appears to care only about duration. Another effect, such as marking new information, may be more sensitive to pitch and intensity because the information processor might be looking for peaks of prominence. As we find more and more prosodic effects on language processing, the next goal should be to gain a better understanding of *why* prosody has these effects. In describing the prosodic effect on semantic processing here, we have attempted to provide an explanation of why duration in particular is important for this kind of process. We hope that this will illuminate both the process of semantic integration itself, as well as our understanding of how prosody affects language processing.

References

Beckman, M. & Ayers-Elam, G. 1997. *Guidelines for ToBI labeling, Version 3*. Ohio State University, ms.

Bever, T. G. & Townsend, D. J. 1979. Perceptual mechanisms and formal properties of main and subordinate clauses. In *Sentence processing: psycholinguistic studies presented to Merrill Garrett*, eds. W. Cooper & E. Walker, 159–226, Hillsdale, NJ: Erlbaum.

Dahan, D., Tanenhaus, M. K. & Chambers, C. G. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47, 292–314.

Kjelgaard, M. M. & Speer, S. R. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic ambiguity. *Journal of Memory and Language* 40, 153–194.

Nespor, M. & Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris.

Pierrehumbert, J. 1980. *The phonetics and phonology of English intonation*. MIT. PhD diss.

Price, P. J., Ostendorf, S., Shattuck-Hufnagel, S. & Fong, C. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 9, 2956–2970.

Schafer, A. 1997. *Prosodic parsing: the role of prosody in sentence comprehension*. University of Massachusetts Amherst. PhD diss.

Selkirk, E. O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Silverman, K., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C., Price P. & Hirschberg, J. (1992). ToBI: A standard scheme for labeling prosody. In *Proc. Second International Conference on Spoken Language Processing (ICSLP)*, Banff, 867–879.

Terken, J. & Nooteboom, S. G. 1987. Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes* 2, 145–164.

Townsend, D. J. 1983. Thematic processing in sentences and texts. *Cognition* 13, 223–261.

Townsend, D. J. & Bever, T. G. 1978. Interclause relations and clausal processing. *Journal of Verbal Learning and Verbal Behavior* 17, 509–522.

Townsend, D. J. & Bever, T. G. 2001. *Sentence comprehension: the integration of habits and rules*. Cambridge, MA: MIT Press.

Townsend, D. J., Hoover, M. & Bever, T. G. 2000. Word-monitoring tasks interact with levels of representation during speech comprehension. *Journal of Psycholinguistic Research* 29, 265–274.

Watson, D. & Gibson, E. 2002. The relationship between intonational phrasing and syntactic phrasing in language production. University of Rochester, ms.

Watson, D. & Gibson, E. 2005. Intonational phrasing and constituency in language production and comprehension. This volume.

Scott R. Jackson
University of Arizona
Department of Linguistics
P.O. Box 210028
Tucson, AZ 85721
USA
scottj@u.arizona.edu

David J. Townsend
Montclair State University
Department of Psychology – DI-257
Upper Montclair, NJ 07043
USA
townsendd@mail.montclair.edu

Thomas G. Bever
University of Arizona
Department of Linguistics
P.O. Box 210028
Tucson, AZ 85721
USA
tgb@u.arizona.edu