Prosody of corrective *but* sentences in English

Danfeng Wu

[1]University of Oxford
danfeng.wu@magd.ox.ac.uk

Abstract

This paper provides prosodic evidence for a syntactic analysis of corrective *but* sentences, and argues that prosodic structure is not completely flat, but can replicate the dominance relations in syntax. Corrective *but* sentences are *but*-coordination that requires negation in the first conjunct, e.g. (1) *Max misses **not** spinach but chard*; and (2) *Max does**n't** miss spinach but chard*.

There is debate about the syntactic analysis of (1): Toosarvandani (2013) analyzed it as DP-coordination (Max misses [**not** spinach] but [chard]), while Wu (2022) argued that it is structurally ambiguous between DP-, vP- and TP-coordination. In a production study, we showed with duration-based evidence that the prosodic boundary following the first conjunct (i.e. following *spinach*) is stronger than the boundary following a typical DP, supporting Wu's analysis.

Sentence (2) is uncontroversially analyzed as vP-coordination plus ellipsis in the literature, and this vP embeds a DP (Max does **not** [miss [spinach]] but chard). In the second part of the study, we took advantage of this recursive syntactic structure in (2) and asked whether it might lead to recursive prosodic structure. Duration-based evidence suggests that it does, as the prosodic boundary following "spinach" is stronger than the boundary following an unembedded DP.

Index Terms: pre-boundary lengthening, phrasing, recursivity, Strict Layer Hypothesis, corrective *but*, coordination, prosodic evidence for syntax, syntax-prosody mapping, English

Introduction

This paper studies the prosody of corrective *but* sentences. These are sentences coordinated by *but* that require presence of negation in the first conjunct and absence of negation in the second conjunct:

a. Max does**n't** miss spinach but chard. (1)
b. Max misses **not** spinach but chard.
(Based on [1])

Negation cannot be absent or present in both conjuncts:

a. #Max misses spinach but chard. (2)
b. #Max does**n't** miss spinach but **not** chard.

Examples like (1a-b) constitute two different types of corrective *but* sentences: there is debate in the literature about the correct syntactic analysis of (1b), while the analysis for (1a) is less controversial and broadly agreed-on. We will show that prosody can adjudicate between the competing syntactic analyses of (1b). Following the uncontroversial syntactic analysis of (1a), (1a) is a great place to study an important theoretical question about the syntax-prosody mapping– whether the prosodic structure can be recursive. Therefore, the prosody of (1a-b) can address two separate research questions: (a) what the correct syntactic analysis of (1b) is, using prosodic evidence; and (b) whether the prosodic structure can be recursive and replicate the dominance relations in syntax, based on the prosodic study of (1a). The rest of this section will describe these two research questions in detail, and lay out competing theories and their predictions. Section 2 will then present the prosodic experiment that tests both research questions at the same time, and section 3 will conclude.

1.1. Competing syntactic analyses of (1b) and their prosodic predictions

There are two competing syntactic analyses of sentences like (1b). The first is what we call *the strictly-DP-coordination approach*, which analyzes (1b) as DP-coordination based on evidence involving scope, word order and locality effects ((3), [1]).

Strictly-DP-coordination analysis of (1b) (3)
Max misses [$_{DP}$ **not** spinach] but [$_{DP}$ chard]

[2] provided additional evidence involving constituency, scope and antecedent-contained deletion suggesting that (1b) has other parses besides DP-coordination (we call this *the ambiguity approach*). One of these parses involves DP-coordination (4a), but the other parses could involve larger coordination (e.g., vP-coordination, (4b), and TP-coordination (4c)) plus ellipsis, a process that makes some syntactic structure silent (marked by struckthrough text).

Ambiguity analyses of (1b) (4)
a. Max misses [$_{DP}$ **not** spinach] but [$_{DP}$ chard].
b. Max [$_{vP}$ misses **not** spinach] but [$_{vP}$ chard$_i$ ~~misses t_i~~].
c. [$_{TP}$ Max misses **not** spinach] but [$_{TP}$ chard$_i$ ~~he misses t_i~~].

To adjudicate between these two analyses, we can use an empirical generalization about English coordination that has been confirmed experimentally (e.g., [2], [3] and [4]): in coordination, the size of the coordinated constituents is correlated with their prosody. For example, (5a) is coordination of two TPs, while (5b) can involve coordination of two DPs.

(5)
a. [$_{TP}$ Lillian will look for Lauren] or [$_{TP}$ she will look for Bella].
b. Lillian will look for [$_{DP}$ Lauren] or [$_{DP}$ Bella] this Saturday.

This difference in syntactic structure is reflected in their prosody: *Lauren* in (5a) is followed by a stronger prosodic boundary than *Lauren* in (5b). (the strength of a boundary can be detected durationally, as we will see later). Following this empirical observation that the size of coordination affects prosody, the two syntactic approaches make different predictions about the prosody of (1b). The strictly-DP-coordination approach predicts that (1b) should have the prosody of DP-coordination. We can test this prediction by comparing the prosody of (1b) with that of a sentence that is uncontroversially DP-coordination, such as (6). We use the collective predicate *mix* in (6) to make sure it involves DP-coordination. The strictly-DP-coordination expects the prosodic boundary in (1b) to be no different from the boundary in (6) (Figure 1).

Max doesn't mix spinach and chard. (6)

In contrast, the ambiguity approach claims that (1b) can involve vP- and TP-coordination. Suppose that when producing a structurally ambiguous sentence, the speaker chooses any one of the possible parses when saying it. This means that the speaker will sometimes produce (1b) as DP-coordination, sometimes as vP-coordination and other times as TP-coordination. If we can look at many speakers' many productions of (1b), and can take an "average" of their prosodic realizations across these many instances of production, then the ambiguity approach predicts that on average, the prosodic boundary in (1b) should be stronger than that of (6) because of previous findings that coordinated TP has a stronger prosodic boundary than coordinated DP. (Figure 2).

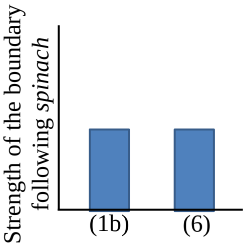

Figure 1: *Prediction of the strictly-DP approach.*

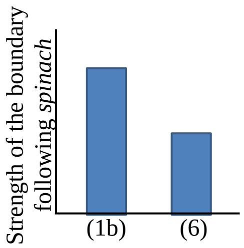

Figure 2: *Prediction of the ambiguity approach.*

1.2. Competing theories of syntax-prosody mapping and their prosodic predictions

Having discussed the competing syntactic analyses of (1b), we now present the uncontroversial syntactic analysis of (1a), and argue that this syntactic analysis can in turn shed light on syntax-prosody mapping. The literature agrees that (1a) should be analyzed as vP-coordination (7) ([1], [2] and [5]).

(7)

The analysis of (1a) according to both approaches
Max does [$_{vP}$ not miss spinach] but [$_{vP}$ chard$_i$ ~~miss t$_i$~~].

It has not been studied before to our knowledge how a syntactic structure like (7) is mapped onto prosody. Specifically, it is not clear how in English, a vP that contains a DP is mapped onto prosody. This is the second research question that this paper wants to address (i.e., what sorts of syntactic phrases are mapped onto prosody).

Different theories on syntax-prosody mapping make different predictions about this question. They fall into two types: one that follows the Strict Layer Hypothesis (e.g. [6], [7] and [8]), where the prosodic structure is flatter than the syntactic structure and does not have nested structure; and the other where the prosodic structure can replicate the dominance relations in the syntactic structure (e.g. [9], [10], [4], [13], [14], [15], [16] and [17]). The first type of theories would neutralize the difference between a vP that contains a DP and a syntactic phrase that doesn't dominate any other phrase, and map the dominating phrase and the non-dominating phrase to prosodic constituents of the same strength. The second type would map the vP that contains a DP to a stronger prosodic constituent than a syntactic phrase that doesn't dominate any other phrase. For concreteness, we discuss an example theory of each type.

Among the theories that respect the Strict Layer Hypothesis, edge-based theory aligns edges of maximal syntactic subclauses (i.e., DP and vP in our case) and clauses (i.e., TP) to edges of prosodic constituents. Assuming that English aligns the right edge of DP and vP to the right edge of a phonological phrase (φ), and following versions of edge-based theory that do not allow recursive prosodic structure (i.e., a φ cannot dominate another φ, e.g. [7]), *spinach* in (7) would be followed by a single φ-boundary (8) because it is at the right edge of a DP and a vP. Thus, these theories derive a prosodic structure that is flatter than the syntactic structure because the two syntactic phrases correspond to a single φ-boundary.

(8)

Prosodic structure of (7) assigned by edge-based theory
Max doesn't miss spinach)$_\varphi$ but chard.

Contrast theories that respect the Strict Layer Hypothesis with those that do allow recursive prosodic structure. For example, Match Theory matches syntactic phrases to φ, and would assign the following prosodic structure to (7), where *spinach* is at the right edge of two φs: one that is mapped from the DP *spinach*, and the other that is mapped from the vP *miss spinach*:

(9)

Prosodic structure of (7) assigned by Match Theory
Max doesn't miss spinach)$_\varphi$)$_\varphi$ but chard.

We cannot directly compare the predictions of these theories (8) and (9) experimentally, but we can test them by comparing the prosody of (7) with that of (6). Both edge-based theory and Match Theory would assign the following structure to (6) because *spinach* is at the right edge of a DP and no other XP.

(10)

Prosodic structure of (6) assigned by edge-based theory and Match Theory
Max doesn't mix spinach)$_\varphi$ and chard.

Edge-based theory predicts that the prosodic boundary following *spinach* is the roughly same for (6) and (7) because *spinach* is at the right edge of a φ in both (Figure 3), while Match Theory puts *spinach* at the right edge of two φs in (7) but only a single φ in (6) (Figure 4).

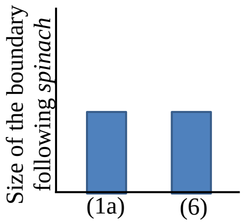

Figure 3: *Prediction of edge-based theory.*

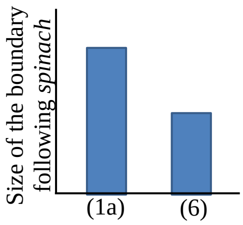

Figure 4: *Prediction of Match Theory.*

While Match Theory allows for recursive φs, to our knowledge no proposal in Match Theory has made explicit predictions about how this recursive structure may lead to gradient phonetic effects that can be detected experimentally such as the degree of lengthening of a segment. Thus, we add the following auxiliary assumption to Match Theory: the more levels a node dominates in the prosodic structure, the phonetically "stronger" this node is. Phonetic "strength" can be reflected by phonetic effects at the left and right edges of this node, such as domain-initial strengthening and domain-final lengthening. By this assumption, a φ must be phonetically stronger than its daughter φ' because the mother φ dominates one more level of φ than the daughter.

2. The experiment

2.1. Materials

The speech materials for the experiment consisted of 8 sets of dialogs in 3 conditions (the two corrective *but* types and the *and*-sentence), exemplified by (1a-b) and (6). Each target sentence was shown to the subjects along with a leading context sentence and an interlocuter, speaker A's utterance, to elicit the intended information structure in the target sentence, speaker B's utterance. For example, the following materials were presented to the speaker to elicit (1a-b) and (6); (1a&b) had the same context and speaker A's utterance.

(11)
Context: Max has been on an all-meat diet, and misses something in particular. They're debating about what Max misses.
A: Max misses spinach.
B1: He misses not spinach but pears.
B2: He doesn't miss spinach but pears.

(12)
Context: Max is particular about his smoothie: he mixes all sorts of ingredients, except a vegetable and a fruit.
A: Which vegetable and which fruit doesn't Max mix?
B: He doesn't mix spinach and pears.

To make sure the difference between the sentences is minimal, we make (6) answer to a double *wh*-question rather than a single *wh*-question, so that all the target sentences have the same focus structure and involve double focus. If the question were a single *wh*-question like *What doesn't Max mix?*, its answer (13B) would put broad focus on the entire conjunction phrase:

A: What doesn't Max mix? (13)
B: Max doesn't mix [spinach and chard]$_F$.

But due to the contrastive nature of the corrective *but* sentences, each conjunct in (1a-b) (i.e., *spinach* and *chard*) is focused separately:

a. Max does**n't** miss [spinach]$_F$ but [chard]$_F$ (14)
b. Max misses **not** [spinach]$_F$ but [chard]$_F$.

Comparing a sentence with broad focus (13B) with ones with double focus (11B1&B2) may create a confound, if focus can affect prosodic boundaries. Therefore, to eliminate this confound and make sure that all the target sentences put focus on each conjunct, we made (6) answer to a double *wh*-question.

The speaker was to read the context silently, and say the dialog in the given order. Every speaker saw all 24 items. There were 100 filler items, which all contained a context, a question and an answer.

2.2. Participants

We conducted a production study with 18 native speakers of North American English (14 female, 4 male, age 19 to 50), who were all university students and working professionals living in Boston, US and Oxford, UK. They were remunerated a small sum for their time, and granted their written consent to being tested.

2.3. Data collection

Recording took place in two events. The first event took place in a sound-attenuated booth at MIT for 3 of the 18 participants, and the second event took place in a quiet, non-reverberant room at the University of Oxford for the other 15 participants. In each event, participants were seated in front of a computer, which displayed one context-question-answer trio at a time. The stimuli plus fillers were presented in pseudo-randomized order, and the order of items was different for every participant. Participants were given instructions about the task at the beginning of the experiment, which asked them to first read each trio quietly to themselves, and only proceed to read it out loud when they were ready. They could take as long as they wanted. They were asked to imagine they were playing three different roles in each trio, and to act out the dialogues naturally rather than reading the sentences mechanically. If the participants were not satisfied with their rendition of an item (a common reason was that they stumbled over some words), they were allowed to say it again. If they asked to repeat an item, we only considered the rendition they were happy with, and discarded the previous renditions.

2.4. Data processing and analysis

The recordings were aligned with the Montreal Forced Aligner [18], using the pretrained acoustic model English (US) ARPA acoustic model [19], and duration was calculated with the forced-aligned boundaries. We measured the duration of the last rime of the word immediately before the prosodic boundary (e.g., for (1a-b) and (6), *ach* of *spinach*). We chose this durational measure because as our auxiliary assumption said, the levels of embedding in a φ are correlated with the phonetic strength of its boundaries. Furthermore, as [18] showed, the phonetic strength of a prosodic boundary can be detected by the degree of lengthening of the final rime before this boundary: the stronger the boundary, the longer the rime. Thus, the duration of the last rime of *spinach* in (1a-b) and (6) is correlated with the strength of the prosodic boundary following *spinach*.

We fitted a linear mixed effects model, with the duration of the last rime as the dependent variable, and item as fixed effects. We calculated p-values using Satterthwaite's degrees of freedom method. The model included random intercepts by speaker and item group, and random slope by speaker.

2.5. Results

The last rime before *but* in corrective *but* sentences with sentence negation (i.e., items like (1a), leftmost box in Figure 5) is 50.6 ms longer than the average duration of the last rime before *and* in *and* sentences (i.e., items like (6), rightmost box in Figure 5; $p < 0.001$). The last rime before *but* in corrective *but* sentences with constituent negation (i.e., items like (1b), middle box in Figure 5) is 52.7 ms longer than the average duration of the last rime before *and* in *and* sentences ($p = 0.044$). Finally, the last rime before *but* in corrective *but* sentences with sentence negation (i.e., items like (1a)) does not differ significantly in duration from that in corrective *but* sentences with constituent negation (i.e., items like (1b)). In Figure 5, the top and bottom of the boxes are the 75th and 25th percentiles, and the middle line is the median. The red dot is the mean, and the red lines are standard error bars.

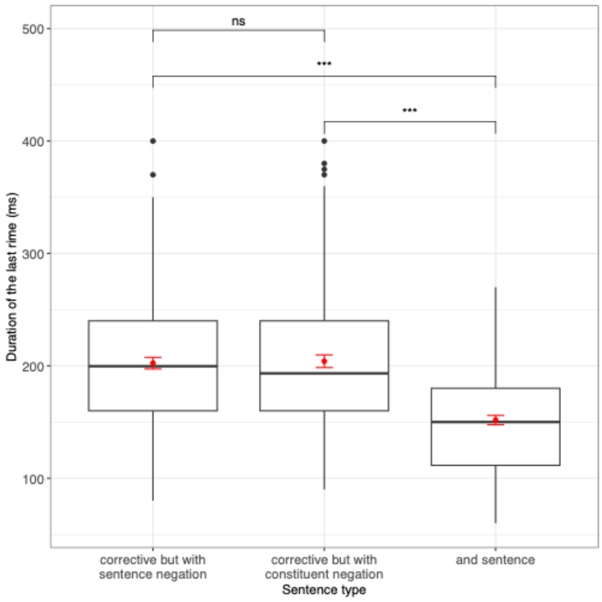

Figure 5: *Duration of the final rime before* but / and.

2.6. Discussion

The durational pattern suggests that the prosodic boundary before *but* does not differ significantly for corrective *but* with sentence negation (e.g., (1a)) and corrective *but* with constituent negation (e.g., (1b)), but those boundaries are stronger than the boundary before *and* (e.g., (6)). This is consistent with the ambiguity approach to (1b) and Match Theory. The fact that the prosodic boundary before *but* in sentences like (1b) is greater than the prosodic boundary before *and* in sentences like (6) supports the ambiguity approach: (1b) is structurally ambiguous, and can not only be analyzed as DP-coordination, but also larger coordination with ellipsis. The fact that a vP that contains a DP (e.g., the vP in (1a)) corresponds to a stronger prosodic phrase than just a DP (e.g., the DP in (6)) suggests that the prosodic structure is not completely flat. One way to implement this is to allow for recursive φs (i.e., a φ can dominate another φ), and boundary strength depends on the number of φ-levels that a φ dominates.

3. Conclusion

This paper presented experiments with consequences for two separate research questions. First, the prosodic realization of corrective *but* sentences supports one syntactic analysis over the other, and suggests that these sentences can involve more underlying syntactic structure than what they appear, and that structure has been obscured by ellipsis. Second, the prosodic realization of some other corrective *but* sentences suggests that at least in English coordination, the prosodic structure tracks the syntactic structure more closely than the Strict Layer Hypothesis claimed, and allows recursive prosodic structure.

These two research questions have broader implications. Evidence for the syntactic structure has traditionally come from sources such as word order and sentence meaning, but evidence from these domains is not always clear. For example, tests based on word order may not be able to detect elided material due to its silent nature. This paper follows a small but growing literature in developing a new syntactic test based on prosody (e.g. [21], [22] and [23]), which will be especially helpful when traditional sources of evidence are not so clear. This paper has also demonstrated that syntactic theory can in turn provide basis for investigating questions about the mapping process between syntax and prosody: due to the recursive syntactic structure of some corrective *but* sentences, they provide a great opportunity to show that the prosodic structure can also be recursive and replicate the dominance relations in the syntax.

4. Acknowledgements

I thank Athulya Aravind, Edward Flemming, Danny Fox, David Pesetsky and Maziar Toosarvandani, participants of NELS 52 and reviewers of Speech Prosody 2024 for helpful comments and feedback.

References

[1] M. Toosarvandani, “Corrective *but* coordinates clauses not always but sometimes,” *Natural Language and Linguistic Theory*, vol. 31, pp. 827–863, 2013.

[2] D. Wu, *Syntax and Prosody of Coordination*, Ph.D. dissertation, MIT, 1995.

[3] M. Wagner, *Prosody and Recursion*, Ph.D. dissertation, MIT, 2005.

[4] M. Wagner, “Prosody and recursion in coordinate structures and beyond,” *Natural Language and Linguistic Theory*, vol. 28, pp. 183–237, 2010.

[5] L. Vicente, “On the syntax of adversative coordination,” *Natural Language and Linguistic Theory*, vol. 28, pp. 381–415, 2010.

[6] M. Nespor and I. Vogel, *Prosodic phonology*. Dordrecht: Foris, 1986.

[7] E. Selkirk, “On derived domains in sentence phonology,” *Phonology*, vol. 3, pp. 371–405, 1986.

[8] J. B. Pierrehumbert and M. Beckman, *Japanese Tone Structure*. Cambridge: MIT Press, 1988.

[9] H. Truckenbrodt, *Phonological Phrases–Their Relation to Syntax, Focus, and Prominance*, Ph.D. dissertation, MIT, 1995.
[10] H. Truckenbrodt, "On the relation between syntactic phrases and phonological phrases," *Linguistic Inquiry*, vol. 30, pp. 219–255, 1999.
[11] E. Selkirk, "On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure," *Gengo Kenkyuu*, vol. 136, pp. 35–73, 2009.
[12] E. Selkirk, "The syntax-phonology interface, 2nd edn," *The Handbook of Phonological Theory*, eds. J. Goldsmith, J. Riggle, and A. Yu, 2011.
[13] E. Elfner, *Syntax-Prosody Interactions in Irish*, Ph.D. dissertation, University of Massachusetts, Amherst, 2012.
[14] E. Elfner, "Recursion in prosodic phrasing: Evidence from Connemara Irish," *Natural Language and Linguistic Theory*, vol. 33, pp. 1169–1208, 2015.
[15] R. Bennett, E. Elfner, and J. McCloskey, "Lightest to the right: An apparently anomalous displacement in Irish," *Linguistic Inquiry*, vol. 47, no. 2, pp. 169–234, 2016.
[16] J. Ito, and A. Mester, "Prosodic subcategories in Japanese," *Lingua*, vol. 124, pp. 20–40, 2013.
[17] J. Ito, and A. Mester, "The perfect prosodic word in Danish," *Nordic Journal of Linguistics*, vol. 38, no. 1, pp. 5–36, 2015.
[18] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced aligner: Trainable text-speech alignment using Kaldi," *Interspeech 2017*, pp. 498–502, 2017.
[19] K. Groman, J. Howell, and M. Wagner, "Prosodylab-aligner: A tool for forced alignment of laboratory speech," *Canadian Acoustics*, vol. 39, pp. 192–193, 2011.
[20] C. W.|Wightman, S. Shattuck-Hufnagel, M.|Ostendorf, and P. J.|Price, "Segmental durations in the vicinity of prosodic boundaries," *Journal of the American Acoustic Association*, vol. 91, pp. 1707–1717, 1992.
[21] J. W.|Bresnan, "Sentence stress and syntactic transformations," *Language*, vol. 47, pp. 257–281, 1971.
[22] L. Clemens and J.|Coon, "Prosodic constituency of verb-initial clauses in Ch'ol," *Proceedings of the workshop on structure and constituency in the languages of the Americas [WSCLA] 20*, vol. 43, pp. 88–100, 2018.
[23] L. Clemens, "Prosodic noun incorporation: The relationship between prosody and argument structure in Niuean," *Syntax*, vol. 22, pp. 337–377, 2019.