Consistency and Variability in Prosody-Syntax Correspondences in English (E)
Amy J. Schafer
University of Hawaii
aschafer@hawaii.edu

1. Prosody-syntax correspondences have long been noted in linguistic analysis (e.g., Selkirk 1984, Cooper & Paccia-Cooper 1980). (Parentheses indicate prosodic phrasing.)

   a. (Jane) (gave a book) (to Mary)
   b. * (Jane gave) (a book to Mary)
   c. (When John leaves Kathy) (we'll be upset)
   d. (When John leaves) (Kathy will be upset)

2. Prosodic structure of Mainstream American English (e.g., Beckman & Pierrehumbert 1986).

   Intonation phrase
   Marked by an edge tone (L%, H%), final lengthening, and other cues.
   |
   Intermediate phrase
   Marked by an edge tone (L-, H-), final lengthening, and other cues.
   |
   Phonological word
   Carry pitch accents (e.g., H*, L+H*, !H*) associated with stressed syllables.

3. What is the relationship between prosody and syntax in production?

   a. Syntax directly determines the phonetic output? (An early view.)

   b. Syntax alone determines a prosodic structure (and the prosodic structure determines the phonetic output)? (A view seemingly assumed by some psycholinguists.)

   c. Speakers produce prosody that indicates the syntax only when they think their listeners need such cues? (A view currently advocated; see below.)

   d. Many linguistic factors determine a prosodic structure (syntax, contrastive focus, given vs. new information, length of constituents, etc.)?

   I assume (a), (b), and (c) are false and (d) is true. With any approach, we need an exact mechanism that correctly predicts the prosodic form for a broad range of linguistic structures and speech situations. To determine this mechanism, we need rich data sources – of many types.

4. How is prosodic information used in comprehension?

   a. A Gricean approach, in which the actual prosody is compared to other prosodies that could have been produced?

   b. Automatic, non-comparative processing, in which the actual prosody simply activates appropriate linguistic structures?

With either approach, we need to know exactly how the details of prosodic form are interpreted. This is especially difficult if the relationship between prosodic form and linguistic form is complex (i.e., if prosody is determined by multiple interacting linguistic factors). Again, we need converging evidence from many sources.

5. Much linguistic research has argued for regular relationships between prosodic form (including intonational form) and syntactic form. However, this research has often relied on data from intuitions or small collections of recorded speech, and the speakers in such studies have often been linguistically trained or professional speakers.

6. Some recent psycholinguistic results have suggested that prosody does not reflect syntactic structure in the speech of "naïve speakers" (speakers without training in linguistics or professional speaking) if the meaning of the utterance is clear in context (if it is "situationally disambiguated"). This research argues that naïve speakers differ from trained speakers in the prosody they produce, and that naïve speakers produce disambiguating prosody only if there is a situational need to disambiguate (e.g., Snedeker & Trueswell 2003, Allbritton et al 1996).

    a. *Tap the frog with the flower* (S&T 2003)
        i. Significant prosodic cues were produced in a situation that supported both interpretations (i.e., a situation with a large flower and two frogs, one of which was wearing a small flower).
        ii. Fewer prosodic cues were produced in situation compatible with just one interpretation (i.e., a situation with one frog and a separate flower).

    b. *When you learn gradually you worry more.*
        i. Clause boundary before or after *gradually.*
        ii. Expert speakers reliably produced prosodic cues (Price et al 1991).
        iii. Naïve speakers did not (Allbritton et al 1996).

7. Our approach: A game-based conversation task using naïve speakers

    a. Pairs of speakers, naïve to the purpose of the experiment, played a game. Players were not told to produce speech clearly, carefully, or in a way that expressed the meaning. They were simply instructed to play a board game while their speech was recorded.

    b. It was a cooperative, interactive task with non-linguistic goals.

    c. Each speaker possessed distinct knowledge that needed to be shared with the other speaker (locations of goals vs. locations of bonuses and hazards).

    d. Speakers used scripted sentence frames plus object names to negotiate game pieces from start positions to goal positions. (E.g., *I want to change the position of the _____.*)

    e. The speech was recorded and detailed acoustic and phonological analyses were performed (duration, F0, ToBI transcription (e.g., Beckman & Ayers, 1997).

    f. The recordings were then played back to new sets of naïve listeners, who either:
        i. Heard complete sentences and choose between two paraphrases, or
        ii. Heard the lexically ambiguous portion of the sentences and chose between two disambiguating continuations.

8. Example of a dialogue in the game

    a. "Driver": I want to change the position of the square with the triangle.
    b. "Slider": Which triangle do you want to change the position of the square?
    c. "Driver": The red one. When that moves the square it should land in a good spot.
    d. "Slider": Good choice. When that moves the square will encounter a cookie.

9. Examples of sentences in our study

    a. High vs. low PP attachment
        i.   *High*: I want to change the position of [the square]$_{NP}$ [with the triangle]$_{PP}$
        ii. *Low*:   I want to change the position of [the square with the triangle]$_{NP}$
        iii. *High*: I am able to confirm the move of [the square]$_{NP}$ [with the triangle]$_{PP}$
        iv. *Low*:   I am able to confirm the move of [the square with the triangle]$_{NP}$

    b. Early closure vs. late closure of a subordinate clause (intransitive vs. transitive analysis of the verb of the subordinate clause)
        i.   *Early*: When that moves, the square will encounter a cookie.
        ii. *Late*:   When that moves the square, it should land in a good spot.

    c. NP- vs. S-complement (not discussed in this talk)
        i.   *NP-complement*: I am able to confirm [the move of the square with the triangle]$_{NP}$
        ii. *S-complement*:   I am able to confirm [the move was the final one]$_S$

    d. Early vs. late gap position (not discussed in this talk)
        i.   *Early*: Which triangle$_i$ do you want e$_i$ to change the position of the square?
        ii. *Late*:   Which triangle$_i$ do you want to change the position of e$_i$ this time?

10. Our PP sentences are 12 or 13 words long (see 9a). Is this longer than typical?
    We analyzed spontaneous speech with V-NP-PP sequences from the Switchboard corpus (Godfrey & Holliman, 1993; recordings and transcripts of spontaneous telephone conversations) to determine the typical length of such constructions.

    a. The mean length (total number of words / number of utterances) was 21.8 words.
    b. The median length (half of the utterances were longer, half were shorter) was 9 words.
    c. The mode length (most frequent length) was 17 words.
    d. Most of the utterances were between 7 and 19 words in length.

11. Examples of spontaneous speech with V-NP-PP sequences from the Switchboard corpus.

    a. Those things must take up a huge amount of space in landfills.
    b. It never hurts to have some kind of a grounding in law.
    c. But I think you you *[sic]* really hit the nail on the head.
    d. We was teasing about you know splitting our house down the middle.
    e. But I feel very good about spending those years with my boys.
    f. I hope I'll be able to do the same with grandchildren.
    g. You sound like you've got some small ones in the background.
    h. I heard it's a law that you're only allowed to have three layers of shingles.
    i. Just to you know have a moment of peace and quiet without little ones at the table.

j. And then I just have the rest of my house with lot of hand painting in it.
k. So I went to a dealership because they had the same year and same model of Prelude on their used car lot.
l. And this man was accused of killing I think it was two people and shooting a policeman point blank in the face.

12. Factors in the PP experiments

a. Syntax: High vs. low PP attachment
Will the prosodic boundary before the PP be stronger for high attachment sentences?

b. Speaker role: Driver (gave directions) vs. Slider (confirmed moves)
Will Drivers, who direct the moves, disambiguate more than Sliders?

c. Gamepiece contrast: Ambiguous Triangle sentences vs. situationally disambiguated Cylinder sentences. (Triangle sentences: combined square-and-triangle piece for Low Attachment vs. separate pieces for High Attachment, in which the triangle is an instrument that moves the square. Cylinder sentences: no combined square-and-cylinder piece).
Will disambiguation be greater for Triangle sentences than for Cylinder sentences?

d. Gameboard configuration contrast: Ambiguous, Biased, or Unambiguous configurations of game pieces on the boards (e.g., only the combined square-and-triangle piece can move in the Unambiguous Low Attachment configuration).
Will disambiguation be greater for Ambiguous gameboard configurations?

e. Sequence in game: First through last position in the game (played for 2 hours).
Will disambiguation increase as the game continues (and speakers potentially become more aware of the ambiguity)?

f. Game ambiguity: Ambiguous Game (game with Triangle ambiguity) vs. Unambiguous Game (game in which all PP sentences were situationally disambiguated).
Will disambiguation be greater in the Ambiguous game?

13. Results of PP experiments

a. Significant effect of syntax: Stronger prosodic boundaries before the PP in High Attached PP sentences than in Low Attached PP sentences. (Speakers generally disambiguate.)

b. Significant effect of speaker role: Longer durations for Driver than Slider *and* Significant effect of syntax within each role: Sliders disambiguate as much as Drivers.

c. As much disambiguation for Cylinder sentences as for Triangle sentences.

d. As much disambiguation for Unambiguous configurations as for Biased configurations and for Ambiguous configurations.

e. General decrease in duration as the game continued (increase in speech rate) *and* Decrease in duration for sentences that repeated the syntactic structure of the previous utterance *but* No clear evidence of increased disambiguation because of greater awareness of ambiguity.

f. Significant disambiguation both in the Ambiguous Game and in the Unambiguous Game.

14. Closure experiments

    a. Syntactic contrast between Early Closure (intransitive analysis) and Late Closure (transitive analysis).

    b. No *situational* need to disambiguate with prosody. Prosodic cues to syntactic structure would be redundant, as the correct syntactic structure could be determined by:
        i. The previous utterance in the discourse.
        ii. Whether the Driver or Slider uttered the sentence.
        iii. The syntactic information following the ambiguous region of "moves the square" (the next word – either "it" or "will" – resolves the ambiguity).

    c. Results: Significant disambiguation of syntax with prosody, and *stronger* use of prosody for the Closure sentence than for the PP sentences.

15. Consistency and variability

    a. Within the PPs, there was a strong effect of syntax on prosodic form, yet not all utterances were produced with the same prosody or even with disambiguating prosody.
        i. For example, only 57% of the High Attached PP utterances were spoken with the strongest prosodic break in the utterance immediately preceding the PP. (In 80% of the utterances the boundary at the PP was one of the strongest boundaries (e.g., multiple intonation phrase boundaries, one of which was at the PP).)

        ii. Although comprehension was significantly better than chance, the listeners achieved only 76% correct choices for High Attachments and 64% correct choices for Low Attachments.

        iii. In a tonal analysis of 79 High Attached utterances produced by 13 speakers, we found 63 different tonal patterns on the sequence "the position of the square". For 101 Low Attached utterances, we found 87 distinct tonal patterns on the phrase.

        iv. Sliders disambiguated as much as Drivers, yet Slider utterances were shorter in duration throughout the critical region than Driver utterances.

        v. Duration decreased as the game continued (indirectly affecting disambiguation).

        vi. Duration decreased in adjacent repetitions (indirectly affecting disambiguation).

    b. Within Closure utterances, disambiguation was again less than 100% but strongly constrained by the syntactic structure.
        i. When transcribers were given full sentences to transcribe, approximately 93% were transcribed with the strongest prosodic boundary at the clause boundary. When transcribers heard only the ambiguous fragments, approximately 77% were analyzed as having the strongest prosodic boundary at the clause boundary.

        ii. Listeners achieved only 76% correct choices in the comprehension test for the Early Closure sentences and 72% correct choices for Late Closure sentences.

        iii. We found 25 distinct tonal patterns for the sequence "moves the square" in 35 Early Closure utterances produced by 13 speakers, and 22 distinct tonal patterns on this phrase for 48 Late Closure utterances.

16. Conclusions

   a. There was no clear support for situationally dependent prosodic form. We argue elsewhere that previous research failed to find evidence of prosodic reflections of syntax in disambiguated situations because of methodological limitations (e.g., Schafer et al, forthcoming).

   b. The results generally support previous linguistic results, by showing strong effects of syntax within each construction. This highlights the value of using intuitions and small data sets in linguistic analysis, especially in preliminary work.

   c. The results show stronger prosodic reflections of syntactic structure in the Closure sentences than in the PP sentences. This further supports the role of grammatical factors in prosodic form and is additional evidence against prosodic form being determined by a situational need to disambiguate.

   d. The results also show substantial variability, both within and across speakers. Some of the variability appears to be due to optionality in the grammar; some appears to be due to performance factors. Variability is particularly apparent in the tonal analyses, but also appears in the analyses of prosodic phrasing. The amount of variability we found highlights the value of larger data sets and laboratory-based methods in filling out the details of linguistic models and in assessing the interaction of two or more factors. For example, we can determine which situations/structures allow more than one prosodic form, and calculate the relative frequencies of each option.

References
Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 714-735.
Beckman, M. E., & Ayers, G. (1997). *Guidelines for ToBI labelling.*Unpublished manuscript, Columbus, OH: Ohio State University.
Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3,* 255-309.
Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.
Godfrey, J. J. & Holliman, E. (1993). Switchboard-1. University of Pennsylvania: Linguistic Data Consortium.
Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956-2970.
Schafer, A. J., Speer, S. R., & Warren, P. (forthcoming). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to Studying World-Situated Language Use: Bridging the Action and Product Traditions*. Cambridge, MA: MIT Press.
Selkirk, E.O. (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
Snedeker, J. & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48,* 103-130.