Default Phrasing in English Relative Clause Attachment Data

Sun-Ah Jun and Molly Shilman

Department of Linguistics
University of California, Los Angeles
`{jun; mshilman}@humnet.ucla.edu`

Abstract

This study investigates the validity of the Implicit Prosody Hypothesis (IPH) by examining the default phrasing of English sentences where a complex noun phrase is the head of a relative clause (RC) (e.g., *the servant of the actress who was on the balcony*). The prosodic phrasing data collected from 36 speakers were transcribed by three labelers. Results show that, counter to the prediction of the IPH, the most common prosodic phrasing was (NP1 NP2) // (RC), the same pattern found in Japanese and Korean, which are known to be high attachment preference languages. The results also showed that the default phrasing was influenced by the length of the RC but not the bias of the RC. The results are discussed in relation to factors affecting default phrasing.

1. Introduction

This study investigates the validity of the Implicit Prosody Hypothesis (IPH) through a production experiment on English sentences. The IPH was proposed by Fodor [8, 9] to explain the difference in attachment preferences found across languages. Speakers of some languages, such as English and Arabic, prefer low attachment (e.g., the relative clause (RC) *who was on the balcony* modifies *the actress* in *Someone shot the servant of the actress who was on the balcony*) while speakers of other languages, such as Spanish, German, and Japanese, prefer high attachment (i.e., the RC modifies *the servant* in the preceding example). [See [7] for a review.]

Fodor claims that this cross-linguistic difference in attachment preference is due to a difference in prosody, especially the default prosodic phrasing of a sentence. The IPH states that a default prosodic contour projected in silent reading favors the syntactic analysis associated with the default prosody for the construction. Since languages differ in their prosody, attachment preferences would differ across languages. This hypothesis was motivated by the effect of RC length on attachment resolution. In a sentence processing experiment based on off-line reading data, an RC tends to be attached low when it is short but high when it is long. Maynell [17] and Lovric et al. [16] claimed that speakers interpret a prosodic break before an RC (in a sequence of NP1 NP2 RC) as a marker of a stronger syntactic boundary, which prompts high attachment. This suggests that speakers of a language with high attachment preference in silent reading tasks would tend to produce a specific pattern of breaks: a big prosodic boundary between an RC and an adjacent head noun in the default phrasing of a sentence, but a weak boundary between the two head nouns (NP1 and NP2). Since the perception of boundary strength at one point is relative to the strength of the preceding boundary [3], we can hypothesize that a prosodic break between an RC and the adjacent head noun would be *bigger* than that between the two head nouns for speakers of high attachment preference languages while the opposite relation would be true for speakers of low attachment preference languages.

This hypothesis was tested using Japanese and Korean data. Assuming that out-of-the-blue reading is close to the default prosody projected in silent reading, Jun & Koike [14] examined the prosodic phrasing of 48 ambiguous target sentences in Japanese produced by 30 speakers of Tokyo Japanese. Speakers read the target sentences 'out-of-the-blue', and three weeks later, participated in an off-line sentence processing experiment. The authors found that Japanese speakers prefer high attachment in the processing experiment, confirming previous results [15], and that speakers also produced a bigger prosodic break between an RC and the following head noun than that between the two head nouns (i.e., (RC) // (NP1 NP2), called an early break), supporting the IPH. They also found an effect of RC length in both production and processing experiments: Longer RCs showed more early breaks than shorter RCs and longer RCs were attached high more often than short RCs. However, the length of NP1 and NP2 (short vs. long NP1; short vs. medium NP2) had no effect on attachment or phrasing.

A similar experiment was performed on Korean data (32 target sentences, 30 speakers) by Jun & Kim [12, 13]. They performed an off-line processing experiment as well as a production experiment. They found that Korean speakers also preferred high attachment, and the early break pattern was predominant in their production, supporting the IPH. Furthermore, long RCs showed more early breaks and more high attachment than short RCs.

In this paper, we examine the default phrasing of English sentences containing the target structure NP1 NP2 RC. Since English is a low attachment preference language, we expect to have a smaller prosodic break between the RC and the preceding head noun than the break between the two head nouns. In other words, the break after NP2 would be smaller than that after NP1, represented in this paper as "NP1 > NP2".

Bergmann & Ito [2] tested a similar hypothesis by examining the default phrasing of English sentences read aloud (24 target sentences, 16 speakers). They hypothesized that the length of NP1 would affect the break at NP2 but the length of RC would not. This is because, at NP2, the length of NP1 and NP2 is known but the length of the RC is not. They cited eye-tracking literature [18, 19] indicating that parafoveal preview does not facilitate resolving attachment ambiguity. To test this, they varied the length of NP1 (one word vs. two words) and RC (verb only vs. verb-complement). They found that subjects preferred low attachment in judgments but that the boundary after NP2 was bigger than that after NP1 in production, contra the IPH. They also found that the boundary at NP2 was not affected by the length of either NP1 or RC.

In our study, we did not vary the length of NP1 or NP2 because in [14] the effect of NP length was not significant. We did vary RC length, which did show significant effects in [14]. In addition to RC length, we varied the bias of RC (i.e., RC is

biased to modify NP1 or NP2 or is unbiased) by manipulating either gender or number agreement. This is to see whether default phrasing is sensitive to meaning as well as RC length.

2. Methods

2.1. Subjects

A total of 36 speakers of American English, mostly undergraduate students at UCLA, participated in the experiment. Twelve subjects each were randomly assigned to three groups (Group 1, 2, 3). The script differed by group, so each subject read only one bias-version of each sentence.

2.2. Material

Twenty seven target sentences (9 short RC sentences, 9 medium RC sentences, and 9 long RC sentences) were chosen from the literature: 14 from Fernandez [7], 7 from Carreiras & Clifton [4], 2 from Frazier [10], and 4 from Deevy [5, 6]. All the target sentences have a structure of Subject-Verb-(NP1 *of the* NP2-RC). Sentences with a short RC had 10-13 syllables before the RC and a 3-5 syllable RC. Sentences with a medium RC had 10-14 syllables before the RC and a 7-10 syllable RC. Sentences with a long RC had 13-15 syllables before the RC and a 13-18 syllable RC.

The target sentences were modified so each sentence type had three versions of bias: No-bias, NP1-bias, and NP2-bias. The bias was controlled by varying number agreement (19 sentences) or gender agreement (8 sentences). Examples of the three RC bias types by number agreement are given in (1).

(1) RC bias types by number agreement (data from [7])
    a. No-bias:
       My friend met the aide of the detective that was fired.
    b. NP1-bias:
       My friend met the aide of the detectives that was fired.
    c. NP2-bias:
       My friend met the aides of the detective that was fired.

Three scripts were made, each one containing one bias type of each target sentence. Each script includes 27 target sentences: 9 per RC length type and 9 per bias type. Therefore, there are 3 no-bias sentences, 3 NP1-bias sentences and 3 NP2-bias sentences in each length category. The target sentences were randomized with 30 filler sentences. The length of the fillers was similar to that of the target sentences: 10 short sentences (12-18 syllables), 10 medium sentences (20-24 syllables), and 10 long sentences (27-35 syllables). The same fillers were used for all scripts.

2.3. Procedures

Subjects read each sentence aloud twice in the sound booth at the UCLA Phonetics Lab. The recordings were digitized at 22 kHz. The first repetition was segmented for transcription unless disfluent; when the 1st repetition was disfluent (about 15% of the time), the 2nd repetition was used. Three ToBI-trained labelers independently transcribed the tones and break indices of each utterance starting at NP1, following the MAE_ToBI conventions for tones [1] and a modified ToBI convention for break indices. For tones, the presence of pitch accent and type of phrase accent were transcribed on NP1, NP2 and the first prominent word in RC, and break indices were labeled (1, 3, 4, 1-, 3-, 4-) after NP1 and NP2 following

the MAE_ToBI conventions. For the mismatch cases between tones and break indices, however, a new labeling was used: '1m' was used when the degree of juncture matched that of a phrase-medial word boundary but a phrase accent was present, and '3m' was used when the degree of juncture matched that of an Intermediate Phrase boundary but there was no phrase accent. When assessing relative break strength, '1m' was treated as '1' and '3m' was treated as '3'.

Labeler 1 transcribed all three types of RC bias from 24 speakers (27 sentences * 8 speakers * 3 groups) = 648 sentences). Labeler 2 transcribed all no-bias items from all 36 speakers (9 sentences * 36 speakers = 324 sentences) plus one third of NP1-bias and NP2-bias items from all 36 speakers (3 NP1-bias sentences and 3 NP2-bias sentences * 36 speakers = 216 sentences). Labeler 3 transcribed all no-bias items from all 36 speakers (9 sentences * 36 speakers = 324 sentences). Thus, 27 no-bias type sentences from 24 speakers were transcribed by all three labelers.

3. Results

There was no Group difference in the phrasing, so results are reported by combining data across Groups. Tonal transcription data did not differ between NP1 and NP2 or by RC length/bias type, so only break index (BI) data are presented. The inter-transcriber agreement rate on raw BI data (based on the 27 no-bias items from 24 speakers transcribed by all three labelers) was 53-66% using a stringent method (agreeing transcriber-word-pairs per total transcriber-word-pairs in percent format). When the break index was simplified into two categories, a break word-size or smaller (BI<=1) vs. a break bigger than a word (BI>1), the agreement rate increased to 72-88%. Specifically, the agreement rate was 79.5% and 87.8% for the BI after NP1 and NP2, respectively, and 71.6% for the relative break strength between NP1 and NP2. In the following subsections, results are based on the simplified BI data.[1]

3.1. Break Index after NP1 and NP2

Since English speakers prefer low attachment, we predicted that the break index after NP1 would be greater than that after NP2. The results show that, overall, when the RC is not biased to attach to either NP (no-bias condition), speakers produced a bigger boundary after NP2 than after NP1, counter to the prediction of the IPH.

Figure 1 shows the proportions of BI<=1 and BI>1 after each NP in the no-bias condition for each labeler. Labeler 1's data is from 24 speakers; the data from Labelers 2 and 3 is from 36 speakers. As shown in Figure 1, about half of the breaks after NP1 are larger than word-size, and about 90% of the breaks after NP2 are larger than word-size. That is, the end of NP2 was more often marked by a bigger break than the end of NP1, for all labelers. The difference between the NP1 break pattern and the NP2 break pattern is significant for all labelers: Labeler 1 (χ^2=72.07, df=1, p<.001), Labeler 2 (χ^2=55.23, df=1, p<.001), Labeler 3 (χ^2=83.16, df=1, p<.001). There was no significant inter-transcriber difference when comparing data from the 24 speakers transcribed by all three.

---

[1] This simplification was introduced because the great majority of break indices were either 1 or 3; the small number of breaks transcribed with some labels made the Chi-square test inappropriate for the data in its original form.
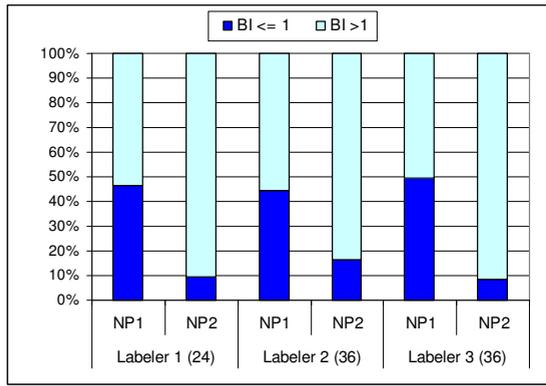
Figure 1: *Frequency (in %) of break index, BI<=1 or BI>1, after each NP in the no-bias condition, for each labeler.*

The relative strength of the NP1 break and the NP2 break within an utterance confirms this. To make this comparison, we divided utterances into three categories of relative break strength: NP1>NP2, NP1<NP2 and NP1=NP2. As shown in Figure 2, the NP1<NP2 type is most common (>50%) and the NP1>NP2 type is least common (<=20%), which is the opposite of the pattern predicted by the IPH. The observed frequency of the three relative BI types was significantly different from the expected frequency for all labelers: Labeler 1 (χ^2=72.07, df=1, p < 0.001), Labeler 2 (χ^2=42.92, df=1, p < 0.001), Labeler 3 (χ^2=91.76, df=1, p < 0.001).

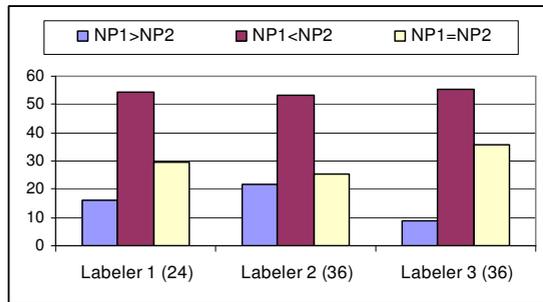Figure 2: *Frequency (in %) of the three relative break index types, NP1>NP2, NP1<NP2, NP1=NP2, in the no-bias condition, for each labeler.*

3.2. The Effect of RC Length on Default Phrasing

The current study differs from Bergmann & Ito [2] in finding a significant effect of RC length on prosodic phrasing for all labelers. Figure 3 shows the proportions of the three relative break index types for each RC length, for all labelers. All data are from the no-bias condition. Here, longer RCs show more NP1<NP2 pattern than shorter RCs, while shorter RCs show more NP1>NP2 pattern than longer RCs. For Labeler 1, long RC items (l) were significantly different from short RC items (s): χ^2=11.05, df=2, p=.004. For Labeler 2, short RC was significantly different from medium RC (m: χ^2=9.77, df=2, p=.008) and from long RC (χ^2=13.94, df=2, p<.001). For Labeler 3, long RC was marginally different from medium RC (χ^2=5.77, df=2, p=.056) and significantly different from short RC (χ^2=7.12, df=2, p=.029).
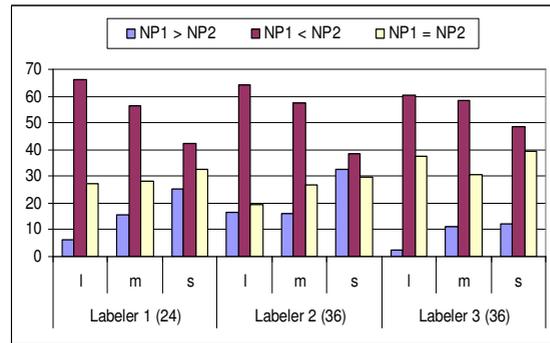

Figure 3: *Frequency (in %) of the three relative break index types, NP1>NP2, NP1<NP2, NP1=NP2, for each RC length (in the no-bias condition only) for each labeler. X-axis labels: l = long RC, m = medium RC, and s = short RC.*

3.3. The Effect of RC Bias on Default Phrasing

We expected that NP2-biased sentences would show a smaller break after NP2 than after NP1, (NP1) // (NP2 RC), because NP2 and RC would form one meaning group. On the other hand, NP1-biased sentences were expected to show a bigger break after NP2 than after NP1, (NP1 NP2) // (RC), so that the RC would not modify NP2. However, the phrasing data did not show any significant effect of RC bias on default phrasing, for either of the two labelers. As shown in Figure 4, speakers produced a similar pattern of default phrasing regardless of the RC bias condition. The frequency of each phrasing pattern across the three RC bias conditions did not differ significantly for either labeler. (Labeler 1: χ^2=5.04, df=4, p = 0.28; Labeler 2: χ^2=1.81, df=4, p = 0.77.) This was true whether the RC bias was due to number agreement or to gender agreement.
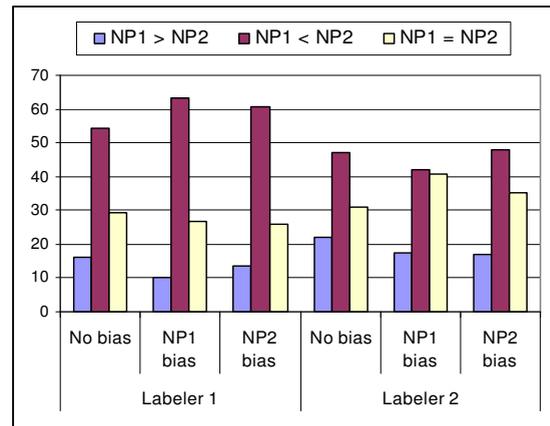

Figure 4: *Frequency (in %) of the three relative break index types, NP1>NP2, NP1<NP2, NP1=NP2, in the three RC bias conditions for Labelers 1 and 2.*

4. Discussion and Conclusion

Earlier studies showed that the default phrasing of Japanese and Korean sentences containing a structure of RC NP1 NP2 supported the IPH. Speakers of these languages, found to prefer high attachment of RC, showed a bigger break after RC

than after NP1. However, in the current study, we have found that the default phrasing of English data did not support the IPH, at least when the default prosody is assessed from out-of-the-blue reading. Assessing how similar this reading prosody is to the implicit prosody requires further research.

Our result showed that the default phrasing was influenced by RC length but not by RC bias, probably because the attachment is disambiguated by morpho-semantics. This is consistent with Hirschberg & Avesani's [11] finding that ambiguities of PP or RC attachment already disambiguated by context were not consistently disambiguated by phrasing across speakers. Taken together, these results suggest that English speakers may not mark attachment differences via prosodic phrasing when other disambiguating information is available. However, as [11] found that speakers did use prosody to disambiguate scope of negation and the domain of focus sensitive operators such as *only*, the effect of disambiguating context may be structure/meaning dependent.

The fact that the string NP1 NP2 RC was predominantly phrased as (NP1 NP2)//(RC) regardless of length or bias suggests that there may be a strong rhythmic or phonological weight constraint on the default phrasing of this structure— one so strong that any difference associated with attachment preferences is overridden. That is, the sequence of two head nouns contains enough syllables to form its own phrase, separate from the RC. This implies that any language with the NP1 NP2 RC (or RC NP1 NP2) structure will have the same pattern of default phrasing: a bigger break between the RC and NP1 than between the two head nouns. If this is the case, the IPH cannot account for the cross-linguistic difference in attachment preferences in NP1 NP2 RC structures. To fully understand the implications of the English results, we need to test more languages with low attachment preference.

The pattern of default phrasing found in our study matches that in Bergmann & Ito's study [2]. However, we found an effect of RC length where they did not. One possible reason for this difference is the length of RC used. In their study, a short RC contained a verb (e.g., *who smokes*), while a long RC contained a verb plus a complement (e.g., *who smokes like a chimney*). These examples show that their long RC is slightly longer than our short RC and shorter than our medium RC. As shown in Figure 3, short and medium RCs differ significantly in only one labeler's data (Labeler 2). Therefore, the difference in RC length in their study may not have been big enough to show an effect.

Another difference between the two studies is the location of the target structure: NP1 NP2 RC was located at the beginning of a sentence in Bergmann & Ito but at the end in our study. In Japanese data [14], where the target structure's location was varied (sentence-initial or –medial), the length effect was significant only when the target structure was in the middle of a sentence. The length of the RC in the Japanese data was comparable to ours. This suggests that default phrasing is influenced by the location of the structure as well as its length.

In sum, we need to investigate what factors affect the default phrasing, and to fully assess the validity of the IPH, we need to examine the prosody of more languages and various constructions.

5. References

[1] Beckman, M.; Ayers-Elam, G. 1997. Guidelines for ToBI Labelling. Unpublished ms. Ohio State Univ. Version 3.

[2] Bergmann, A.; Ito, K. 2007. Attachment of Ambiguous RCs: A Production Study. Talk given at the 13th Annual Conference on Architectures and Mechanisms for Language Processing. Turku, inland. Aug.24-27, 2007.

[3] Carlson, K.; Clifton, C.; Frazier, L. 2001. Prosodic boundaries in adjunct attachment. *Journal of Memory and Languages* 45, 58-81.

[4] Carreiras, M.; Clifton, C., 1993. Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36, 353-372.

[5] Deevy, P. L. 1999. The Comprehension of English Subject-Verb Agreement. Appendix L: Word by Word Self-Paced Reading Study materials: Agreement in Subject Relative Clauses

[6] Deevy, P. L. 2000. Agreement checking in comprehension: Evidence from relative clauses. *Journal of Psycholinguistic Research* 29 (1), 69-79.

[7] Fernandez, E. (2002) *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Ph.D. diss. CUNY.

[8] Fodor, J. D. 1998. Learning to Parse. *Journal of Psycholinguistic Research* 27(2): 285-319.

[9] Fodor, J. D. 2002. Prosodic Disambiguation in Silent Reading. *NELS* 32: 113-32.

[10] Frazier, L., 1990. Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In *Comprehension Processes in Reading,* D. A. Balota, G. G. Flores d'Arcais & K. Rayner (eds.). Hillsdale, NJ: Lawrence Erlbaum, 303-330.

[11] Hirschberg, J.; Avesani, C. 2000. Prosodic disambiguation in English and Italian. In *Intonation: Analysis, Modelling, and Technology*. A. Botinis (ed.). Kluwer Academic Publishers. Pp. 87-96.

[12] Jun, S.-A.; Kim, S. 2004a. Default phrasing and attachment preferences in Korean. *Proceedings of INTERSPEECH-ICSLP*, Jeju, Korea.

[13] Jun, S.-A.; Kim, S. 2004b. The Intermediate Phrase in Korean intonation: Evidence from Sentence Processing. A talk given at the first International Conference on Tone and Intonation in Europe. A paper to appear in *Tones and Tunes: Studies in Word and Sentence Prosody*, C. Gussenhoven and T. Riad. (eds.). Mouton de Gruyter.

[14] Jun, S.-A.; Koike, C. 2003. Default Prosody and RC Attachment in Japanese, a talk given at the 13th Japanese-Korean Linguistics Conference. Proceedings from CSLI, Stanford, in press.

[15] Kamide, Y.; Mitchell, D. C. 1997. Relative Clause Attachment: Non-determinism in Japanese Parsing. *Journal of Psycholinguistic Research* 26: 247-54.

[16] Lovric, N.; Bradley, D.; Fodor, J. D. 2001. Silent Prosody Resolves Syntactic Ambiguities: Evidence from Croatian. Presented at the SUNY/CUNY/NYU Conference. N.Y.

[17] Maynell, L. A. 1999. Effect of Pitch Accent Placement on Resolving Relative Clause Ambiguity in English. Poster at the 12th CUNY Conference, New York.

[18] Rayner, K. & Liversedge, S.P. (2004). Visual and Linguistic Processing During Eye Fixations in Reading. In F. Ferrreira & J. Henderson (Eds). *The Interface of Language, Vision and Action: Eye Movements and the Visual World,* 59-104, Psychology Press.

[19] White, S.J., Rayner, K., & Liversedge, S.P. (2005). The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic Bulletin & Review, 12,* 466-471.