Jemmin Chang 12/08/16
LING 1580
Prof. Warren

Term Paper: The Implicit Prosody Hypothesis

Introspecting on the curious human process of interpreting written language, many have noted a perception or feeling somehow intuitively analogous to speech production, an “inner voice” employed while silently reading. Huey was one of the first to put it to paper, writing “The simple fact is that the inner saying or hearing of what is read seems to be the core of ordinary reading, the 'thing in itself,' so far as there is such a part of such a complex process” (Huey cited in Breen, 2014). The aim of this paper is to examine the empirical evidence that bears on the phenomenon expressed by this intuition, demonstrate its effects on silent reading comprehension, and explain a theoretical interpretation of the process underlying it. Breen is right to ask, “Is the producing, or hearing, of words and phrases during reading simply epiphenomenal — a by-product of the fact that language has been spoken far longer than it has been written, or does it enhance the reader's processing and understanding of the written word?” (Breen, 2014). The theoretical formulation of particular interest here is Fodor's Implicit Prosody Hypothesis (IPH), which posits that readers project prosody (i.e. intonation and/or rhythmic meter) onto written sentences which influences syntactic parsing (Fodor, 2002). We examine the slightly broader question: does implicit prosody play a significant part in sentence comprehension during silent reading? We begin by explaining what the IPH is and how it originated from Fodor's attempt to explain Cuetos and Mitchell's 1988 finding about Spanish and English attachment preferences. We then consider evidence for implicit prosody effects from Ashby and Clifton's 2005 work on lexical stress, and Hwang and Schafer's 2009 experiments demonstrating the effects of constituent length on overt prosody and syntactic parsing in Korean. Lastly, we look at Hwang and Steinhauer's 2011 ERP study suggesting a cognitive process indicative of implicit prosody, and consider what questions remain about the IPH and how we might seek their answers.

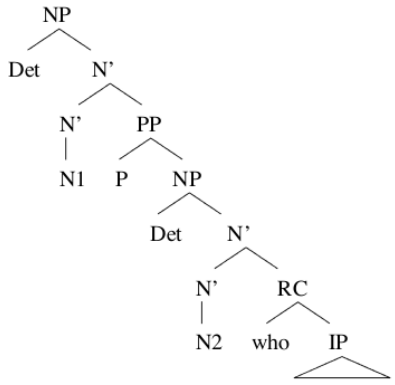
The Implicit Prosody Hypothesis proposes that during silent reading, a prosodic contour is projected onto the sentence which can determine the resolution of a syntactic ambiguity. In her seminal 2002 paper, Fodor anticipates the two natural logical objections. Firstly, since prosody must be projected onto a sentence based solely on the lexical items of the string and its perceived syntactic structure, how can it then influence the result of syntactic parsing? This chicken-and-egg problem only holds so long as the development of the chicken and the egg are assumed to each be singular, disjoint processes; what if, as Fodor suggests, sentence processing instead involves an interleaved process where some lower level syntactic parsing occurs first, informing prosodic projection, which in turn informs higher level syntactic parsing? Secondly, we must wonder: if the projected prosody is derived from the lexical items and syntax, any information it contributes must be redundant, so how could it make any difference with respect to resolution of the syntactic ambiguity? Perhaps implicit prosody exists, but in the end the final parse and interpretation would be the same as if it didn't. Fodor explains that the projected prosody can contribute non-redundantly in cases where not only the syntax, but also the interpretation of prosodic features is ambiguous. In this situation, prosodic structure can be projected based on one syntactic feature (e.g. Fodor's Same Size Sister (SSS) principle, which prefers adjacent phrases to have similar weight), but then interpreted differently as a marker of syntactic

structure. Why might readers “trick themselves” this way? Fodor hypothesizes that given a syntactically ambiguous sentence, a perceiver has a natural goal to resolve the ambiguity; because interpreting a prosodic break as the result of some other constraint does not help the ambiguity, the perceiver prefers to interpret it as a syntactic marker which decides the syntactic ambiguity one way or the other. This idea bears resemblance to the Minimal Attachment principle, which guides interpreters to resolving syntactic ambiguities by assuming the simplest tree structure.

With these philosophical objections out of the way, we can now examine the curious phenomenon which led to the formulation of IPH in the first place. In 1988, Cuetos and Mitchell reported a surprising difference in syntactic ambiguity resolution preference between Spanish and English (Fodor, 2002). In constructions with a relative clause (RC) attached to a complex NP with two possible noun hosts, English prefers local (“low,” in the syntactic tree sense) attachment of the relative clause, whereas Spanish prefers non-local (“high”) attachment, even though both parses are grammatical in both languages (see Figure 1). This appears to violate the Late Closure principle, which Fodor and others took to be a basic principle of sentence parsing universal across languages. Fodor also noted that, across all languages examined, the attachment preference was dependent on constituent length. This could be explained well enough by memory limitations, but the variance in preference between Spanish and English remained a mystery — any effect of limited memory capacity should presumably be equal across languages.

Prosodic effects, however, match this profile nicely: they vary with syntax and constituent

(3) a. *Local RC attachment (preferred in English)*



b. *Non-local RC attachment (preferred in Spanish)*

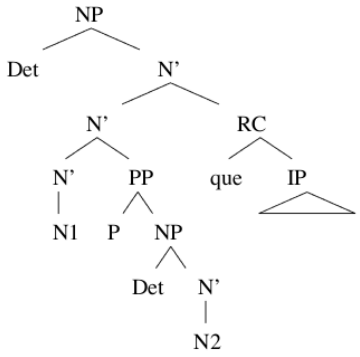


Figure 1: Trees for low and high attachment (Fodor, 2002).

length somewhat universally, yet retain some language-specific differences. A prosodic break is cross-linguistically more likely before a long RC than a short RC. It has been shown, at least in English, that the presence of a prosodic break before an RC increases the rate of high attachment interpretations (Maynell cited in Fodor, 2002). This neatly explains the dependence on constituent length. As for the variance between languages, Fodor concludes that the languages which prefer high RC attachment are exactly those whose “interface constraints” favor a prosodic break before an RC (Fodor, 2002). The dichotomy (e.g. Brazilian Portuguese, Egyptian Arabic, English, Swedish vs. Afrikaans, Dutch, Russian, Spanish) is not easily explained along typical typological boundaries, but Fodor presents one plausible analysis of constraints in English and French which give rise to a difference in preference for a prosodic break before RCs and, hence, a difference in syntactic attachment preference. Fodor and others have demonstrated similar phenomena supporting the IPH in many other languages (see e.g. Hirose, 2003; Lovric et. al., 2000). The effects of prosody in silent reading interpretation may prove varied and robust, and Fodor suggests that research must immediately begin questioning the prevailing procedure of ignoring them.

Ashby and Clifton's 2005 work demonstrates that the number of stressed syllables in a word — a prosodic property — affects reading time during silent reading. While not addressing exactly the same category of phenomena Fodor originally sought to explain with IPH, their study lends further evidence to the claim that some form of prosodic processing occurs during silent reading. The authors chose to study lexical stress because it is a much simpler prosodic feature than sentence-level prosody. It is not uniquely determined by lexical content, and can vary substantially according to individual preference, yet bears important effects on spoken language production and perception. (For example, unstressed syllables are more likely to be omitted than stressed syllables in adult speech; mispronounced stressed syllables impair lexical access more than mispronounced unstressed syllables; etc. (Ashby and Clifton, 2005).) Lexical stress shares with prosodic breaks the feature that it is not encoded at all in writing, and so makes a suitable place to look for implicit prosody effects on reading.

The authors chose eye movements as the medium of measurement, since fixation durations have previously been linked to word recognition processes (Rayner cited in Ashby and Clifton, 2005). Participants in their study read pairs of sentences containing words with either one or two stressed syllables, where the word pairs were matched for number of letters, number of syllables, part of speech, and frequency category (high or low). Ashby and Clifton took great care in their experimental design. They confirmed that there was actually a difference between the two categories of words in overt prosody (reading aloud) by measuring their duration and amplitude, established indices of stress (Beckman, 1986 cited in Ashby and Clifton, 2005). They confirmed that test subjects rated the two categories with similar levels of familiarity and naturalness of fit within their sentences. They even considered the effect of derivational morphology, ensuring that both categories had the same number of derived, non-derived, and derived with a stress shift words.

The three metrics Ashby and Clifton gathered were first fixation time, gaze duration, and number of fixations. First fixation times did not differ significantly between words with one and two stressed syllables, but gaze duration and number of fixations were significantly larger for words with two stressed syllables than one. All three metrics showed frequency effects for both categories, i.e. values were lower for more frequent words; frequency effects were especially significant for gaze duration and number of fixations.

Ashby and Clifton's experimental results, while small and simple, lead them to important conclusions. Considering that other factors which affect the speed of word recognition (e.g. frequency, homophony, ambiguity, predictability) generally affect first fixation time, the fact that lexical stress did not suggests that lexical stress information is not employed in the earliest stages of lexical access. Gaze duration, however, considered an indicator of the time to complete lexical access, was significantly affected by lexical stress, implying an application of lexical stress information in a later stage of lexical access (Ashby and Clifton, 2005). In a separate experiment, the authors found that naming duration also increased with number of stressed syllables, and on this basis hypothesized a connection between speech production processes and prosodic processing in silent reading.

Although their study was extremely well designed and controlled, there is perhaps one remaining question about Ashby and Clifton's methodology. While they did ensure that the average frequency of the one and two stressed syllable words was not significantly different, they did not (at least in their paper) examine the difference in frequency between each individual pair of words. While each pair was matched within the broad categories of "low" and "high" frequency, we might consider that the frequency differences within those categories could still be significant enough to confound metrics between the pair. (There could be significant frequency differences between particular pairs in opposite directions, and still no significant difference in average frequencies, so this possibility would best be investigated directly.)

Hwang and Schafer's 2009 study of the effects of constituent length on syntactic ambiguity resolution in Korean contributes classic Fodorian evidence for the IPH. While previous work had shown that features of overt prosody affect syntactic parsing and that overt prosody is itself influenced by constituent length and syntactic structure, Hwang and Schafer wanted to investigate directly the effect of constituent length on ambiguity resolution. In particular, they sought to control for more confounding factors than Fodor and others had done in manipulating constituent length, such as referential and syntactic complexity of constituents. Their syntactic test construction of choice is a sentence permitting attachment of a dative NP to either a matrix clause or relative clause position (see Figure 2). Both are grammatical in Korean, although prior work shows that the matrix clause attachment is generally preferred (Koh, 1997; Kiaer, 2007 cited in Hwang and Schafer, 2009). The authors predicted that a short subject NP would lend itself to intonational grouping with the dative NP, making a matrix clause attachment more likely (than usual), while a long subject NP would tend to be

(2a) Matrix clause association
Phigules-i Lopin-eykey [[Phwuwu-ka tta-cwu-n] pelcip-ul]
unkunsulccek phala-pelye-ss-ta.
Piglet-NOM Robin-DAT [[Pooh-NOM pick-BEN-REL] honeycomb-ACC]
stealthily sell-complete-PST-DECL.
'Piglet stealthily sold Robin [the honeycomb [that Pooh picked]].'

(2b) Relative clause association
Phigules-i Lopin-eykey$_i$ Phwuwu-ka t_i tta-cwu-n pelcip-ul
unkunsulccek phala-pelye-ss-ta.
Piglet-NOM Robin-DAT$_i$ Pooh-NOM t_i pick-BEN-REL honeycomb-Acc
stealthily sell-complete-PST-DECL.
'Piglet stealthily sold [the honeycomb [that Pooh picked for Robin]].'

Figure 2: Ambiguous Korean test sentence from Hwang, 2009.

produced with its own intonational boundary, making a relative clause attachment more likely.

To test their hypothesis, Hwang and Schafer performed a production experiment and a comprehension experiment. In the first part of the production experiment, to test first pass reading, they had participants read aloud ambiguous sentences with long or short matrix subjects. They isolated the variation of subject NP length by using long and short forms of names of well-known cartoon characters, such as (the Korean equivalents of) "Pooh" and "Winnie the Pooh," which presumably do not exhibit differences in syntactic, referential, or semantic complexity. For comparison, they looked at pairs of sentences which were identical except for the length of the subject. They used Praat to record and analyze the utterances, using word duration as an indicator of phrase-final lengthening (a prosodic boundary) and checking these boundaries with a trained human annotator. In the second stage of the experiment (second pass reading), they asked participants to indicate their preferred syntactic interpretation of the sentence they read, and (after explaining the ambiguity if necessary) they had them produce overt utterances for each of the interpretations.

In first pass reading of sentences with short subjects, the authors found that 88% of participants grouped the subject and dative NPs together with a prosodic boundary, following their prediction. With long subjects, 95% of participants produced an intonational boundary at the end of the subject, although most of them also produced a boundary at the end of the dative NP — something of a neutral prosody. (Interestingly, there was less neutral prosody in the second pass reading, suggesting that the prosody of planned speech differs from the prosody of first pass, online reading — the latter of which is probably more analogous to implicit prosody.) Thus it is clear that when reading aloud, speakers' overt prosody depends on constituent lengths. The second stage of the experiment showed that participants' preferred interpretation was significantly affected by the length of the matrix subject: 82% matrix attachment for short subjects, compared to 58% matrix attachment for long subjects (significant at $p < 0.01$). The authors suppose that the participants' interpretation was biased by the overt prosody they produced based on constituent length — a conclusion consistent with the IPH.

In the comprehension experiment, Hwang and Schafer used much the same setup: four experimental conditions, crossing ambiguous/unambiguous syntax with long/short subject. Again, they used long and short forms of names of well-known characters to vary subject length without affecting other factors. The sentences presented included a ditransitive verb and either one dative NP and an adverb or two dative NPs. The cases of interest are those with two dative NPs, as this disambiguates the syntax to a relative clause attachment (whereas the dative NP and adverb allows either matrix or relative clause attachment). The authors sought to observe whether the disambiguation point was more or less surprising depending on the length of the subject NP, as measured by self-paced reading times. To ensure that participants read the sentences normally (i.e. for comprehension), they inserted comprehension questions between trials.

As they predicted, Hwang and Schafer found that reading times for the disambiguation point were significantly longer for short subjects (which prefer matrix attachment of the dative NP) than long subjects (which prefer relative clause attachment), indicating that the length of the subject influenced participants' syntactic predictions. Reading times for the same segment of ambiguous sentences, with either subject length, were much shorter; this was expected since the ambiguity allows the reader to take their preferred interpretation without needing to perform any backtracking correction.

The results of these two experiments provide strong evidence that constituent length affects

(overt and presumably implicit) prosody, and that constituent length affects syntactic parsing preferences. However, they do not *directly* demonstrate that the generated prosody is the medium through which constituent length affects parsing. This seems altogether reasonable, and, as Fodor suggests, is the only explanation which also accounts for the language-specific differences in parsing preferences. We now turn to Hwang's later work to provide physical evidence of cognitive activity that suggests this account is correct.

Hwang and Steinhauer's 2011 work investigates the IPH through a novel methodology, the measurement of event-related potentials (ERPs). While the studies previously discussed provided strong evidence for an effect on syntactic parsing that seems well explained by implicit prosody, arguing in effect that "this sure looks like implicit prosody's shadow," the present study seems to give us a glimpse of the thing itself. The ERP effects of primary interest here are the closure positive shift (CPS) and the P600. The authors cite previous work demonstrating that the CPS is a reliable indicator of prosodic processing, both in reading (punctuation marks), listening, and production (e.g. prosodic breaks in speech). Previous work has also shown that the P600 corresponds to syntactic parsing difficulties, including those induced by garden path effects (Hwang and Steinhauer, 2011).

The procedure for this experiment follows most of Hwang's 2009 experiment, using the same sentence materials. The main difference is that the sentences are presented in rapid serial visual presentation (RSVP) instead of self-paced reading, since it is more typical for ERP studies. The authors conducted a pretest to confirm that this method would replicate the length effects and comprehension level of the self-paced reading study. Based on the results of that study, Hwang and Steinhauer predict that a CPS would be found after the subject in the long subject NP condition, indicating the insertion of a prosodic boundary there (as occurred in the overt prosody of 95% of first pass reading with long subjects). They also predict a higher likelihood of a CPS after the first dative NP in the short subject condition, since the subject and first dative NP can group together. Lastly, they predict a P600 effect at the point of disambiguation in the disambiguated condition that is stronger with a short subject (since the short subject disprefers the resolution to a relative clause attachment).

As expected, a CPS was observed at the end of long subject NPs, but not short ones. A small CPS was also observed at the end of the first dative NP with a short subject, also predicted, although the authors could not conclude whether a CPS was observed in this position with a long subject as well. They found a P600 after the disambiguating second dative NP with a short subject, but none with a long subject; this corresponds to the longer reading times found in the 2009 study. Hwang and Steinhauer also analyze other, smaller ERP effects and carefully consider the interactions between them, ultimately concluding that the CPS effects found do seem to correspond to prosodic boundaries generated during silent reading. This finding brings uniquely compelling evidence that implicit prosody has a measurable cognitive correlate, supporting Hwang and Schafer's conclusions and hence Fodor's IPH.

Fodor's Implicit Prosody Hypothesis still provides the only surviving explanation for the results of Cuetos and Mitchell 1988, and the clear effects of constituent length on overt prosody and syntactic parsing in Hwang and Schafer 2009, plus the ERP picture of implicit prosody we see in Hwang and Steinhauer 2011, offer strong support for Fodor's formulation of this subtle but fascinating phenomenon. With Ashby and Clifton 2005, we see that prosodic effects during silent reading are not limited to prosodic boundaries, but include lexical stress information as well. Given this, it seems most

likely that the projection of implicit prosody is in general — as many have previously intuited — a process very close to reading aloud, bringing all of its effects to bear on the processing of the read stimulus. It seems likely that more investigation would reveal observable effects of many other prosodic features on silent reading.

The non-redundancy of the information encoded in implicit prosody yields an interesting question. We have certainly seen that implicit prosody makes contributions which can change our interpretation of the written stimulus (from what it might have been if read with a specific overt prosody, for instance). Yet we can also objectively say that this information is not encoded; it is somehow generated, based on the stimulus but also on other knowledge of the language. To what extent, then, do the individual reader's prosodic habits determine the effect of implicit prosody? We might begin to investigate this question by performing versions of the experiments discussed here with children of varying ages, to see if they project different prosody and hence show different results. It would be fascinating to study whether deaf readers' results vary, given that they would necessarily not project auditorily-encoded prosodic information such as intonation, phrase-final lengthening, etc. As Fodor suggested in 2002, the effects of implicit prosody are significant and the process still mostly mysterious, leaving open a plethora of interesting questions and research directions.

References

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. Cognition, 96(3), B89-B100.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. Language and Linguistics Compass, 8(2), 37–50.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. Proceedings of Speech Prosody 2008, Aix-en-Provence, France.

Hwang, H., & Schafer, A. J. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. Journal of Psycholinguistic Research, 38, 151–175.

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean "garden path" sentences. Journal of Cognitive Neuroscience, 23(11), 3555-3575.