Prosodic focus strengthens semantic persistence

Word count: 10908

Abstract

Research on post-repair representations of garden path sentences has found that readers systematically arrive at misinterpretations even after displaying evidence of reanalysis (Christianson et al. 2001; Ferreira et al. 2001; Tabor & Hutchins 2004; Sturt 2007). These comprehension errors have been attributed to the incorrect but locally coherent analysis lingering past disambiguation, but the exact processes that are involved have been unclear (Slattery et al., 2013). A speeded auditory comprehension experiment examined the depth of semantic processing in the incorrect local parse as an independent factor driving the strength of the semantic persistence effect, drawing on known effects of pitch accent on the processing of focus-related semantic meaning (Watson et al., 2008; Fraundorf et al., 2010). Participants listened to sentences with local attachment ambiguity where either the adjunct subject or verb received a sharply rising pitch accent and answered a comprehension question. Analysis of responses after controlling for previously reported local coherence effects (Trueswell et al., 1994; Macdonald et al., 2004) found that a pitch accent on the verb significantly lowers accuracy. A post-hoc analysis of response times additionally revealed that listeners were garden-pathed in both pitch accent conditions, suggesting that a pitch accent on a syntactically ambiguous word gets interpreted with respect to the parse being pursued at any given moment. Findings show that a deeper semantic processing of the erroneous parse can increase its resistance to revision without interfering with structure-building, suggesting a greater level of independence between semantic and syntactic reanalysis.

Keywords: pitch accent, focus, sentence processing, garden-path effect, reanalysis

1. Background

1.1 Misinterpretations of garden path sentences

Despite the complexity of information in speech, comprehenders can process utterances rapidly, accurately, and effortlessly in conversations. This remarkable ability has been attributed in part to the incrementality of the human sentence processing mechanism which allows the structural representation of an utterance to be built up piece-by-piece as it unfolds in real time (Kimball 1973; Abney 1989; Crocker 1996). In other words, listeners do not wait until they finish hearing the entire sentence to start processing it. Rather, they are guided by their linguistic knowledge in using the early fragments of the input to generate expectations about what follows.

While this model of incremental parsing is powerful enough to explain people's ability to comprehend most utterances without much difficulty, a class of sentences called *garden path* sentences have been reported to disrupt this process by creating a temporary, local syntactic ambiguity for the parser (e.g., Bever 1970; Frazier & Fodor 1978; Frazier 1979). For example, in (1), the complementizer *while* projects the head of an adjunct clause. Because the presence of an adjunct clause signals the presence of a matrix clause that dominates it, the beginning of the sentence triggers an expectation for a structure which is minimally composed of two clauses – an adjunct clause followed by the matrix clause. However, up through the second NP headed by *baby*, it is ambiguous whether the boundary between the two clauses comes early after *dressed* or late after *cute*, as both (1a) and (1b) are grammatical continuations of the sentence fragment. Here, if the parser commits to the late-closure analysis of the adjunct clause VP, but the actual structure of the sentence is one where that VP closes early, this presents a serious challenge.

(1) While Anna dressed <u>the baby that was small and cute</u> …
   a. … spat up on the bed. [early closure]
   b. … the neighbor visited. [late closure]

Fortunately, the parser is not deterministic such that it simply breaks down when it is led down the garden path. Instead, it has the capacity to recover from an initial misparse and build the correct underlying structure, though often with great effort (Marcus et al. 1983). In reading studies, comprehenders have been reported to systematically display patterned behavior while processing ambiguous regions of the sentence that reflect this process of recovery from being garden-pathed, also called *reanalysis*.

For example, Frazier & Rayner (1982) examined the on-line behavioral responses to parsing difficulty in sentences like (2) in an eye-tracking-while-reading paradigm.

(2) Since Jay always jogs a mile seems like a very short distance to him.

Frazier & Rayner suggest that the parser initially interprets *a mile* as the direct object of *jogs*, although the correct analysis requires the adjunct clause VP to be closed early at the verb, with the NP *a mile* being analyzed as the subject of the matrix clause instead. This violation of the parser's expectation is reflected in longer looking times at the disambiguating region *seems like*, followed by a regressive eye movement to an earlier region of the sentence. This sequence of behaviors has been taken to be a sign of a successful reanalysis because readers do not appear to experience any further difficulty reading the rest of the sentence. More specifically, this reanalysis process appears to allow the parser to both discard the initially incorrect analysis from further consideration and also build the globally correct analysis. In sum, reanalysis was understood to be a mechanism that allows comprehenders to arrive at the correct interpretation of garden path sentences despite the initial difficulty processing them.

However, this all-or-nothing view of reanalysis has been challenged by studies that more closely examine the post-repair representations of garden path sentences. For example,

Christianson et al. (2001) report that after reading sentences like (3), comprehenders often failed to correctly answer simple questions like (4) even after reporting high confidence in their answers.

(3) While Anna dressed the baby played in the crib.

(4) Did Anna dress the baby?

If reanalysis indeed facilitated the parser's recovery from its initial mistake of attaching *the baby* as the direct object of *dressed*, then the post-repair representation of the sentence should no longer license the interpretation that Anna dressed the baby. Therefore, these comprehension errors provide evidence that the process of reanalysis as traditionally understood could sometimes silently fail at yielding the correct interpretation of the sentence. At the same time, however, comprehenders consistently answered questions like (5) correctly, suggesting that the failure of reanalysis is very particular and systematic.

(5) Did the baby play in the crib?

In light of this finding, early works by Ferreira and colleagues have made a broader claim about the relationship between sentence processing, the syntactic structure, and comprehension (e.g., Ferreira et al. 2001). In Ferreira et al. (2002), they claimed that the syntactic structure is "fragile" and not by itself a reliable source of sentence meaning for comprehenders, in contrast to the traditional view of sentence meaning as compositionally derived from the syntactic structure during processing. Instead, they argued that the communicative context in which the sentence is situated is what ultimately helps make a sentence's meaning available to comprehenders. In most natural conversations, the context of utterance reinforces the interpretation generated from the sentence's underlying structure, which allows people to comprehend utterances rapidly and accurately in real time. However, when comprehenders are deprived of the context, or if the available contextual information interferes with the interpretation generated from the underlying

structure, then the normal routines of comprehension are disrupted, resulting in representations of sentences that are merely "good enough."

Ferreira & Patson (2007) formalize this notion in what they call the Good Enough approach to human language comprehension, founded on the view that the language processing system, by design, builds representations of utterances that are simply "good enough" for the task at hand (e.g., comprehension in a conversations). More specifically, they propose that when the processing of a sentence becomes costly, comprehenders may resort to constructing a shallow representation of the sentence that may sometimes conflict in interpretation with the one generated from the faithfully-parsed syntactic structure. In this way, the Good Enough approach attempts to unify the phenomena of comprehension errors beyond just those involving sentences with local syntactic ambiguity, such as thematic role reversals in passive sentences and Moses illusions (Ferreira 2003).

However, there are strong reasons to maintain the belief that the parser is capable of building complex structures under pressure, and that the syntactic structure still plays a significant role in determining the possible interpretations that are made available to comprehenders. Decades of psycholinguistics research on the role of grammatical constraints in parsing have found that comprehenders are sensitive to the structural relationships between abstract linguistic units in a sentence that cannot be represented at a "shallow" level of processing. For example, readers have been shown to respect island constraints when positing gaps (Stowe 1986) and Binding Principles when forming pronoun-antecedent dependencies (Sturt 2003) in real time, neither of which are easily reducible to lexical identity, word order, pragmatic inference, and the like. Given these observations, the claim from early formulations of the Good Enough approach – that the syntactic structure can become a fragile source of meaning for comprehenders when the sentence is too complex – warrants greater scrutiny.

This is what motivated Slattery et al.'s (2013) study into the syntactic representations that underlie "good enough" interpretations. They lay out two competing hypotheses about the nature of post-repair syntactic representations when misinterpretations are observed. The first hypothesis is that the parser fails to build the correct structure in reanalysis, resulting in a degenerate structure that tolerates misinterpretations. The second hypothesis is that the correct structure is fully represented, but the locally coherent analysis lingers in some form and interferes with comprehension. Early proposals of the Good Enough approach reviewed in the previous section (e.g., Christianson et al. 2001; Ferreira et al. 2001) is most consistent with the first possibility, where an ill-formed and/or underspecified syntactic representation can drive comprehenders to become more reliant on external heuristics that operate quickly at the surface level of the sentence. However, if the second possibility is true and the correct syntactic structure is fully represented after undergoing reanalysis, that presents a potential challenge to this account: it would need to make an even stronger claim about the ability of surface-level heuristics to dominate structural information as the primary source of sentence meaning.

Slattery et al. test both possibilities in two separate eye-tracking-while-reading experiments. In the first experiment, they exploit the gender mismatch effect (Sturt 2003) as a diagnostic for whether the parser successfully builds the correct syntactic structure, using a pair of sentences such as in (6).

(6) After the bank manager telephoned David's <u>father/mother</u> grew worried and gave himself approximately five days to reply.

In the absence of the post-adjunct comma, the classic garden-path effect is predicted at the matrix verb *grew* for both sentences. But critically, if the parser is able to recover the correct structure after undergoing reanalysis, an additional difficulty at the pronoun *himself* is predicted when the

matrix subject does not agree in gender (i.e., when the matrix subject is *David's mother*) because the parser's search for the antecedent of the reflective should be constrained by the c-command constraint on reflexive binding (Chomsky 1981). In other words, if the post-repair syntactic representation specifies the c-command relationship between the matrix subject and the reflexive, and if the matrix subject is *David's mother*, then the parser will attempt to interpret *himself* as *David's Mother* and experience difficulty doing so. On the other hand, if reanalysis is "good enough" and simply yields a coarse-grained representation of the sentence such that the antecedent search is not syntactically constrained, then the mere presence of the noun *David* preceding the reflexive pronoun should be a sufficient cue to erroneously license the interpretation of *David* as the antecedent of *himself*. Interestingly, Slattery and colleagues observe a slowdown at the pronoun only when it does not agree in gender with the matrix subject, suggesting that the parser does indeed build rich and detailed representations of the syntactic structure even when comprehension is degraded.

At this point, if we still maintain the assumption that interpretations are constrained by the syntactic structure, then the parser must have built a structure that can license a misinterpretation at some point during processing, in addition to ultimately building the globally correct structure. This assumption bears out in Slattery et al.'s second experiment, which investigated whether the post-repair interpretation is consistent with that associated with the locally coherent parse by manipulating the material past disambiguation, such as in (7).

(7) While Frank dried off the truck that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

The adjunct verb *dried off* in the first sentence of (7) is temporarily ambiguous in transitivity until the matrix verb *was* signals that the NP headed by *truck* is not the embedded object – i.e., that the

transitive analysis of *dried off* is no longer viable. If that incorrect analysis nevertheless persists past disambiguation, comprehenders should experience difficulty processing *himself* in the following sentence, since the interpretation that Frank has finished drying himself off presupposes that he had been drying himself off before – an interpretation that is only available if the matrix verb *dried off* in the preceding sentence had been interpreted as a semantic reflexive at some point. Indeed, readers slow down at the pronoun well after the garden path effect in the first sentence had subsided, suggesting that some aspect of the initial misanalysis is persisting past the point of disambiguation.

Thus, Slattery et al. provide strong evidence that it is not the failure to build the correct structure, but rather the failure to discard the initial misanalysis, that leads to misinterpretations. This is consistent with previous findings for the semantic persistence effect as described in Sturt (2007), which studied reading behavior in garden-path sentences when they followed a context sentence. Sturt reports that given a context sentence like (8), late reading measures indicative of semantic integration were longer at the disambiguating region (*was actually*) when the content of the second sentence the was semantically inconsistent (8b) compared to when it was consistent (8a). This suggests that the initial semantic interpretation from the preceding sentence must have been available at the point of reanalysis to interfere in the process of semantic integration.

(8) The Antarctic expedition had been going on for months.

a. The explorers found the South Pole was actually right at their feet.

b. The explorers found the South Pole was actually impossible to reach.

1.2 Influences on the strength of semantic persistence

Two qualities of the merely locally-coherent analysis have been reported to contribute to the difficulty of syntactic ambiguity resolution: semantic fit and structure probability. In the context of the local attachment ambiguity illustrated in (3), these factors have often been discussed with respect to the appropriateness of the post-verbal NP (*the baby*) as the direct object of the adjunct verb (*dressed*) and the subcategorization frequency of that verb, i.e., how often *dressed* occurs in transitive and intransitive argument structures.

Early research on the effect of semantic fit on sentence processing have examined the ways in which the interpretation of an entity's thematic role can facilitate or impede its attachment as an argument of a particular interpretation of a verb. In an eye-tracking-while-reading experiment, Trueswell et al. (1994) found that for a minimal pair of sentences like *The defendant/evidence examined by the lawyer turned out to be unreliable*, reading times at the disambiguating post-verbal region *by the lawyer* were longer when the subject was *defendant* than when it was *evidence*. The increased difficulty found in the former condition is attributed to the fact that the subject *defendant*, which is initially assigned the role of the Agent, conflicts in the event structure of the correct relative-clause reading, which requires an analysis of *defendant* as the Patient of *examined*. Because this mismatch presented an immediate penalty to comprehenders, Trueswell and colleagues argue that thematic roles are satisfied in the earliest stage of processing, acting as a strong constraint on syntactic structure building. A consequence of this conclusion, that local structures are interpreted once they are built, is that there should also be differences that can be observed at the level of semantic processing. Pickering & Traxler (1998) argues this to be the case, reporting that the overall semantic coherence of the incorrect local analysis made recovery from garden paths more difficult. They suggest that readers make stronger semantic commitments to the

incorrect interpretation when it is highly plausible, which in turn makes it more resistant to revision (see also Tabor et al. 2004; Hare et al. 2009)

The second factor reported to affect ambiguity resolution, frequency, is another extensively studied constraint on parsing. MacDonald et al. (1994) proposes that the frequency of a verb's argument structures can bias which structural analysis the parser initially pursues. For example, given a pair of contrasting sentences like *The rancher knew that the nervous cattle pushed/moved into the crowded pen were afraid of the cowboys*, MacDonald (1994) observes a garden pathed effect only when the embedded verb was *pushed*, which is frequently transitive, although both *pushed* and *moved* have a transitive interpretation. This and other empirical evidence for the pervasive effect of the frequency of argument structure frames on structure-building (e.g., Garnsey et al. 1997) have motivated models of parsing such as that outlined in Gibson & Pearlmutter (1998; 2000), in which the structural probability of a local analysis affects its activation level in competition with alternative analyses over the region of ambiguity. In this architecture, a possible local parse needs not to have been actively pursued for the interference to be observed. Although the parser may only build the correct structure, there can be differences in the difficulty of doing so, affected by the activation levels of competing incorrect analyses that cause interference over a region of ambiguity.

These ideas were adopted in later models that explicitly aim to capture comprehension difficulties associated with the semantic persistence effect, such as Tabor & Hutchins' (2004) dynamic self-organized (DSO) parsing model and Ferreira and colleague's lexicalized tree-adjoining grammar (LTAG) model (Ferreira et al., 2004; Lau & Ferreira 2005). These proposals share the capacity for the parser to not only build and maintain multiple local structures over a point of ambiguity, but also to allow for a local structure to linger in the syntactic representation

if it is sufficiently activated during the process of ambiguity resolution. In other words, a locally coherent structure could fail to decay after disambiguation even if the parser ultimately pursues the globally correct structure. If the incorrect local structure can be represented alongside the globally correct structure in this a way, it could explain how comprehenders arrive at misinterpretations despite having successfully built the correct structure, as reported in Slattery et al. (2013).

In sum, prior findings suggest that the overall semantic fit and structure probability of the incorrect local analysis can conspire to strengthen that analysis in processing, making disambiguation more difficult and misinterpretations more likely. Yet, it is still unclear whether the interference from the lingering misanalysis is with respect to semantic or syntactic reanalysis. This is because both semantic fit and structure probability, while very different in their nature, are ultimately about the content of a particular analysis and do not provide straightforward information about the level of representation at which they facilitate processing. One possibility, consistent with DSO and LTAG models, is that the parser uses information about both factors in the evaluation of all possible local structures and ranks them as part of the ambiguity resolution process. This assumes that the parser has not made yet a commitment to a particular structural analysis when both semantic interpretations are available. Under this interpretation, the structure of the incorrect analysis may linger and interfere with overall comprehension when it is highly activated and fails to decay over time. Another possibility, consistent with Sturt's (2007) original formulation of semantic persistence, is that the incorrect local analysis is first pursued before the globally correct analysis is made available to the parser, and that it is merely the semantic interpretation from that initial misanalysis which lingers. Under this interpretation, the parser may make stronger semantic commits in the initial misparse if the structure is very probable and the

interpretation is very coherent, strengthening its resistance to revision during semantic reanalysis. However, no additional difficulty is expected for syntactic reanalysis, which can simply proceed without an accompanying revision of the semantic representation.

1.3 Prosodic prominence and semantic focus

In the Autosegmental Metrical (AM) model of English intonational phonology, the pitch contour of an utterance is understood to be derived from interpolation between discrete tonal targets which are anchored to syllables that are phrase-level peaks of prominence (Pierrehumbert 1980; Beckman and Pierrehumbert 1986). Phrasal prominence is determined from a hierarchically organized metrical structure, which defines a strong-weak patterning over words at the phrase level. In English, the default prominence pattern has an obligatory prominence on the rightmost content word in the prosodic phrase. For a simple sentence with no embedded clause, produced in a neutral speaking style, the verb and its (non-clausal) complements constitute one prosodic phrase, and a non-pronominal subject NP may optionally constitute a separate prosodic phrase, or may be integrated into the phrase containing the verb (Büring 2016). The obligatory prominence at the right edge marks the location of the main phrasal stress and is termed the *nuclear* prominence (Chomsky & Halle 1968). If a phrase is long enough, there may be additional *prenuclear* prominence assigned to words in metrically strong positions earlier in the phrase (Calhoun 2006). A word with phrasal prominence licenses a tonally specified pitch accent (Truckenbrodt 1995), which is obligatorily assigned to the word with nuclear prominence and is optionally assigned as a "rhythmic" or "ornamental" feature of words with prenuclear prominence (Calhoun 2006; Büring 2016). The following example illustrates the default metrical structure and accent assignment for the sentence *Northwestern is a university in Evanston*.

H* L- H* L- L% — *tonal sequence*
(x) — *metrical grid*
(x) (x) — *metrical grid*
(x) (x) (x) — *metrical grid*

(9) [S [NP Northwestern] [VP is [NP a university] [PP in Evanston]]].

In addition to metrical structure, pragmatic and semantic factors also play a role in the distribution of phrasal prominence. For instance, accent placement is sensitive to distinctions in the information status of words within a prosodic phrase; specifically, words that are accessible from the prior discourse context (*discourse-given*) are generally not eligible for nuclear prominence. In conversations, the canonical ordering of sentence constituents is given information before new information, with the result that the new information is located in the default position for nuclear pitch accent (Birner & Ward 1998). For example, in (10), the italicized portion of Speaker B's response is new in the discourse because it is the answer to Speaker A's question, while the underlined portion is old in the discourse because it was presupposed in the question.

(10) A: What did John eat for dinner?

B: <u>He ate</u> *cake*.

Prominence assignment is also conditioned by semantic focus. A word with contrastive, narrow, or corrective focus is preferentially assigned nuclear prominence, even if it occurs in the non-final position in the sentence (Büring 2016). For example, in the sentence *SUZY showed up to class*, the prominence on *SUZY* can be interpreted as a marker of contrastive focus. In the framework of alternative semantics, focus is understood as a semantic notion which introduces references to a set of focus alternatives (f-alternatives) and triggers a set of prepositions which contrast in the focused element (Rooth 1992).

In English, tonally distinct pitch accents are also used to mark these graded distinctions in givenness and semantic focus (Baumann & Riester 2012). In particular, new and/or contrasting

information is often marked with an H* or L+H* (Watson et al. 2008). For example, in (11), Speaker B's response has the nuclear prominence on *cake,* and would typically be assigned a H* or L+H* pitch accent marking narrow focus (as the answer to the prompting question). [1]

(11) A: What did John eat for dinner?

B: He ate CAKE.

Experimental works on prosodic focus over the last several decades provide supporting evidence of a systematic relationship between pitch accents and focus, both in production and comprehension (see review in Cutler et al. 1997; Wagner & Watson 2010; Wagner 2019; but see also Chodroff & Cole 2019 and Roettger, Mahrt & Cole 2019 for evidence of a weaker relationship). For example, studies using eye-tracking in the visual world paradigm have found that listeners can rapidly access information-structural meanings from pitch accents in complex comprehension tasks (Ito & Speer, 2008; Kurumada et al., 2014). Furthermore, interpretations of focus-related semantic meaning have been found to trigger deeper semantic processing that has consequences for encoding and retrieval in memory. In a series of recognition tasks, Fraundorf et al. (2010) found that listeners were better able to recall a word from a prior discourse when the word received prosodic prominence. Critically, listeners performed better on the recognition task when the choice was between contrasting referents, meaning that the processing of focus-related meaning from contrastive pitch accent facilitates both identification and discrimination.

These findings indicate that listeners rapidly and automatically compute the rich semantic and pragmatic meanings of pitch accents in sentence processing. However, as already noted, not all pitch accents mark semantic focus on the accented word. Rather, prosodically licensed semantic focus scopes over syntactic constituents in a particular derivation process known as focus

[1] We use the term "prominence" to broadly refer to both the H* and L+H* pitch accents in the ToBI framework given that their domains of interpretation overlap (Watson et al., 2008; Calhoun, 2012).

projection (Chomsky 1972; Gussenhoven 1992; Selkirk 1995). Revisiting the example in (11), because the most prominent pitch accent in Speaker B's answer (by virtue of being the only one present) occurs on *cake*, it specifically marks that NP as the answer focus. The focused constituent (also called the domain of focus) is marked with $[]_{FOC}$ brackets in (12), following the notational conventions used in Selkirk (1995). It is due to this ability for focus to project from the accented word and percolate up the syntactic structure that the same response by Speaker B in (12) is also an appropriate answer to a broader, VP-focus question like (13).

(12) A: What did John eat for dinner? [narrow NP-focus question]

B: He ate $[_{NP}$ CAKE$]_{FOC}$.

(13) A: What did John do? [broad VP-focus question]

B: He ate $[_{NP}$ CAKE$]_{FOC}$.

Theories of the intonational grammar of American English generally agree that, to mark semantic focus, the last pitch accent in the focus domain must bear the nuclear pitch accent and all subsequent material in the sentence must be deaccented (Jackendoff 1972; Büring 2016). This focus-to-accent relationship becomes clearer in sentences where focus and default accenting locate the nuclear pitch accent on different parts of the utterance. For example, the default accenting pattern in (14) assigns prominence to *CLASS*, the rightmost word in the intonational phrase eligible to receive a pitch accent. However, an interpretation of contrastive focus over *Suzy* requires that the nuclear pitch accent occurs within the domain of focus, which is realized in (15) with an accent on the word *Suzy* itself. In turn, the accent on *class* is removed, as per the rule of post-focal deaccenting (Jackendoff 1972).

(14) Suzy showed to CLASS.

(15) No, not John. $[SUZY]_{FOC}$ showed up to class.

That the domain of focus must minimally contain the nuclear pitch accent is also reflected in the observation that focus is not realized by simply placing a pitch accent just anywhere in the domain of focus. For example, given Speaker A's question in (16), the verb phrase *visited a friend in Chicago* in Speaker B's answer constitutes the domain of focus. This may be realized prosodically with a pitch accent on *Chicago* (16a), but it would be inappropriate to have the prominence fall on any other word in the focus domain, such as on *visited* (16b). Because the verb is phrase-medial and is a predicate that precedes an argument, a pitch accent on the verb makes only the narrow verb-focus interpretation available, as shown in (17).

(16) A: What did you do last weekend?

a. B: I [$_{VP}$ visited a friend in CHICAGO]$_{FOC}$.

b. ?B: I [$_{VP}$ VISITED a friend in Chicago]$_{FOC}$

(17) I [$_{V}$ visited]$_{FOC}$ a friend in Chicago.

Given these constraints on the interpretation of focus from prosodic prominence and previous findings on the effect of pitch accent on real-time semantic processing, we can make predictions about the interpretation of prominence in the processing more complex sentences. Consider again the garden-path sentence *While Anna dressed the baby played in the crib* from Section 1.1. Over the course of processing this sentence, the parser is expected to have built both the incorrect and the correct structure – either simultaneously (parallel) or in succession (serial) – due to the local attachment ambiguity at the post-verbal noun and the general parsing preference for a late closure of the adjunct clause VP. As noted earlier, the incorrect local analysis that was temporarily entertained over the region of ambiguity may persist in memory depending on the degree of its coherence, but the likelihood and strength of persistence may also be affected by prosodic factors as well.

For example, the sentence can be uttered with a sharply rising pitch accent on the adjunct verb, as in *While Anna DRESSED the baby played in the crib*. Here, the pitch accent may be immediately interpreted with respect to the parse(s) that are under consideration at that moment, strongly biasing an interpretation of narrow focus on the verb in the incorrect late-closure analysis since the verb occupies a non-canonical, phrase-medial position in that structure. However, the same pitch accent may be interpreted as simply licensing the prosodic phrase in the correct early-closure analysis, since it occurs in the canonical position for nuclear stress in that structure: it is the last word in the prosodic phrase eligible to receive a pitch accent.

This yields an interesting prediction: the prominence on the adjunct clause verb should result in an *asymmetrical interpretation of focus* between the two local parses where only the incorrect parse is enriched for focus-related semantic meaning. The differences in semantic representation at the level of both the ordinary ("event") semantic meaning (Carlson 1984) as well as the focus ("alternative") semantic meaning (Rooth 1992) between the early-closure and late-closure analyses when the adjunct verb is accented is illustrated in (18) and (19), respectively.

(18) While Anna $[_{VP}\,[_{V}$ DRESSED$]_{FOC}$ the baby], …

Ordinary semantic meaning: $\exists e$[dress(e) & Agent(e, Anna) & Theme(e, the baby)]

Focus semantic meaning: $\exists e$[P(e) & Agent(e, Anna) & Theme(e, the baby)]

(19) While Anna $[_{VP}$ DRESSED], the baby …

Ordinary semantic meaning: $\exists e$[dress(e) & Agent(e, Anna)]

Focus semantic meaning: $\emptyset$

If the pitch accent on the verb facilitates an interpretation of semantic focus only in the incorrect local parse, and if that in turn makes the erroneous interpretation more likely to persist, listeners should have difficulty accurately comprehending the sentence. This should lead to lower accuracy rates on comprehension question that ask whether the second NP is the embedded object (e.g., *Was*

the baby dressed?) compared to when the pitch accent falls on the adjunct subject instead. In the baseline subject accent condition, the pitch accent is simply interpreted for the referential meaning of the subject (*Anna*) before the parser encounters the local syntactic ambiguity and is thus predicted to have no differential impact on the process of ambiguity resolution down the line. Critically, if this effect of pitch accent on the depth of semantic processing drives the persistence effect independently of the extent of locally determined coherence, then the location of the pitch accent should emerge as a significant predictor of accuracy on comprehension questions after controlling for the item-level effects from semantic fit and transitivity bias.

2. Experiment

Two off-line norming experiments and a speeded auditory comprehension experiment were created and hosted on PCIbex (Zehr & Schwarz 2018) and administered through Prolific, a platform for online subject recruitment. Thirty adult English monolingual speakers between 18 and 65 years of age residing in the United States were recruited for each of the two norming experiments, and sixty-one participants of the same demographic background – with the additional qualification of having no known hearing issues – were recruited for the speeded auditory comprehension experiment. All experiments reported in this study were approved by the Northwestern University Institutional Review Board as compliant with the ethical standards for research on human subjects.

2.1 Design and norming studies

The critical items for the auditory comprehension experiment consisted of twenty-four garden-path sentences with closure ambiguity like those in (20) below. The adjunct clause consisted of a complementizer (*Since*, *While*, or *When*) followed by a two-syllable name with a strong-weak

stress pattern and an ambiguously transitive verb in the past tense. The matrix clause consisted of a subject noun phrase that was also a possible direct object candidate for the adjunct clause verb, followed by a matrix verb phrase.

(20) While Anna dressed the baby stopped crying.

An initial set of twenty-six sentences were normed for biases in *semantic fit* (how often the interpretation of the second NP as the embedded object is preferred) and *transitivity bias* (how often the adjunct verb is followed by a direct object) in two independent off-line experiments.

Semantic fit was normed in a three-alternative forced-choice task, in which participants were asked to choose the more plausible event between the early-closure interpretation such as (21), the late-closure interpretation such as (22), or an option for both being equally plausible. For each item, the proportion of responses preferring the early-closure interpretation against those preferring the late-closure interpretation was calculated (*median* = 0.53*, inter-quartile range* = 0.38). Semantic fit scores of each item were normalized by centering all the responses at 0.5 and downweighing the mean by the number of "both" responses.[2]

(21) While Anna dressed, the baby stopped crying.

(22) While Anna dressed the baby, he stopped crying.

Transitivity bias was normed in a sentence completion task, in which participants saw fragments of each sentence up to the adjunct clause verb, such as in (23), and were asked to type in a continuation that completes the sentence. Responses that failed to complete the sentence were excluded from analysis and the rest were coded for the presence of a direct object following the adjunct clause verb. The proportion of responses involving direct objects was calculated (*median* = 0.38, *inter-quartile range* = 0.60) and again centered at 0.5.[3]

[2] Normalized semantic fit score = early-closure/(early-closure + late-closure) * (1 – "both"-response/total)
[3] Normalized transitivity bias score = direct object/(direct object + no direct object)

(23) While Anna dressed _____ .

The collected measures of semantic fit and transitivity bias were z-scored to be entered into the statistical model as independent item-level predictors (i.e., nuisance variables). Figure 1 shows the distribution of semantic fit and transitivity bias scores for all items used in the experiment. The variability among items along both dimensions allows the model to statistically control for their effects on comprehension accuracy and isolate the effect of pitch accent placement (further detailed in the Analysis section).

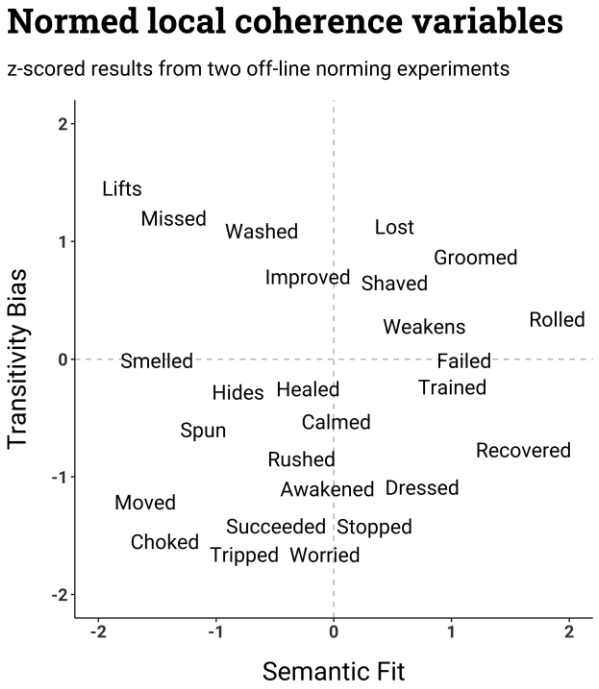


Figure 1: Norming scores for transitivity bias and semantic fit of the sentences used as stimuli for the speeded auditory comprehension experiment.

2.2 Audio stimuli

The stimuli for the speeded auditory comprehension experiment created for a final set of twenty-four normed sentences.[4] A trained male speaker recorded each sentence with two accent

[4] Two items were removed to maximize variation of local coherence effects among the items.

patterns: one with prominence on the adjunct subject, such as (24), and another with prominence on the adjunct verb, such as (25).

(24) While ANNA dressed, the baby stopped crying. [subject accent, early closure]

(25) While Anna DRESSED, the baby stopped crying. [verb accent, early closure]

In the construction of the auditory stimuli, we considered the possibility that an interpretation of a prosodic phrase boundary after the adjunct verb may immediately resolve the local attachment ambiguity, much like the role of a comma in reading. The effect of an intervening intonational phrase boundary on the blocking of local attachment is well-documented and can be signaled via a combination of a pause, pitch reset, and pre-boundary lengthening (Kjelgaard & Speer 1999; Schafer et al. 2000). In order to make these cues ambiguous, so as to induce the garden path effect, a set of late closure variants of the stimuli, such as (26) and (27), were also recorded by the same speaker. From the late closure recordings, the durations of the words over the region of ambiguity that are potential targets of pre-boundary lengthening – the adjunct verb (*dressed*) and the post-verbal noun (*the baby*) – were measured and referenced in the resynthesis of the early-closure stimuli used in the experiment.

(26) While ANNA dressed the baby, he stopped crying. [subject accent, late closure]

(27) While Anna DRESSED the baby, he stopped crying. [verb accent, late closure]

In sum, a total of four recordings were created for each sentence. Figure 2 shows the distribution of durations for the two words that are potential candidates for pre-boundary lengthening in the four raw recordings (in grey) as well as in the resynthesized stimuli (in black).

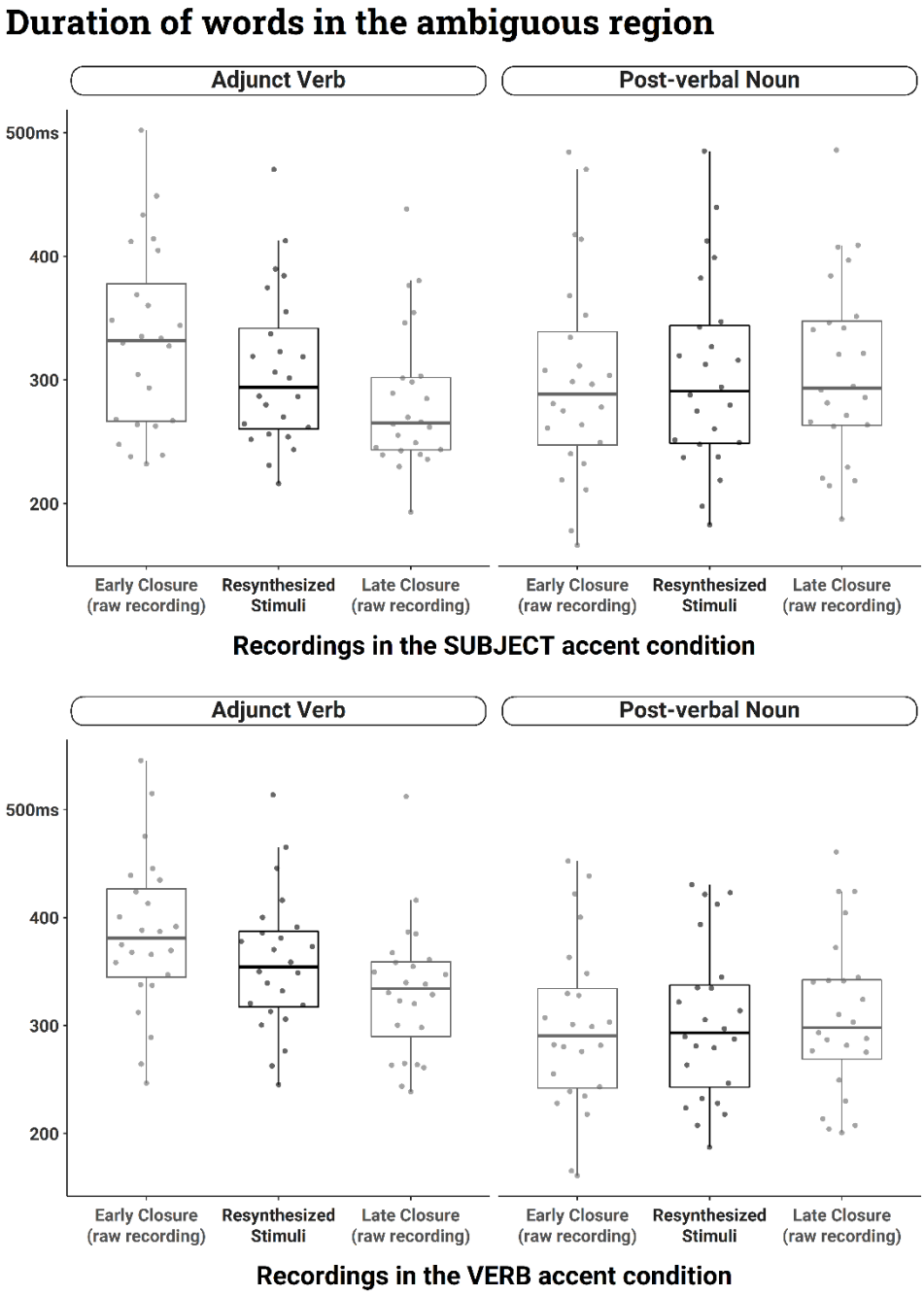


Figure 2: Distribution of durations (in milliseconds) for words in the ambiguous region that are candidates of pre-boundary lengthening: the adjunct verb (left) and the post-verbal noun (right). The duration distributions are plotted for recordings in the subject accent condition (top) and the verb accent condition (bottom). The raw recordings are plotted in grey and the resynthesized stimuli which averages between the raw recording are plotted in black and placed in the middle.

We used the PSOLA feature in Praat (Boersma & Weenik 2020) to resynthesize the duration of the adjunct clause verb and the post-verbal noun to their averages between the early closure and

the late closure variants, using the early-closure recording as the source recording. After controlling for pre-boundary lengthening cues in this way, the end-of-sentence region (material after the post-verbal noun) from the subject-accent recordings was spliced into the verb-accent recordings to control for any sentence wrap-up effects between the two accent conditions. Lastly, pitch movement over the main clause was flattened to induce an unambiguous interpretation of prominence in the adjunct clause. The pitch contours of the subject accent stimuli and the verb accent stimuli for a sentence used in the experiment are shown in Figures 3 and 4 respectively, where the red dotted line after the ambiguous region marks the location of the splice.

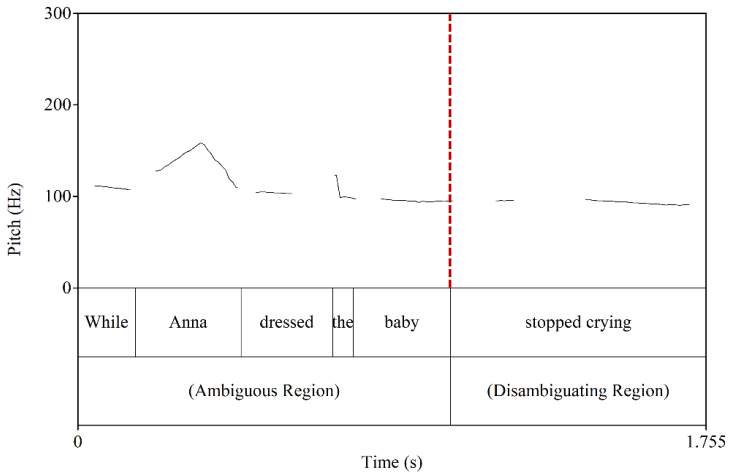


Figure 3: An example pitch contour of a subject-accent condition trial.

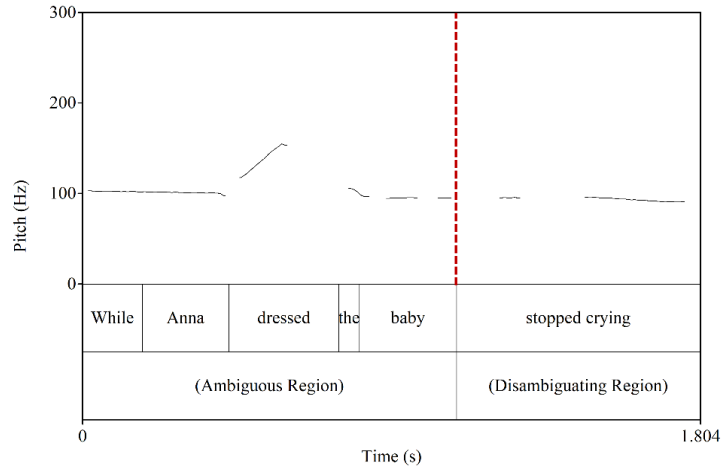


Figure 4: An example pitch contour of a verb-accent condition trial.

Because prominence on a word has consequences for the prosodic structure of the entire phrase, acoustic measures between the conditions were not further controlled in order to preserve naturalness. Table 1 reports the unaltered measurements for pitch (max f0) on the adjunct verb and the post-verbal noun in both conditions from the resynthesized stimuli.

Table 1: The mean, standard deviation, and distribution of max F0 values on the adjunct subject and the adjunct verb for the resynthesized stimuli in the two accent conditions.

		Max F0		
Condition	Word	Mean	Std. Dev.	Distribution
Subject Accent	Adjunct Subject	155.2	9.5	
	Adjunct Verb	118.7	14.1	
Verb Accent	Adjunct Subject	109.0	8.1	
	Adjunct Verb	150.1	10.5	

For each sentence, a corresponding comprehension question similar to that used in Christianson et al. (2001) was constructed. These questions asked whether the post-verbal noun was the direct object of the adjunct clause verb, such as in (28). Because the post-verbal noun in the critical sentences were always the matrix clause subject, the correct answer to all the comprehension questions for the critical trials were *NO*.

(28) Was the baby dressed? *YES/NO*

In addition to the critical items, forty-eight distractor sentences were recorded by the same speaker to balance the transitivity bias of the adjunct verb (ambiguous, obligatorily transitive, obligatorily intransitive), the location of prosodic prominence (first noun, first verb, second noun), and the sentence's syntactic structure (early closure, late closure) across all trials presented to the participant in the experiment. The comprehension questions were also overall balanced in the correct answer choice and varied in the syntactic position of the entity being asked about. Lastly,

two catch trials which simply instructed participants to answer *Yes* or *No* to the following question were included to filter out inattentive listeners.

As discussed previously, our hypothesis predicts a significant negative effect of verb pitch accent on comprehension accuracy. If a pitch accent on the verb leads to an asymmetrical interpretation of focus that triggers deeper semantic processing in the incorrect analysis, listeners should be more likely to answer the comprehension question in a way that is consistent with that erroneous interpretation. Importantly, this effect of pitch accent location on accuracy should reach significance after controlling for transitivity bias and semantic fit. If previous studies reviewed earlier are correct about these two factors contributing to the incorrect of the merely local parse, we expect an additional negative effect of both on accuracy. The presence of an adjunct verb that is transitively biased and a following NP that is a plausible direct object candidate to that verb should conspire to strengthen the misanalysis, making misinterpretation more likely.

2.3 *Procedure*

Participants were given 5000ms after each sentence played to answer the comprehension question that followed. If an answer was not given within the time window, the next trial automatically proceeded. An exposure phase consisting of two sentences with the same structure as the critical sentences was presented at the beginning to get participants accustomed to the task. In the experiment phase, participant heard a total of seventy-four sentences consisting of the twenty-four critical trials, the forty-eight distractor trials, and the two catch trials. The experiment was run in a within-participant design and participants heard the critical sentences in twelve subject-accent and twelve verb-accent conditions, presented in random order with the other trials.

The experiment lasted around ten to fifteen minutes on average. The answers and the response times to the comprehension questions were recorded for analysis.

2.4 *Results*

Prior to analysis, critical trials where participants timed-out or answered the comprehension question under 300ms were excluded (less than 2 percent of trials).[5] Additionally, out of sixty-one participants, one was excluded due to an accuracy rate on the distractor trials that was lower than 75 percent. Data from the remaining trials from the sixty participants were included in the analysis.

Statistical analysis was conducted by fitting a logistic mixed-effects regression model to the accuracy of the responses using the *lme4* package in R (Bates et al. 2015), with pitch accent condition, semantic fit, and transitivity bias as fixed effects without interaction terms, and subject and item as random effects.[6] Convergence issues due to overparameterization of the random effects structure were addressed through the iterative model reduction process suggested in Bates et al. (2015), by which the correlation between random effects were first dropped and individual random effects were removed in the order of lowest variance until the model converged to keep the random effects structure as informative as possible. In the converging maximal model, intercepts were allowed to vary by item, and intercept and slope of the pitch accent condition were allowed to vary by subject. To test significance, the full model was compared via a log likelihood test to a depleted model where the predictor of interest was removed.[7] Table 2 reports the log-odds

[5] Given that the fastest response time on the attention-checking filler trial (which simply instructs the listener to answer Yes or No to the following question) across all participants was around 400ms, we reasoned that it would be improbable for a participant to read and answer an actual comprehension question under 300ms.

[6] Accuracy ~ PitchAccent + SemanticFit + TransitivityBias + (1 + PitchAccent + SemanticFit + TransitivityBias | Subject) + (1 + PitchAccent | Item)

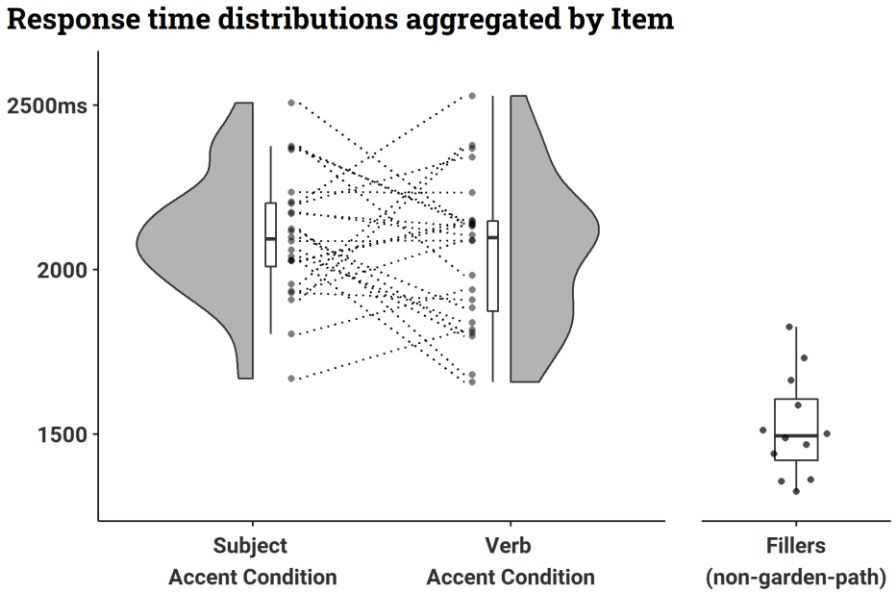
[7] For example, with pitch removed: Accuracy ~ SemanticFit + TransitivityBias + (1 + SemanticFit + TransitivityBias | subject) + (1 | item)

estimates and standard errors from the converging model[8] and the chi-squared statistics and *p*-values from the log likelihood tests.

Table 2: The point estimate and standard error (in brackets), chi-squared value, and the *p*-value (bold for reaching significance $p < 0.05$) from the logistic mixed effects model fitted to accuracy.

	Estimate (SE)	$\chi^2(1)$	*p*
Pitch (Verb)	-0.19 (0.08)	5.41	**0.020**
Semantic Fit	-0.44 (0.15)	7.45	**0.006**
Transitivity	-0.19 (0.16)	1.39	0.238

A post-hoc analysis of response times was also conducted to confirm that listeners were indeed garden-pathed in both pitch accent conditions, as intended in the design. The majority of the comprehension questions were answered well within the five-second window (*mean* = 1.8, *std. dev.* = 0.7). Figure 8 shows the distribution of response times aggregated by item between the two conditions, where each line tracks an item's mean response time across conditions. The response time distribution for a subset of the filler items that were non-garden-path sentences containing an unambiguous marking of a prosodic phrase boundary is also plotted to the right for comparison.



[8] Accuracy ~ PitchAccent + SemanticFit + TransitivityBias + (1 + PitchAccent | Subject) + (1| Item)

Figure 5: Response time distributions of the critical trials by condition (raincloud plot to the left) and a subset of the filler trials (boxplot to the right). Each dot represents the mean response time of an item. The slopes of the lines connecting the dots represent the effect of pitch accent condition on response time.

In addition, a linear mixed-effects regression model was fitted to log-transformed response time with pitch accent placement, semantic fit, transitivity bias, and accuracy as fixed effects without interaction terms, and subject and item as random effects with varying intercepts.[9] Table 3 reports the estimates and standard errors from the model and the chi-squared statistics and *p*-values from the likelihood ratio tests.

Table 3: The point estimate and standard error (in brackets), chi-squared value, and the *p*-value (bold for reaching significance $p < 0.05$) from the linear mixed effects model fitted to response time.

	Estimates (SE)	$\chi^2_{(1)}$	*p*
Pitch (Verb)	-0.009 (0.008)	1.13	0.289
Semantic Fit	0.031 (0.014)	4.50	**0.034**
Transitivity	0.035 (0.015)	5.26	**0.022**
Accuracy	0.065 (0.021)	13.34	**<0.001**

3. Discussion

The auditory comprehension experiment investigated the independent effect of the depth of semantic processing on the persistence of the erroneous, locally coherent interpretation in sentences with local attachment ambiguity. Depth of semantic processing was manipulated through the marking of contrastive focus, under the assumption that a sentence with contrastive focus undergoes a deeper level of processing that involves the identification of semantic alternatives to the focused constituent. Contrastive focus was prosodically marked by a sharply rising pitch movement on the adjunct clause verb, which facilitates an interpretation of contrastive

[9] LogResponseTime ~ PitchAccent + SemanticFit + TransitivityBias + Accuracy + (1 | Subject) + (1| Item)

focus only in the incorrect, late-closure parse due to the non-canonical, phrase-medial position of the accent in that structure. This contrastive focus in the incorrect parse was predicted to trigger deeper semantic processing of that analysis, thereby strengthening its persistence in memory and lowering overall accuracy on comprehension questions. In comparison, sentences with a pitch accent on the adjunct clause subject facilitate an unambiguous interpretation of focus at this earlier position in the sentence, outside the region of ambiguity, and was thus predicted to have no impact on ambiguity resolution.

The results show a significant negative effect of verb pitch accent on accuracy (β = -0.19, *std. err.* = 0.08, χ^2 = 5.41, p = 0.02), providing evidence that the processing of focus-related semantic meaning within the incorrect analysis strengthened semantic persistence. Critically, this effect is significant after controlling for item-specific local coherence effects from semantic fit and the transitivity bias. Therefore, we find that pitch accents can have immediate semantic consequences in the processing of syntactically ambiguous material, complementing previous findings on the rapid interpretation of focus-related semantic meaning signaled by prosodic prominence in complex comprehension tasks.

It may be possible for there to be alternative explanations for the effect of pitch accent on accuracy that are based on an account of the pitch accent as forcing the parser to commit to a particular structural analysis. For example, the effect of pitch accent placement on accuracy could be that, for whatever reason, the verb-accent condition leads the parser down the garden-path while the subject-accent condition does not. If issue is about whether the parser was garden-pathed or not in this sense, we expect a strong corresponding effect of pitch accent on response time given the well-attested causal relationship between being garden-pathed and increased processing time due to reanalysis (Frazier 1979; Frazier & Rayner 1982). Specifically, we would predict shorter

response times in the subject-accent condition that are similar to the response times for the non-garden-path filler trials if the parser pursued the correct analysis from start to finish without encountering any ambiguity. Instead, as shown in Figure 5, we find that both pitch accent conditions have longer response times than the non-garden-path distractor trials. This indicates that the parser was garden-pathed regardless of the location of the pitch accent in the adjunct clause, as intended by the design of the resynthesized stimuli that controlled for the multiple phonetic cues to a prosodic phrase boundary to be uninformative.

At this point, we have strong evidence for two claims. First, misinterpretations are driven by the interference from the semantic interpretation associated with the incorrect structural analysis, the strength of which can be modulated by prosodically marked focus-related semantic meaning. Second, independent of this effect of pitch accent placement on comprehension, the local syntactic ambiguity is still present at the post-verbal noun and the parser must eventually build the structure that is faithful to the input. These conclusions from the data are broadly consistent with the account of semantic persistence as proposed in Sturt (2007) and Slattery et al. (2013), as well as our hypothesis that a pitch accent on a syntactically ambiguous word is immediately interpreted with respect to the syntactic structure of the parse(s) being considered, at least in the absence of clear phonetic cues to a prosodic phrase boundary.

With regards to the issue of whether it is the syntactic structure as a whole or the just the semantic representation of the incorrect local parse that persists, the response time analysis is more consistent with the latter account. This is because the formal account predicts the processing of semantic focus in the incorrect local parse to increase the activation level of that parse as a whole, making overall process of ambiguity resolution more costly. This is suggested by Ferreira & Lowder (2016) in a recent formulation of the Good Enough approach. Under the assumption that

"successful comprehension is about successfully integrating new information", they claim that the comprehension system specifically targets new and focused information and gives it an advantage in semantic processing (p. 239). If the parser prioritizes structures that contain focused information in the allocation of attentional resource during ambiguity resolution, the incorrect parse should become a stronger competitor to the globally correct parse, resulting in not only lower accuracy but also longer response time.

Again, contrary to this prediction, we find that the placement of the pitch accent neither facilitated nor impeded the process of syntactic ambiguity resolution. This finding is more consistent with the latter account and suggests that the pitch accent was simply interpreted with respect to each parse that was being considered before and after syntactic reanalysis. In other words, the presence of the attachment ambiguity and the preference for local attachment led the parser to first construct the late-closure analysis of the adjunct VP, within which the pitch accent was interpreted as contrastive focus in the verb accent condition. When the disambiguating material downstream forces reanalysis and the parser recovers the correct parse that closes the adjunct clause at the verb, the pitch accent on the verb is *reinterpreted* as a canonical, "default" accent, with respect to this new structure. In this sense, the success of syntactic reanalysis is independent from the success of semantic reanalysis, where the failure of the latter can affect the accuracy of comprehension without necessarily affecting the time course of syntactic structure building.

The semantic processor failing to discard the initial semantic commitments to the incorrect parse despite the success of syntactic reanalysis can be best captured in a serial parsing architecture that also allows some degree of incremental semantic processing. Sturt's (2007) proposal that syntactic reanalysis always succeeds but the semantic processor can independently fail to revise its initial analysis can account for our finding that the depth of semantic processing in the incorrect

parse does not necessarily interfere with the process of syntactic ambiguity resolution itself. We find that a deeper semantic processing in the incorrect parse can simply make the erroneous interpretation more resistant to semantic reanalysis at the point of disambiguation, similar to how the depth of processing has been found to affect the resolution of other types of ambiguities (Sanford & Sturt 2002). Interference of the initial misanalysis in memory, as opposed to the degradedness of the underlying syntactic structure, appears to drive misinterpretations in garden-path sentences.

4. Conclusion

In this study, we examined the depth of semantic processing as a factor influencing the semantic persistence effect that is independent of previously reported local coherence effects. Depth of semantic processing was examined through known effects of pitch accent on the processing of focus-related semantic meaning. Since pitch accents are interpreted with respect to the syntactic position of the accented word, we reasoned that a pitch accent on a temporarily syntactically ambiguous word could lead to an asymmetrical interpretation of focus-related semantic meaning between the possible analyses that are considered by the parser. Specifically, in garden-path sentences with local attachment ambiguity, we predicted that a pitch accent on the adjunct clause verb would be interpreted as marking contrastive focus only in the incorrect late-closure parse, facilitating deeper semantic processing for the associated erroneous interpretation and making it more likely to persist in memory and resulting in lower accuracy on comprehension questions.

In a speeded auditory comprehension experiment, we found that a nuclear pitch accent on the verb indeed lowers accuracy compared to the baseline condition with a nuclear pitch accent on the adjunct subject, which is positioned outside the region of ambiguity. This effect is significant after

controlling for previously reported local coherence effects from transitivity bias and semantic fit, consistent with our hypothesis that the depth of semantic processing is an independent factor affecting the strength of semantic persistence. A post-hoc response time analysis suggests against alternative explanations for the effect of pitch accent on accuracy that are based on the pitch accent facilitating or impeding the process of syntactic ambiguity resolution itself, as those accounts would predict a corresponding effect of pitch accent on response time. Combined, the data suggest that the pitch accent is immediately interpreted for its focus-related semantic meaning *in the garden path* before the syntactic ambiguity is fully resolved, strengthening the initial erroneous interpretation's resistance to revision in semantic reanalysis.

These findings on the effect of local coherence present a challenge to prior accounts of semantic persistence claiming that the locus of misinterpretations is the failure of syntactic reanalysis, where a sufficiently activated local structure can fail to decay past disambiguation and linger alongside the globally correct structure in the syntactic representation. Instead, results are more consistent with accounts in which syntactic reanalysis may successfully proceed without being accompanied by necessary semantic revisions that would require the semantic processor to discard the initial misinterpretation from memory (Sturt 2007). In sum, not only is semantic persistence about the failure of reanalysis as opposed to the ill-formedness of the underlying syntactic structure (Slattery et al. 2013), but more specifically, it appears to be about the failure to sever the semantic commitments to the initial misinterpretation and not about the failure to prevent the locally coherent structure from lingering in the syntactic representation.

These results are interesting not only for sentence processing research, but for prosody research as well. Studies in the real-time processing of pitch accents have traditionally centered round the interpretation of discourse-pragmatic meanings in structurally unambiguous sentences, often in the

context of referential ambiguity resolution. In fact, only very recently have pitch accents been studied in relation to syntactic structure building, such as Carlson & Tyler's (2018) study reporting that the location of pitch accent biases the height of adverbial-PP attachment when the attachment site is globally ambiguous. This study contributes to this new area of research on pitch accents by examining how they are interpreted when the syntactic position of the accented word is temporarily ambiguous but globally unambiguous. Our analysis posits an incremental semantic processor that makes stronger semantic commitments to the erroneous interpretation associated with the initial incorrect parse when it is marked for semantic focus via pitch accenting. This analysis is consistent with previous findings for the rapid and immediate processing of focus-semantic meaning from contrastive pitch accents, suggesting that "semantic processing" as talked about in prosody research and sentence processing research may be more interlinked than previously thought. This study serves as a yet another testament to the fruitfulness and necessity of synthesizing prosody and sentence processing in psycholinguistics research, as Janet Dean Fodor (2002) emphasized nearly two decades ago.

Acknowledgements

The authors would like to thank Dan Turner for the recording of the stimuli, and members of the Prosody & Speech Dynamics Lab and the Syntax Semantics and Sentence Processing Lab at Northwestern University for valuable feedback on the project.

Funding Information

This study was supported by the Northwestern University Office of Undergraduate Research Summer Undergraduate Research Grant and the Weinberg Academic Year Undergraduate Research Grant.

References

Abney, Steven T. 1989. A computational model of human parsing. *Journal of Psycholinguistic Research* 18. 129-144. DOI: https://doi.org/b10.1007/BF01069051

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. *arXiv*. arXiv:1506.04967

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1-48.

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and meaning*. 119–162. Berlin and New York: Mouton De Gruyter.

Beckman, Mary E. & Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (eds.), *Cognition and Language Development*. 279-352. New York: Wiley & Sons.

Birner, Betty J. & Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English*. Amsterdam: John Benjamins.

Boersma, Paul & David Weenink. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.06. http://www.praat.org

Büring, Daniel. 2016. *Intonation and Meaning*. Oxford: Oxford University Press.

Carlson, Greg. 1984. Thematic roles and their role in semantic interpretation. *Linguistics* 22(3). 259-280.

Carlson, Katy & Joseph C. Tyler. 2018. Accents, Not Just Prosodic Boundaries, Influence Syntactic Attachment. *Language and Speech* 61(2). 246-276.

Chodroff, Eleanor & Jennifer Cole. 2019. The phonological and phonetic encoding of information structure in American English nuclear accents. In *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Holland: Foris Publications. Reprint. 7th ed. Berlin and New York: Mounton de Gruyter, 1993.

Christianson, Kiel, Andrew Hollingworth, John F. Halliwell & Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42(4). 368-407.

Crain, Stephen & Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. In David R. Dowty, Lauri Karttunnen & Arnold M. Zwicky (eds.), *Natural Language Parsing*. Cambridge, UK: Cambridge University Press.

Crocker, Matthew, W. 1996. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Cutler, Anne, Delphine Dahan & Wilma van Donselaar. 1997. Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech* 40(2). 141-201.

Erickson, Thomas D. & Mark E. Mattson. 1981. From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior* 20. 540-552.

Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47. 164-203.

Ferreira, Fernanda, Karl G. D.Bailey & Vittoria Ferraro. 2002. Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science* 11(1). 11-15.

Ferreira, Fernanda, Kiel Christianson & Andrew Hollingworth. 2001. Misinterpretations of garden-path sentences: implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research* 30. 3-20.

Ferreira, Fernanda, Ellen F. Lau & Karl D. G. Bailey. 2004. Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science* 28. 721-749.

Ferreira, Fernanda, & Matthew W. Lowder. 2016. Prediction, Information Structure, and Good-Enough Language Processing. *Psychology of Learning and Motivation* 65. 217-247.

Ferreira, Fernanda, & Nikole D. Patson. 2007. The good enough approach to language comprehension. *Language and Linguistics Compass* 1, 71-83.

Fodor, Janet D. 2002. Psycholinguistics cannot escape prosody. In *Proceedings of the SPEECH PROSODY 2002 Conference*, Aix-en-Provence, France.

Fraundorf, Scott H., Duane G. Watson & Aaron S. Benjamin. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language* 63(3). 367-386.

Frazier, Lyn & Janet D. Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6(4). 291-325.

Frazier, Lyn. 1979. *On comprehending sentences: Syntactic parsing strategies.* Storrs, CT: University of Connecticut dissertation.

Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14(2). 178-210.

Frazier, Lyn & Keith Rayner. 1987. Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language* 26(5). 505-526.

Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers & Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37(1). 58-93.

Gibson, Edward & Neal J. Pearlmutter. 2000. Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research* 29(2). 231-240.

Gussenhoven, Carlos. 1999. On the limits of focus projection in English. In Peter Bosch & Rob van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Hare, Mary, Jeffrey L. Elman, Tracy Tabaczynski & Ken McRae. 2009. The wind chilled the spectators, but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science* 33. 610-628.

Ito, Kiwako & Shari R. Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Memory and Language* 58. 541–573.

Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2(1). 15-47.

Kjelgaard, Margaret M., & Shari R. Speer. 1999. Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity. *Memory and Language* 40. 153-194.

Kurumada, Chigusa, Meredith Brown, Sarah Bibyk, Daniel F. Pontillo & Michael K. Tanenhaus. 2014. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition* 133(2). 335-342.

Lau, Ellen F. & Fernanda Ferreira. 2005. Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes* 20. 633–666.

MacDonald, Maryellen C., Neal J. Pearlmutter & Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4). 676-703.

Marcus, Mitchell P., Donald Hindle & Margaret M. Fleck. 1983. D-theory: Talking about talking about trees. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*. 129-136.

Pickering, Martin J., & Matthew J. Traxler. 1998. Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24. 940-961.

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Roettger, Timo B, Tim Mahrt & Jennifer Cole. 2019. Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience* 34(7). 841-860.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75-116.

Sanford, Anthony J. & Patrick Sturt. 2002. Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Science* 6(9). 382-386.

Shafer, Amy J., Shari Speer, Paul Warren & David S. White. 2000. Intonational Disambiguation in Sentence Production and Comprehension. *Journal of Psycholinguistic Research* 29. 169-182.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7. 141-177.

Selkirk, Elisabeth. 1995. Sentence Prosody: Intonation, Stress, and Phrasing. In John A. Goldsmith (eds.), *The Handbook of Phonological Theory*. London: Blackwell.

Slattery, Timothy J., Patrick Sturt, Kiel Christianson, Masaya Yoshida & Fernanda Ferreira. 2013. Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language* 69. 104-120.

Stowe, Laurie A. 1986. Evidence for on-line gap-location. *Language and Cognitive Processes* 1. 227-245.

Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48. 542–562.

Sturt, Patrick. 2007. Semantic re-interpretation and garden path recovery. *Cognition* 105(2). 477-488.

Tabor, Whitney & Sean Hutchins. 2004. Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2). 431-450.

Tabor, Whitney, Bruno Galantucci & Daniel Richardson. 2004. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language* 50. 355–370.

Townsend, David J. & Thomas G. Bever. 2001. *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: The MIT Press.

Truckenbrodt, Hubert. 1995. *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Trueswell, John C., & Michael K. Tanenhaus. 1994. Toward a lexical framework of constraint-based syntactic ambiguity resolution. In Charles Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum Association.

Wagner, Michael. 2020. Prosodic Focus. In Daniel Gutzmann, Lisa Matthewson, Cecile Meier, Hotze Rullmann & Thomas E. Zimmermann (eds.), *The Wiley Blackwell Companion to Semantics*. Hoboken, NJ: Wiley-Blackwell.

Wagner, Michael, & Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25. 905-945.

Watson, Duane G., Michael K. Tanenhaus & Christine A. Gunlogson. 2008. Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*. *Cognitive Science* 32. 1232-1244.

Zehr, Jeremy & Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832