9

Pronouncing and comprehending center-embedded sentences

JANET DEAN FODOR

9.1 The cognitive basis for center-embedding difficulty

The extreme processing difficulty of doubly center-embedded relative-clause constructions played an important role in Bever's argument in the 1970s that the perceived acceptability of sentences is in some cases attributable not to the mental grammar but to "performance mapping rules." These performance rules reflect general cognitive strategies which can be observed at work in nonlinguistic perception also (Bever 1970).[1] Nowadays, such a conclusion would be hailed by linguists working in the Minimalist Program (Chomsky 1995a *et seq.*), whose aim is to minimize the richness of the "narrow faculty of language" by allocating more of the descriptive and explanatory burden to general cognitive mechanisms within the "broad faculty of language" (e.g., Hauser et al. 2002). But the linguistics of the 1970s was still steeped in the heady atmosphere of Chomsky's early arguments that language is unique, shaped by intricate grammatical principles not found in other cognitive functions. This suggested (though it didn't entail) that the processing mechanisms for language use would also be special, closely tailored to the unique formal properties of language structure. A significant embarrassment for this position was that it had no good *linguistic* reason to offer for the striking difficulty of the doubly center-embedded relative-clause construction (henceforth 2CE-RC for short), as in (1).

(1) The dog the cat the fox was chasing was scratching was yelping. (Bever 1970)

[1] Except where noted, references in this paper to Bever's work will be to Bever (1970). Additional evidence supporting this position was presented in a number of his subsequent publications.

This construction embodies the two quintessential formal properties of human language that are celebrated in generative linguistics: recursive hierarchical phrase structure and transformational operations (*wh*-movement in (1)). So the 2CE-RC construction could have been expected to provide a brilliant illustration of the virtuosity of naïve speaker-hearers, well equipped with the innate faculty of language. Clearly, it does not. It fails abysmally compared with innumerable other constructions that have multiple clauses and multiple transformations. These include examples such as (2)–(5), which people can parse with remarkably little difficulty, including the right-branching paraphrase of (1) in (5).

(2) Which of the papers did the secretary decide to file without bothering to read first?

(3) What are boxes easy to store in?

(4) If Susan wasn't crying because she'd lost her iPod, what was it that was bothering her?

(5) The fox was chasing the cat that was scratching the dog that was yelping.

Several attempts were made to identify a simple and intuitively plausible performance constraint that would hold the infinite generative power of the grammar in check at just two levels of center-embedding. But Bever was dismissive of two of the most prominent proposals. Miller and Chomsky (1963) had proposed that a perceptual principle may interrupt itself no more than once, but Bever rejected this as an arbitrary stipulation: why should one interruption be acceptable, as in (6), but not two or more?

(6) The dog the cat was scratching was yelping.

Fodor and Garrett (1967) identified the source of 2CE-RC difficulty as its structural density: the high ratio of the number of underlying sentential units to the number of words in the surface string. But Bever countered with an easily comprehended example, in (7), for which this ratio was even higher.

(7) The fox was chasing the cat scratching the yelping dog.

In place of these notions, Bever proposed a general cognitive "double-function" constraint, which he illustrated in visual processing as well as language. His Principle I is as follows:

> *Principle I: In a sequence of constituents x, y, z, if x has an internal relation Ri to y and y has the same internal relation to z, and x, y, and z are superficially identical, then the stimulus is relatively complex, due to y's double function in the perceptual strategy, Si.*

$$Si: x\ y \rightarrow x\ Ri\ y$$
$$si$$

In the case of the 2CE-RC construction the perceptual strategy in question is Strategy J, which extracts the internal (i.e., deep) relation between two adjacent NPs followed by a VP, as in "the cat the fox was chasing."

> *Strategy J: In . . . NP_1 NP_2 (VP) . . . sequence in the external structure, NP_1 is the internal object of an internal structure sentence unit of which NP_2 is the subject.*

Bever observed that Strategy J applies twice in the double RC structure (NP_1 NP_2 NP_3 VP_1 VP_2 VP_3). Hence there is difficulty due to Principle I, because one application of Strategy J (to NP_1 NP_2. . .) maps NP_2 as a subject while the other application (to NP_2 NP_3. . .) maps it as an object.

In addition to the range of supporting data that Bever presented (e.g., on double negations), what is satisfying about this explanation is that it does not have to *stipulate* that a major step up in parsing difficulty sets in between one center-embedded RC and two center-embedded RCs. The potential for confusion when a parsing strategy applies in different ways to the same item is intuitively clear. And as it happens, that occurs with double center-embedding of RCs but not with single center-embedding of RCs, or with double center-embedding of other clause types such as clausal complements to nouns. Gibson and Thomas (1999) give example (8), where an RC is embedded inside a noun complement, which they deem to be "quite processable." Strategy J does not apply twice in (8).

(8) The fact that the employee who the manager hired stole office supplies worried the executive.

In this chapter I offer yet another explanation of 2CE-RC difficulty which, like Bever's, has this property that the fact that the difficulty sets in suddenly at two degrees of embedding does not need to be stipulated but is an inevitable indirect consequence of the explanatory factors at work. My story, at its most basic, is that while natural language syntax thrives on recursion, prosodic phrasing does not,[2] and that a sentence cannot be efficiently parsed if it cannot be assigned a supportive prosodic contour. This is thus an explanation that is specific to language, unlike Bever's, but it falls within the broad faculty of language, since it concerns the interface between prosody and syntax (the "articulatory/perceptual" interface with

[2] This is compatible with the evidence that natural language prosody *permits* recursion; see discussion in section 9.4.3.

syntax; Chomsky 1995a and elsewhere). I will approach it in two steps: an early "phrasal packaging" account, then translated into a prosodic phrasing account.

9.2 The phrasal packaging account of parsing difficulties

Frazier and Fodor (1978) offered an account of the processing difficulty of center-embedded sentences within a proposed model of syntactic parsing (the so-called *Sausage Machine* model) which employed the rules of the competence grammar[3] but which, like Bever's explanation, also called on a general cognitive constraint (in our case, memory-related) to explain cases of differential processing difficulty. Before tackling center-embedded RCs, we worked through examples of several other tricky constructions (some from Kimball 1973).

(9) Joe bought the book that I had been trying to obtain for Susan.
(preferred: obtain for Susan)

(10) John read the note, the memo and the newspaper to Mary.
*(preferred: *newspaper to Mary)*

(11) John threw the apple that Mary had discovered was rotten out. / out of the window and into the rosebush.
*(preferred: *rotten out. ✓rotten][out of the window and into the rosebush.])*

We noted that the preferred readings of an ambiguous example such as (9), and the parsing difficulty of unambiguous examples such as (10) and the short form of (11), could be explained on the assumption that a first-stage parser divides a sentence into chunks (approximately six words, often a syntactic phrase but not always exactly), computing the internal structure of each; and then a second-stage parser follows along stitching the chunks together to form a complete structure for the sentence.

We proposed that the source of the first-stage packaging procedure was the need to limit the amount of working memory needed at any point in the parse. This idea of dealing with one portion of a sentence at a time was common in other prominent parsing models of the time. In Kimball's (1973) model and in Fodor, Bever, and Garrett's (1974) model, it was proposed that phrasal or clausal units are "shunted" out of the syntactic working memory

[3] Following Kimball (1973), Frazier and Fodor (1978) focused on phrase structure parsing and did not attempt to get to grips with the parsing of transformed sentences, which was addressed in Fodor (1978).

store as they are processed, to make room for subsequent units. Since the shunted units had at some point to be reintegrated, these were all, in that respect, "two-stage" models of syntactic parsing, and they all attributed the need for a two-stage design to memory limitations, prompted by Miller's (1956) famous treatise on a human memory span of seven plus or minus two units.

9.3 Phrasal packaging as prosodic phrasing: Late closure

While other aspects of the Sausage Machine (henceforth SM) model survived into subsequent research, the first-stage "preliminary phrase packager" was not widely embraced, and little or no experimental work tested its specific predictions which therefore remained dangerously perched on mere intuitions.[4,5] But I resurrected phrasal packaging later, in the new guise of prosodic phrasing (Fodor 1998b), as a potential explanation for an apparent exception to the principle of Late Closure (Frazier 1978; also called Recency by Gibson 1991). Late Closure (LC) favors *local* attachment of an incoming constituent, structurally close to the words that immediately precede it in the input string; this entails *low* attachment in a right-branching construction. While an LC tendency had been widely observed in constructions tested in English,[6] Cuetos and Mitchell (1988) notably found a preference for high attachment of relative clauses in Spanish, an apparent violation of LC. Familiar examples are shown in (12) and (13).

(12) Someone shot the servant of the actress who was on the balcony with her husband.
(English: preferred RC attachment to lower noun *actress*)

[4] Wanner (1980) presented an outright counterexample to our claim that the packaging process was the *sole* source of the Late Closure strategy. Experimental evidence is accruing that there is *both* a syntactic LC tendency, very early online, and also a later packaging (prosodic) effect; see discussion and references in Augurzky (2006) and Maia et al. (2006).

[5] Those were days in which one could go far in spinning a general claim on the basis of one's own intuitions of processing difficulty—not even the aggregated intuitions of a group of native judges, let alone behavioral data such as response times and error rates, or ERPs. Empirical standards have moved onward and upward in the twenty-first century, and we do now have some experimental evidence (Fodor and Nickels 2011) supporting the role of phrase lengths in facilitating syntactic parsing of 2CE-RC constructions, as predicted by Frazier and Fodor. See section 9.5.

[6] A notable exception was PP attachment to a preceding noun or to the verb, as in *She hit the man with the stick*, which generally favors high attachment in English. Possible explanations included an argument/adjunct contrast, or dominance of Minimal Attachment over Late Closure. There may be a prosodic explanation, but this remains unresolved; see discussion in Shafran (2011).

(13) Alguien disparó contra el criado de la actriz que estaba en al balcón con su marido.
(Spanish: online preferred RC attachment to higher noun *criado*; garden path at *con su marido*)

Importantly, though, Spanish shows the same general LC preference as other languages do, for ambiguities other than RC attachment. Subsequent work by Igoa (1995) showed a preference for attachment of a prepositional phrase into a subordinate clause rather than a main clause in Spanish. So the high RC attachment preference is apparently an aberration even within-language. A number of clever and plausible explanations were proposed for this anomaly, but all ran into empirical difficulties in face of experimental data that were being gathered from a variety of languages. (See Lovrić 2003: chapter 3 and Augurzky 2006: chapter 2, for summaries and references.)[7]

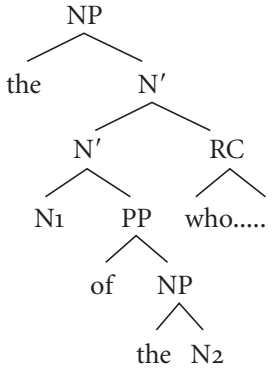
An explanation of high RC attachment in terms of the SM packaging mechanism suggested itself, since a long (i.e., "normal" length) RC would be expected to be packaged separately from the preceding double-noun complex. The second-stage parser would then join the two chunks together as simply as possible. We may assume that for maximum efficiency of the two-stage design, the second stage cannot easily look *inside* a package it had previously received from the first stage processor.[8] The result of packaging the RC separately from the nouns would thus be high RC attachment, as in tree (14a), overriding the usual LC preference for low attachment, as in (14b).[9] In both cases the relative pronoun is linearly adjacent to the second noun, but in high attachment it is far from that noun in the tree structure, as an indirect consequence of the two words being in different packages.

[7] There is one nonprosodic explanation which cannot be discounted on the basis of current data. This attributes the different parsing preferences across languages to differences in the information structure of sentences (Hemforth et al. 2006). Influences of prosody and information structure can be difficult to dissociate.

[8] In the 1978 paper we did permit one package to be inserted into a preceding one, since this seems to be necessary for building normal syntactic trees, but it could still be the case that simply concatenating successive packages is the least-effort option for the parser. See Wagner (2010) and discussion in section 9.4.3.

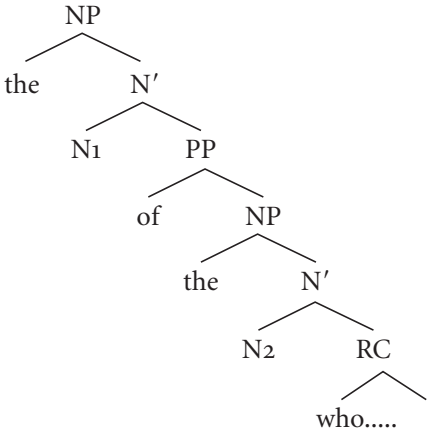
[9] Syntacticians would posit more elaborate structural representations than this, but many details can be dispensed with here. For displaying the syntactic structure relevant to prosody assignment a simpler traditional mode of representation generally suffices; see also (16).

(14) a.



HIGH ATTACHMENT
(preferred in Spanish)

b.



LOW ATTACHMENT
(preferred in English)

A contrast between high RC attachment and low PP attachment in Spanish would follow from the packaging-relevant fact that RCs but not PPs are clausal units, and that RCs tend to be longer than PPs—both of which could make the first-stage parser more inclined to make a packaging break before an RC than before a PP. The packaging approach also suggested that among RCs, short RCs should be less inclined to attach high than long RCs. This was confirmed by Fernández (2003), who found less high attachment for short than long RCs in Spanish; and a comparable effect of RC length has been found in other languages tested since.

These within-language observations concerning the variability (category-based, length-based) of preferred modifier attachments thus provide some encouragement for the idea that they originate in "first-stage" packaging operations. However, any such explanation must confront the fact that high RC attachment is more frequent in Spanish than in English. Whereas Late Closure predicted low attachment in both (all) languages, packaging considerations would apparently predict high attachment of long RCs in both languages—unless the size parameters of packages (how long counts as long?) could differ across languages. But clearly, if packaging divisions are made in order to avoid short-term memory overload, their placement should not differ across languages. (It would not do to suggest that people who speak Spanish have shorter memory spans than people who speak English.) But now, suppose instead that the packaging mechanism is not a memory-saving device but is the prosodic component of the grammar at work, dividing word

strings into prosodic phrases for pronunciation. On this assumption, many things fall into place.

Cross-language differences are explicable and even expected. Phonological studies of prosodic phrasing had long observed that though it is roughly congruent with syntactic phrasing it is also open to other influences, such as phrase length (Selkirk 1984; Nespor and Vogel 1986). Cross-language prosodic differences had also been noted. Selkirk and Shen (1990) argued that alignment constraints on prosodic phrasing are parameterized; later in an Optimality Theory framework (Selkirk 2000) it was proposed that the constraints are ranked differently in the grammars of different languages. Thus it was well established that languages can differ to some extent in where they place their prosodic boundaries. Therefore, the cross-linguistic differences in syntactic parsing preferences observed by Cuetos and Mitchell might be attributed to a stronger tendency in Spanish than in English to place a prosodic boundary at the beginning of an RC. Intuitive judgments tended to support this idea, though the subsequent experimental data suggested that the difference is in the quality rather than the frequency of a boundary: in Spanish there is a pre-RC rise in fundamental frequency while in English there is a fall (Fernández 2005), making the boundary more prominent in Spanish than in English.

The explanatory potential of the linguistic relationship between preferred syntactic phrasing and prosodic phrasing was thus promising, but it had one further hurdle to clear. The data on attachment preferences in parsing came largely from experiments in which sentences were visually presented, for silent reading by participants. Similarly, the examples that originally motivated the SM packaging mechanism were usually presented to informants in written form, and probably read silently. Therefore, in order for prosody to supplant working memory as the explanatory basis for packaging effects in syntactic parsing, it had to be shown that prosodic phrasing is projected onto sentences during silent reading and can affect syntactic parsing in ways parallel to the effects of overt prosody in spoken sentences. This is the message of the Implicit Prosody Hypothesis (IPH; Fodor 1998b, 2002).

Substantial support for this hypothesis has accumulated over the last few years and is too extensive to report in detail here. Experimental studies have been conducted on a variety of structures in addition to RC attachment (e.g., AP attachment in Japanese: Hirose et al. 1998; Hirose 1999, and clause-boundary ambiguities in Japanese and Korean: Hirose 2003; Hwang and Steinhauer 2011), with results showing parallel syntactic attachment preferences cued by overt prosodic patterns in listening and by assumed default prosodic patterns

in silent reading. On RC attachment there are experimental data linking attachment and prosody for Bulgarian (Sekerina et al. 2004), Croatian (Lovrić 2003), Egyptian Arabic (Abdelghany, 2010), French (Pynte and Colonna 2000), German (Augurzky 2006), Hebrew (Shaked 2009), Jabberwocky (Wijnen 2004), Japanese (Jun and Koike 2008), Portuguese (Brazilian and European; Maia et al. 2006), Spanish (Maia et al. 2006), and Russian (Sekerina 2003), with manipulation of a variety of factors including RC length, noun length, and PP vs RC modifiers among others. For additional references see Augurzky (2006: chapter 3). One finding of interest, for example, is that a small lexical difference in a sentence can affect the prosodic phrasing, and a difference in syntactic parsing ensues. Specifically, omitting the preposition between the two nouns in the RC ambiguity construction (Lovrić 2003, for Croatian; Augurzky 2006, for German) reduces the probability of a prosodic break between them, which increases the probability of a break between the second noun and the RC, yielding (as predicted) more high RC attachment than when the preposition is present.

In short, the general idea has held up. Within- and cross-language differences in attachment tendencies in parsing have repeatedly been found to correlate with within- and cross-language differences in preferred prosodic phrasing.[10] And this occurs even in silent reading, where there is no prosody in the physical stimulus at all. Apparently, the reader mentally imposes a prosodic contour on a word string seen on the page or screen, and then at least sometimes treats that prosody as a cue to syntactic structure. This reconciles cross-language differences in parsing RC attachment with the strong assumption that the parsing mechanism is biologically shaped, hence universal. We might hope, therefore, to be able to resolve other stubborn problems in parsing theory in a similar way by reference to mentally imposed prosody—problems such as a disconnect between what the syntactic component "ought" to find easy (e.g., recursion, transformational operations) and what the processing routines in fact find easy.

Frazier and Fodor (1978) actually made a start on a packaging explanation of the difficulties posed by multiple center-embedding. In light of Bever's

[10] Contrary to the general trend, two recent papers have reported a dissociation between prosodic boundaries and syntactic attachment preference in English RC attachment constructions. However, both ascribe this at least in part to the unnaturalness of the prosody elicited in their reading-aloud tasks. Bergmann et al. (2008) report data showing a high proportion of Intonational Phrase (IPh) breaks preceding the RC, which is unexpected in English. But it is to be noted that 40 percent of the utterances with an IPh boundary there *also* had an IPh boundary before the preceding NP2, and apparently some of those had a break before the previous NP1 as well. This strongly suggests that readers were applying an unnaturally slow and cautious prosody, as noted also by Jun (2010).

(1970) paper, it was clear that any credible parsing model would need to have something illuminating to say about 2CE-RC sentences, so this was an important challenge for the explanatory power of SM packaging. Frazier and Fodor undertook to show that the 2CE-RC construction poses an almost insurmountable problem for their proposed two-stage parser. The overarching hypothesis was that the linear sequence of packages created by the first-stage parser would correspond very poorly to the heavily nested syntactic tree structure that needs to be built in the second stage. That is: the 2CE-RC construction is inordinately difficult to parse because it is the construction in which the surface chunking and the true syntactic structure are most at war. In the next section I consider whether it is possible to maintain and even improve on this account of the severe difficulty of center-embedded sentences, once the old SM packages are recast as prosodic phrases.

9.4 Prosodic phrasing and center-embedding difficulty

Frazier and Fodor (1978) estimated the optimal length of an SM package to be about six words. Shorter and longer packages were possible if called for, but shorter would require chopping a sentence into more chunks than necessary, thus making more work for the integrative processes of the second-stage parser, while longer packages were in danger of exceeding some sort of assumed maximum capacity limit for the first-stage parser. Now, looking back, one can see that a string of approximately six English words is compatible with the typical length of a phonological phrase (an *intermediate phrase*, ip; also called *major phrase*). The BinMin and BinMax constraints of Selkirk (2000) specify the optimal length of an ip as two prosodic words, though these are not hard constraints but can be overridden by other factors. A prosodic word may contain more than one lexical item, since an unstressed function word such as a determiner is grouped together with an adjacent open-class word.[11] Intuitive judgments (in the absence of hard data) suggest that English tends towards longer phrases, while other languages such as Hebrew favor shorter phrases (e.g., for sentences with an odd number of prosodic words, in which at least one ip must consist of either one or three

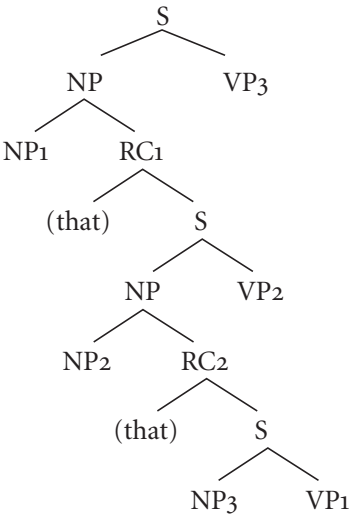
[11] Frazier and Fodor (1978: 293) considered the possibility that package length should be measured in some units other than lexical words ("syllables or morphemes or conceivably in terms of time") but did not mention prosodic words. We also speculated that the capacity of a package might be measured in terms not just of length but also syntactic complexity (cf. Fodor and Garrett's 1967 proposal, in section 9.1), so that some notion of "weight" rather than length is what matters. This could be true also for the prosodic phrasing version of the proposal, but I will not address that here.

prosodic words, English often opts for a length of three, while other languages might chop that into 1 + 2, or 2 + 1). Therefore, one natural prosodic phrasing of sentence (15) would take the first six words to be one prosodic phrase, dividing up the string as NP1 NP2 NP3 || VP1 VP2 VP3.[12] (I use the || notation to mark ip boundaries.)

(15) The woman the man the girl loved met died.[13]

This prosodic phrasing runs completely counter to the correct syntactic structure for (15), which is [NP1 [NP2 [NP3 VP1] VP2] VP3], shown in (16).

(16) Syntactic structure of a 2CE-RC sentence[14]



The phrasing NP1 NP2 NP3 || VP1 VP2 VP3 is compatible with a common misparse of such examples into two coordinate sequences [[NP1 NP2 NP3] [VP1 VP2 VP3]], despite its ungrammaticality in English (Blumenthal 1966). However, we need not dwell on this since the coordination misanalysis can be averted by inserting relative pronouns as in (17), or by blocking coordination

[12] This grouping might be abetted in (15) by the repetitive syntactic form of the constituents, as Bever noted.

[13] In the examples that follow, I sometimes include a relative pronoun (or complementizer) *that*, and sometimes do not. The naturalness of its presence or absence interacts to some extent with the prosodic phrasing: Fox and Thompson (2007) show that the probability of an overt *that* increases following a prosodic boundary.

[14] In (16), as in (14) above, the structural representation is simplified for present purposes. In particular, DP structure is omitted. A (restrictive) relative clause is standardly grouped with its head noun to the exclusion of the determiner, but this is not an issue for prosodic phrasing since a determiner and adjacent head noun are typically combined into a single prosodic word.

either morphologically as in (18), or semantically as in example (19) from Bever (1970) to be compared with (1) above.

(17) The woman who the man that the girl loved met died.

(18) The woman someone I loved met died.[15]

(19) The dog the destruction the wild fox produced was scaring will run away fast.

As Frazier and Fodor noted, this can make the beginning of the sentence easier to parse (much easier for (18) but less so for (17) and (19) in my judgment), but a temptation to mispackage the three VPs remains. (We took this as evidence that the second package is computed locally, without access to the internal structure of preceding packages, because if the internal structure of the first package remained accessible, it ought to be clear to the parser that each of the three correctly structured NPs needs to be associated with one of the three following VPs.) Taking a more positive tack, Frazier and Fodor then considered what kind of packaging would facilitate recognition of the correct syntactic structure. In retrospect, it turns out to be very much as a prosodic account would predict, and from here on I will frame the discussion in prosodic terms.

The SM phrasal packager created a sequence of medium-sized units, which the second-stage parser connected together in relatively simple fashion into larger clausal and sentential units. Translating from SM packages into intermediate prosodic phrases, the prosodic parser creates a sequence of ips, which are then connected into an Intonational Phrase (IPh) for the sentence as a whole. Maintaining the SM assumption that once a package is created, it is inefficient for the second-stage parser to have to attach new material inside it (cf. Kimball's *Fixed Structure* principle), I will make the strong assumption here that the ips are (optimally) connected as syntactic sisters. This is what gives the impression that a complex syntactic structure must be "flattened out" in order for a congruent prosody to be assigned.[16] (See discussion in section 9.4.3 of a possible multilevel phrasing.)

The task of the prosodic phrasing component, proceeding (as I will assume) from left to right through a sentence alongside the syntactic parser,

[15] Gibson and Thomas (1999), noting that "doubly nested RC structures are easier to process when a first or second-person pronoun is in the subject position of the most embedded RC," attribute this observation to Bever (1970). The proper reference may be Bever (1988: 124). Gibson and Thomas explain this amelioration by pronominal NP_3 in terms of the number of new discourse referents introduced. Bever's Principle I accounts for it, since the principle applies only in case of superficial similarity of the NPs, which the pronoun eliminates. As will become clear below, there is also a very natural prosodic explanation.

[16] See Chomsky and Halle (1968) and Langendoen (1975) on syntactic readjustment rules which "flatten" the structure of *house that Jack built* examples, with recent discussion by Wagner (2010).

is thus to identify a sequence of successive intermediate phrases which ideally are of roughly optimal prosodic length, and of roughly equal prosodic length,[17] while doing the least possible damage to the syntactic structure in the process. In an ideal case, the imposed prosodic groupings provide transparent guidance for syntactic parsing and comprehension. This is a tall order for multiply center-embedded RC constructions, where the prosodic priorities are at odds with the syntactic priorities. Taking the "least damage" condition seriously implies that elements should not be grouped together in an ip unless they constitute a coherent syntactic unit. For example, *the woman the man* would not be a syntactically helpful ip, because the two NPs would later have to be prised apart, or else added to in order to create a complete syntactic phrase (e.g., *the woman the man admires*). I will call this the *syntactic coherence* condition.[18] The obverse of this is that elements that do constitute a syntactic unit should be grouped together prosodically where possible (cf. the Wrap constraint of Truckenbrodt 1999). For sentence (15), for example, an ip consisting of the RC *the girl loved* is helpful, but not so much as an ip containing the whole NP *the man the girl loved*, or even the larger RC *(that) the man the girl loved met.* The larger the syntactic phrase that can be packed into an ip, the less combinatory syntactic work there is to be done when the ips are combined. However, the rhythm-based constraints of optimal length and balance, which the prosodic phrasing must also satisfy, may impede satisfaction of these prosody–syntax congruence conditions. In fact, as I will show, there are many patterns of phrase lengths that oppose syntax–prosody alignment in a 2CE-RC construction, and very few that cooperate with it.

9.4.1 *The battle between length and syntactic coherence*

In English a 2CE-RC sentence contains a minimum of six words (three subjects and three predicates, e.g., *Children teachers parents trust praise thrive*) and usually more (see examples above and below). It is therefore unlikely to be phrased as a single ip (two or three prosodic words). It could be divided into two ips in several different ways, as shown in (20) for the example (15) above. Underlined sequences violate the syntactic coherence condition: the indicated prosodic phrase does not constitute a syntactic phrase.

[17] This is a "balance" or more generally a "uniformity" condition on prosodic phrasing; see Gee and Grosjean (1983), Ghini (1993), Fodor (1998b), Sandalo and Truckenbrodt (2002).

[18] This relates to Selkirk's semantically defined (1984) *Sense Unit Condition* (since abandoned; Selkirk 2005), but I cannot dwell on exact comparison here.

(20) a. The woman || the man the girl loved met died.
b. The woman the man || the girl loved met died.
c. The woman the man the girl || loved met died.
d. The woman the man the girl loved || met died.
e. The woman the man the girl loved met || died.

Only the last of these, in which the prosodic grouping parallels the NP–VP division of the matrix clause, is congruent with the syntactic phrasing. This example (20e) is wildly imbalanced rhythmically, so a reader is unlikely to think of phrasing the sentence this way. But Frazier and Fodor showed that a sentence can become easier to parse if some constituents are lengthened so that the resulting lengths are more conducive to a phrasal packaging suited to the syntactic structure. In fact, this turns out to be a useful tool for investigating syntactic constraints on prosodic phrasing without interference from rhythmic constraints. The strategy is to create examples in which phrase lengths would be rhythmically acceptable for a syntactically congruent prosodic phrasing and then see whether that phrasing is acceptable; if not, it must be a syntax–prosody alignment constraint that rules it out. In the case of (20e) we would need to shorten the overlong matrix subject NP and lengthen the one-word VP to balance it. This is done in (21), which intuitively is both more pronounceable and more comprehensible than (20e), even though it is longer overall.

(21) The woman someone I love met || died on a cruise ship in Maine.

However, (21) makes use of the pronoun maneuver, for which there are possible nonprosodic explanations (see footnote 15 above). For any 2CE-RC sentence that has three full lexical NPs in addition to the two verbs in RC1, it is very difficult to compress the matrix subject NP sufficiently to fit it into a single ip.

Thus, except in fortunate cases like (21), the primary syntactic division is not suitable as the one and only prosodic division. What could the parser do instead? Cutting down one more syntactic level, between NP1 and the RC that modifies it, keeps as much of the phrase structure intact as possible (see the tree diagram (16)), while breaking up the overly long subject into two ips.[19]

[19] The sentence could be prosodically divided into three ips in ten different ways as shown in (i) for example (20).

(i) a. The woman || the man || the girl loved met died.
b. The woman || the man the girl || loved met died.
c. The woman || the man the girl loved || met died.
d. The woman || the man the girl loved met || died.

There would be three prosodic phrases: NP1, RC1, and VP3.[20] For sentence (15) this is still not ideal rhythmically, because the NP1 and VP3 are shorter than RC1 (*The woman || (that) the man the girl loved met || died*), but it is not impossible, at least if NP1 and VP3 are pronounced with heavy stress to balance the weight of the RC1.[21]Adjusting the proportions of words per phrase also helps, as expected. In (22) the outer constituents, NP1 and VP3, have been lengthened to balance the weight of the RC1 in the middle. Although the sentence as a whole is now longer and more complex compared with (20e), syntactic parsing becomes (at least intuitively) quite natural.[22]

(22) The elderly Frenchwoman || that the man the girl loved met || died last year in Maine.

I will argue that this three-way prosodic phrasing, NP1 || RC1 || VP3 as in (22), is the most facilitative for syntactic parsing, as long as it can be achieved—crucially, as long as the constituent lengths and natural stress patterns for the word string permit it.[23] Thus I predict that a 2CE-RC construction is easiest to

e. The woman the man || the girl || loved met died.
f. The woman the man || the girl loved || met died.
g. The woman the man || the girl loved met || died.
h. The woman the man the girl || loved || met died.
i. The woman the man the girl || loved met || died.
j. The woman the man the girl loved || met || died.

Of these, all except (d) creates at least one noncoherent phrase. Note that (d), which is in principle acceptable, would be improved by insertion of *that* after the prosodic break at the beginning of RC1, for the reason in footnote 13.

[20] Though this three-phrase division clearly can assist the rhythmic alignment of prosodic phrasing with syntactic structure, it presupposes the legitimacy of a prosodic break between NP1 and RC1, which needs some justification. Syntactically, a restrictive relative clause is generally regarded as bracketed together with the head noun that it modifies. However, at least in English an RC can be separated prosodically from its head when there is some rhythmic reason for doing so. For instance, a long NP1 can trigger a following break even in sentences without center-embedding such as (i).

(i) The very youngest ERP laboratory assistant || that Professor Tomlinson was prepared to hire || died on a cruise ship in Maine.

Wagner (2010) argues that when a RC is prosodically separated from its head, it is also syntactically separated from the head; see discussion of RC extraposition in section 9.4.2.

[21] RC1, which is comprised of two clauses, is also syntactically more complex than NP1 or VP3, which may contribute to the relevant notion of prosodic weight; see footnote 11.

[22] Frazier and Fodor showed that not just any added length facilitates parsing, but only length in the outer phrases, not the inner one (with overall sentence lengths matched). For instance, the phrase lengths in example (34) in section 9.4.3 facilitate processing while those in (35) impede it.

[23] Though I cannot substantiate it in this paper, I believe stress patterns can be as relevant to prosodic weight in English as word count is. There are contextual differences with respect to the likelihood of a word bearing phrasal stress (stress on new content, not on old), and also inherent lexical differences (Ladd 1986). For instance, the verb *loves* in *the woman that the man the girl loves met* is more susceptible to reduction than the verb *hates* is (*the woman that the man the girl hates met*),

parse (indeed, is quite *easy* to parse) if the NP1 and VP3 are each heavy enough to stand alone, and the whole RC1 (containing NP2, NP3, VP1, and VP2) is light enough to not need dividing. The first condition is not too hard to satisfy because, as noted, a too-short phrase can be pronounced with heavy stress (or with a following pause, or just very slowly) to give it extra weight. The second condition is more challenging, because cramming too much material into one prosodic phrase has practical limits; it can take serious effort to pack up all of RC1 as a single ip. It is a feat unlikely to be achieved, by however expert a reader, on a first reading without any preview. However, once a reader gets the hang of it, it is impressive how well it can be done.[24] English (though perhaps not every language; this needs to be looked into) offers several techniques of phonological compression that can be brought to bear. Because English is a stress-timed language, it is primarily the stressed syllables that count for prosodic weight; the unstressed syllables in between can be much reduced if needed. Also, a tendency toward alternating stress patterns means that even full lexical items may receive reduced stress. For example, the two nouns and two verbs in the RC1 (*the man the girl loved met*) of example (22) are not all produced with equal stress—or if they are, then they do not fit into one ip.

My claim that the three-way prosodic pattern NP1 || RC1 || VP3 is the most supportive for the parser is based on the observation that dividing a 2CE-RC sentence into even smaller prosodic chunks is counterproductive. Why should that be? If it is effortful to package up a long RC1 as a single ip, why shouldn't the parser, in projecting a prosodic contour onto the written sentence, simply make one more prosodic cut at the next level down in the syntactic tree? That would divide the subject and verb phrase of RC1 (see (16)), creating the four-way pattern NP1 || NP2 NP3 VP1 || VP2 || VP3.[25] Applying our operational tactic, we adjust constituent lengths to be compatible with this phrasing, so that it becomes possible to judge how well it assists syntactic processing once

which can affect the likelihood that the RC *that the man the girl loves/hates met* will be grouped into a single prosodic phrase.

[24] Tom Bever has offered a triple embedding which can also be pronounced with the three-phrase prosody—given sufficient determination, plenty of look-ahead, and the aid of a pronoun in the middle: *The pictures || that the reporter everyone I met trusts took || showed that the fire was set by an arsonist.* In fact even the first break here can be omitted.

[25] I am assuming here the impossibility of making a cut between *that* and the remainder of the RC, since the *that* would be absorbed into a prosodic word with the following subject. There are ten potential prosodic analyses into four ips. Of these, the analysis NP1 || NP2 NP3 VP1 || VP2 || VP3 is the only one that does not group together items that do not constitute a possible syntactic phrase. The fact that the four-way analysis obeys this prosody–syntax interface condition makes it even more interesting that it is not a syntactically cooperative phrasing of the sentence.

rhythmic needs are met. Accordingly in (23), keeping all else constant, we lengthen VP2 (which was very short in (22), just *met*) so that it can comfortably stand alone.

(23) The elderly Frenchwoman || that the man the girl loved || met on a Mediterranean cruise || died last year in Maine.

By extreme squeezing, the parser might pack all of *that the man the girl loved met on a Mediterranean cruise* into a single prosodic phrase, but rhythmically the sentence would more naturally be phrased as shown in (23). However, this rhythmically more comfortable phrasing does not make the sentence easier to parse syntactically. The VP2 || VP3 sequence in (23) (*met on a Mediterranean cruise* || *died last year in Maine*) is intuitively awkward despite its supportive constituent lengths, making the latter part of the sentence more difficult to process than in (22).[26] It is not fair, of course, to compare sentences that differ in overall length, as (22) and (23) do. But other examples above and below show that total sentence length is not the primary factor. What matters is how that length (or weight) is distributed. To confirm that here, (23) can be compared with (24), which has as many words as (23) does but its distribution of phrase lengths is compatible with the prosodic phrasing NP1 || NP2 NP3 VP1 VP2 || VP3 and intuitively it is easier to process than (23). (See also (34) and (35) in section 9.5.)

(24) The elderly and eccentric French doctor || that the man the girl loved met || died last year in northern Maine.

Thus we arrive at the conclusion that prosodic facilitation of the 2CE-RC syntactic structure is effective only for sentences whose phrase lengths fit the three-phrase analysis, with RC1 in its entirety as the middle phrase. Once VP2 escapes from RC1 and is phrased separately, its structural role in the sentence becomes harder for the parser to grasp. This calls for explanation, and especially so because standard alignment constraints (Align_{R}XP; Selkirk 2000) require an ip boundary at the end of a relative clause; thus a boundary could be expected between VP1 and VP2 (and another between VP2 and VP3),

[26] Frazier and Fodor suggested that the less than fully helpful four-package analysis as in (23) was the best division the human parser could be expected to achieve, on grounds that it would be risky for the parser to embark on packaging up the complete RC1 in case it found at the last moment that it couldn't squeeze in the last word or two, creating worse problems. We had in mind the parsing of novel text online, by readers with no prior familiarity with the content and only the usual limited amount of look-ahead, so that the parser would not be able to anticipate where its next opportunity for a package break would be. We have yet to test whether, even when phrase lengths are cooperative, it is true that the optimal three-phrase prosody is not often achieved without prefamiliarity with the text.

reinforcing the pressure from phrase lengths in typical examples to split off VP2 as a separate prosodic phrase.

9.4.2 *Why is separation of VP2 unhelpful?*

For a properly explanatory account, we must understand why the four-phrase analysis is less conducive to comprehension than the three-phrase analysis is. A straightforward account might be that, as noted, the more lower-level phrasal packages there are, the more effort the parser must expend in stitching them back together at the integrative stage; this might transgress some optimal trade-off of workload between the two stages. But that may not be the only factor at work in the present case. In an example like (23) the second ip is not a complete constituent. It is comprised of a relative pronoun/complementizer (null or overt) and a subject NP, which do not stand in any grammatical relation to each other because the predicate that would connect them is missing. Suppose the first-stage parser is able to build a relative-clause structure over these words, with an empty slot for the missing VP2 (as it might for a single-embedding example such as *The elderly Frenchwoman* || *that Ellen's favorite cousin Timothy* ||...). Then, when the VP2 is processed next, the second-stage parser cannot connect it as a sister to that preceding unit (the incomplete RC1). Instead, the VP would need to be inserted internally to the RC, requiring the second-stage parser to delve inside the previous ip package—which, by hypothesis, is just what the second-stage parser does not like to do.

Let us take this line of thought one step further. Suppose that the second-stage parser, which is disinclined to embed VP2 into RC1, instead does what comes more naturally to it: it simply concatenates them. That is, it attaches the new unit (VP2) as a sister to the previous unit (the incomplete RC1). That would yield the syntactic phrasing [NP1] [incomplete RC1] [VP2]..., which makes VP2 also a sister to NP1. However, the only VP that can legitimately be a sister to NP1 is the matrix VP, i.e., VP3. The parser would thus have inadvertently put VP2 into the tree position reserved for VP3. When VP3 is encountered next in the word string, it must oust this cuckoo VP2 which is usurping its position—but then the VP2 is left with no attachment site at all. Since it cannot fit into the structure, VP2 is in danger of being ignored and forgotten by the parser. This offers an explanation for the "missing VP illusion," in which readers judge a 2CE-RC sentence with VP2 omitted to be equally acceptable or more so than the full sentence with VP2 present. For example, in a study by Gibson and Thomas (1999), readers

judged examples like (26) to be as acceptable as those like (27), even though (27) is grammatical and (26) is not.[27]

(26) *The ancient manuscript that the graduate student who the new card catalog had confused a great deal was missing a page.

(27) The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library was missing a page.

One last piece of this explanation needs to be set in place. My tale about why VP2 is often overlooked by the parser is based on the assumption that once VP2 has been packaged as a separate prosodic unit, it is difficult for the second-stage parser to reintegrate it where it belongs inside RC1. Yet the facilitative three-phrase prosodic analysis, as in (22) above, apparently does not suffer from any such problem in reintegrating a separately phrased RC1 into the matrix subject NP. Thus, my account is committed to the claim that dividing RC1 from NP1 does *not* overstress the integrative processes of the second-stage parser, even though a restrictive RC modifier is standardly analyzed as subordinate to its head. I propose that this is because the grammar of English allows a noun head and a restrictive RC that modifies it to be properly integrated semantically even when the surface syntactic tree does not display that relationship between them. RC extraposition, as in (28), is not uncommon in English. There are even more elaborate examples, such as (29), which has two RCs both of which are undeniably restrictive and both of which are undeniably extraposed.

(28) The children e_i were weeping RC_i[who the principal had scolded].

(29) Nobody e_i puts anything e_j into this sink RC_j[that would block it] RC_i[who wants to go on being a friend of mine].

Wagner (2010), following Chomsky and Halle (1968), notes that RC extraposition applying string vacuously would explain the presence of prosodic breaks preceding the RCs in the famously puzzling case of (30), a right-

[27] Gibson and Thomas's materials have phrase lengths that are not rhythmically supportive of the correct parse. The account presented here clearly predicts that when phrase lengths are compatible with a nonseparated VP2, the missing VP2 illusion does not occur. It is worth noting that the Gibson and Thomas materials have strict selection restrictions pairing up the NPs with the VPs (e.g., the student can study, but neither the manuscript nor the card catalog can), but these evidently did not succeed in protecting against the missing VP2 illusion. The simple example (i) below was given by Frazier and Fodor to illustrate the powerlessness of semantic selection to overcome the difficulty of 2CE-RC sentences when phrasal packaging is unhelpful: (i) The snow the match the girl lit heated melted.

branching structure in which there would otherwise be no syntactic positions appropriate for a prosodic boundary:

(30) This is the cat that caught the rat that stole the cheese.

String-vacuous extraposition would raise an RC out of its subordinate position to become sister to its head, creating what is essentially a paratactic syntactic construction, but with some sort of anaphoric link between the RC and the head to encode their true semantic relationship. This would provide the syntax–prosody interface with an adjusted syntactic structure to which the prosodic structure can be aligned. In 2CE-RC sentences such as (22), where NP1 and the RC that modifies it cannot be prosodically phrased together for length reasons, RC extraposition would allow the RC to be moved out and adjoined as a sister to NP1, yielding prosodic phrasing of them as two separate units without detriment to the semantic relation between them.

Following Wagner's lead here, we may find an explanation for why the four-phrase prosodic analysis of 2CE-RC sentences, as in (23), does not further facilitate syntactic parsing. There is no comparable extraposition operation in English that could separate VP2 prosodically from the relative clause that it belongs to. In contrast to RC extraposition in (28) and (29), extraposition of a finite VP as in (31) is strongly ungrammatical—either because finite VPs do not extrapose at all, or because they cannot extrapose out of relative clauses, which are extraction islands.[28]

(31) *The children RC[who the principal e_i yesterday] VP_i[had scolded] were weeping.

Therefore there is no legitimate syntactic structure for the four-phrase prosodic analysis to align with. However tempting the parser may find that phrasing online, it does not aid comprehension of the 2CE-RC construction because it implies an incorrect syntactic structure.[29]

[28] Wagner admits string-vacuous right-node raising of nonfinite VPs in English, to account for the acceptability of prosodic breaks dividing a series of infinitives as in (i).

(i) Jane tried to begin to learn Spanish.

[29] This form of argument may need to be treated with caution. Langendoen's foundational work on readjustment rules concluded that they differ in formal respects from "standard" syntactic transformations and are immune to the usual constraints (Langendoen 1975). See den Dikken and Lahne (forthcoming) for a summary of recent arguments that syntax proper must disallow vacuous movement.

9.4.3 *A better prosody?*

So far I have been considering only one level of prosodic phrasing for 2CE-RC sentences. Two levels could be achieved by mingling Intonational Phrase boundaries (IPh) and intermediate phrase boundaries (ip) but the former seem too extreme for this construction; typically only the entire utterance would constitute an IPh. But multiple levels of phrasing are possible if prosodic structure is recursive. In what Selkirk (2009) now calls "the standard theory of prosodic constituent structure," the Strict Layer Hypothesis of Selkirk (1984) prohibits recursion: a prosodic unit contained in another prosodic unit is required to be of a type lower in the prosodic hierarchy. But this postulate has been reconsidered in recent theorizing, initially by Ladd (1986), more recently by Wagner (2010) and also by Selkirk herself (e.g., Selkirk 2009). Prosodic recursion has been introduced, allowing embedding of units within units of the same prosodic type, and thereby offering more scope for close prosodic mapping of deeply embedded syntactic structure. Evidence of multiple levels of prosodic phrasing has indeed been found for a German 2CE-RC construction tested by Féry and Schubö (2010). So we should consider the possibility that a multilevel phrasing would provide an optimal compromise between the competing tugs of alignment and phrase length in the English 2CE-RC construction.

An analysis at just two levels could suffice: NP1 || NP2 NP3 VP1 | VP2 || VP3 would retain all of RC1 intact at the higher level, while a weaker lower-level break inside RC1, aligned with the end of VP1, would create two shorter units in place of one large indigestible one. Note especially that this would *not* require VP2 to be extraposed out of the relative clause; VP2 can be sister to the incomplete RC1 (consisting of NP2 NP3 VP1) *within* the higher-level prosodic phrase that spans the whole of RC1. In a reading-aloud task (Fodor and Nickels 2011) this two-level contour was used somewhat less than a fifth of the time in utterances of English 2CE-RC sentences. An example is shown in (32).

(32) The old high-school friend || that the student Bill teaches | wrote to || jogs every morning by the river.

However, this rhythmic pattern apparently did not occur to most participants, even though in this study they were thoroughly familiarized with the content of the target sentence before they pronounced it. Possibly the pattern of phrase lengths that would suggest this phrasing to a reader is too subtle. Or perhaps it is difficult to control two levels of prosodic phrasing at once, even though the grammar permits it in principle. It is apparently not difficult in

German 2CE-RC sentences. And it is not difficult in English in coordinate structures such as (33) (based on examples in Wagner 2010; see also references there).

(33) Who went to the forest?

(a) Lysander | and Helena || or Demetrius.

(b) Lysander || and Helena | or Demetrius.

However, for coordination each shift in prosodic level has semantic import, which may help speaker/hearers keep track of the prosodic structure. For 2CE-RC sentences, prosodic groupings do not relate to the message conveyed.

9.5 Nonprosodic explanations

This prosodic understanding of the extreme parsing difficulty of 2CE-RC sentences does not preclude other explanations or the relevance of other contributing factors. Easily processable examples like (22) do speak against the class of explanations that assume the parser is unable to reapply a rule or parsing subroutine or to hold more than one clause in working memory. But there are other nonprosodic approaches which might be able to account for the relevance of phrase lengths as noted here. Gibson and Thomas (1999) offer an explanation of center-embedding difficulty in terms of Gibson's Syntactic Prediction Locality Theory (SPLT; Gibson 1998), which points to an exceptionally heavy memory load at the NP3 position, where the parser must store predictions of all items necessary to complete the sentence, i.e., VP1, VP2, and VP3. It is proposed that the memory costs at a given word in the sentence increase with the distance between that word and how far back in the sentence the predictions stored there were initiated (except in the case of prediction of the matrix VP, which is argued to be cost-free). At the NP3 position, prediction of VP2 is costly because it was first made back at NP2, but prediction of VP1 is less costly because it was made locally at NP3. Thus, VP2 is the most vulnerable constituent online.

Because SPLT costs are distance-sensitive, reducing the length of intervening material could make processing less onerous. And that might mirror, at least in part, the length-sensitive predictions of the prosodic phrasing account. The distance measures of the two approaches are brought closer together by the fact that SPLT quantifies distance in terms of the number of new discourse entities introduced between one sentence position and another, where in practice this amounts, at least approximately, to the number of nouns and verbs between them. On the prosodic account, the number of

nouns and verbs is an approximate index of the number of stressed lexical items that intervene, and hence of the (im)probability that they can be grouped together into a prosodic phrase. So these measures of the distances between syntactically related elements may converge. Differences in other regards remain: the distances of interest to SPLT are not the same as the distances that matter to prosodic phrasing. Nevertheless, a processing load account is by no means excluded by the considerations adduced here. Possibly prosody and memory costs collude in 2CE-RC parsing. That remains to be seen as investigation is extended to a wider range of examples of the 2CE-RC construction that can distinguish their predictions, such as lengthening the outer constituents NP1 and VP3, which seems to aid prosody assignment but would not reduce memory costs for SPLT.

We are making a start on disentangling these factors. In Fodor and Nickels (2011) we reported data from a reading-aloud paradigm which shows that participants' judgments of pronounceability and comprehensibility, and expert judges' ratings of appropriateness of prosodic contours, correlate highly with the distribution of long and short phrases in 2CE-RC sentences. Sentences like (34) with long outer constituents (NP1 and VP3) and short inner constituents (NP2, NP3, VP1, VP2) were rated significantly higher in all these respects than sentences like (35) with short outer constituents and longer inner constituents (controlled for overall sentence length).

(34) The rusty old ceiling pipes that the plumber my dad trained fixed continue to leak occasionally.

(35) The pipes that the unlicensed plumber the new janitor reluctantly assisted tried to repair burst.

A next step will be to establish whether the "missing VP2 illusion" waxes and wanes depending on phrase lengths. The prosodic phrasing explanation predicts that when VP2 fits comfortably into the RC1 prosodic phrase, as in (34), both VP2 and VP3 can find their proper tree positions, so there is no need for VP2 to be shunned by the parser and the illusion should vanish.