

Center-embedded sentences: What's Pronounceable is Comprehensible

Janet Dean Fodor, Stefanie Nickels and Esther Schott

1. Introduction

Doubly center-embedded relative clause constructions (henceforth 2CE-RC), with the structure shown in (1), are notoriously difficult to process. This is so for classic examples as in (2), whose difficulty seems disproportionate to their brevity, and equally for longer examples such as (3), tested in a much-cited experiment by Gibson & Thomas (1999).

(1) [NP1 [NP2 [NP3 VP1] VP2] VP3]

(2) The girl the man the cat scratched kicked died.

The rat the cat the dog chased killed ate the malt. (from Chomsky & Miller, 1963:286)

(3) 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.

Fodor, Bever and Garrett (1974) made a virtue of this unwieldy construction, by using 2CE-RC sentences as their experimental materials in a number of studies of how the parsing mechanism extracts cues from surface sentences in order to establish their deep structure. Their foundational work in experimental psycholinguistics was achieved at a time when tools for stimulus presentation and response measurement were primitive: DMDX didn't yet exist; event-related potentials (ERPs) hadn't even been dreamed of; some responses were timed with stop-watches. Making fine distinctions of syntactic processing difficulty with the blunt instruments to hand could be tricky and frustrating. But by working with a sentence type so difficult that comprehension often failed, Fodor, Bever and Garrett were able to expand the scale of response measures so that performance differences of interest could be observed.¹

The 2CE-RC construction has three well-established peculiarities.

- I. First is its unusually difficult comprehension. Such sentences have been deemed incomprehensible, unacceptable, even ungrammatical. Intuitively, the increment of processing cost due to embedding one object-gap RC inside another one is much greater than the cost of embedding the same RC inside a main clause.
- II. Second is an observation by Bever (1988), who first noted the ameliorating effect of using a pronoun as NP3, as in *The man the girl I scratched kicked died*.

¹ "Self-embedded sentences... exhibit features that are relevant to testing the significance of certain types of surface clues to deep structure configurations. We have employed them in the present experiments because, with iteration of the self-embedding operation, Ss have difficulty in understanding them. This provides an opportunity for the presumed facilitatory effects of surface structure clues to be revealed more strongly than in the case of sentences which Ss find easy to understand." (Fodor & Garrett, 1967:291)

- III. Third is that a 2CE-RC sentence may be perceived, wrongly, as equally or more grammatical if VP2 is omitted, as in *The girl the man the cat scratched died*, which may be judged acceptable. This is the ‘missing-VP illusion’. (References to experimental data in several languages are in section 3.2 below.)

Many explanations have been offered over the years since Miller & Chomsky (1963) first drew attention to this recalcitrant construction type. We summarize a handful of them in Table 1.

.....

Table 1: A sample of proposed explanations for the processing difficulty of 2CE-RC sentences

- The parser cannot recursively call the same sub-routine (Miller & Isard 1964).
- A three-NP sequence with no relative pronouns is misparsed as coordination (Blumenthal 1966).
- Exponential increase in number of potential grammatical relationships (Fodor & Garrett 1967).²
- The parser cannot assign both subject and object roles to NP2 (Bever 1970).
- The Sausage Machine parser can’t correctly ‘chunk’ the word string (Frazier & Fodor 1978).
- ‘Disappearing’ syntactic nodes in complex tree structures (Frazier 1985).
- Syntactic prediction locality theory (SPLT, Gibson & Thomas 1999).

Along with these different accounts of the source of the difficulty, there are corresponding proposals about how the difficulty can be minimized, thus acknowledging the considerable range of variation in the severity of the problem that is observed across examples. Hudson (1996) ran a series of informal experiments in which students had to recall a spoken sentence; he reports error rates for 2CE-RC constructions ranging widely, from 7% for sentence (4) to 81% for sentence (5), though matched for number (if not frequency) of words.

(4) The shells that the children we watched collected were piled up.

(5) People that politicians who journalists interview criticise can't defend themselves well.

We will argue for a significant role of prosodic phrasing in creating the difficulty of the 2CE-RC construction, and correspondingly a role for prosodic phrasing in facilitating its processing. Specifically, we propose that there is an alignment problem at the syntax-prosody interface, consisting of a mismatch between the heavily nested syntactic structure and the flat

² “Given one embedding, two nouns must be assigned to each of two verbs as subject and two nouns must be assigned to each of two verbs as object. Hence, we have four possible analyses of N1 N2 VI V2 into NVO assuming no noun is both subject and ob[ject] of the same verb. However, given two embeddings, three nouns must be assigned [to] each of three verbs as subject and three nouns must be assigned to each of three verbs as object. Still assuming no noun may be assigned as both subject and object of the same verb, we have 18 possible analyses of the double-embedded case (if the final verb is intransitive, there are two possible analyses for the single embedding and 12 for two embeddings).” (Fodor & Garrett 1967:296)

structure required by prosodic phrasing.³ We predict as a corollary that if the prosody can be made natural, the syntax will be computable without the usual extreme difficulty. Of course such sentences will never be very easy to parse and comprehend. They contain two relative clauses, each of which modifies a subject and contains an object 'gap', properties well-known to increase processing difficulty; prosody cannot eliminate these complexities. But our data suggest that the difficulty of double center-embedding *per se* can be tamed by cooperative prosody. We present examples showing that selective shrinking and lengthening of phrases can coax the prosodic processor into creating rhythmic packages that do fit well with the nested syntactic tree structure. Short inner phrases help with that, while short outer ones hinder. The appropriate prosody is difficult to achieve, for reasons that will be explained, and typical syntactic phrase lengths in 2CE-RC sentences do not cooperate in this regard, which may be why this prosodic phenomenon has not been widely recognized. We will show that the prosodic approach offers explanations for all three distinctive peculiarities of the 2CE-RC construction listed above: the near-incomprehensibility of most standard examples; the pronoun effect; and (perhaps) the missing-VP effect.

2. A facilitative prosodic phrasing⁴

Suppose a speaker wishes to tell a friend “The girl the man the cat scratched kicked died.” The syntactic structure of this 2CE-RC sentence is sketched in Figure 1, with some details omitted so as to focus attention on the main configurational relations.

³ That prosodic structure must be flat was entailed by the Strict Layer Hypothesis of Selkirk (1981 and elsewhere), which forbade recursion in prosodic structure: one prosodic unit could be embedded in another only if they were at qualitatively different levels of the prosodic hierarchy. More recently this constraint has been recast as a violable condition which may be outweighed by other constraints on prosodic structure (Selkirk 1995; Truckenbrodt, 1999; Myrberg 2013). Wagner (2010) presents robust evidence for recursive prosodic phrasing in English coordination constructions, and Féry and Schubö (2010) demonstrate it in RC constructions in German, but there is scant evidence for it in our data for English 2CE-RC. A possible reason is offered in section 4 below.

⁴ Fodor & Garrett (1967) compared 2CE-RC stimuli (such as *The pen the author the editor liked used was new*) pronounced with neutral prosody and pronounced with ‘expressive’ prosody (details not specified) and found little benefit from the latter, compared with the benefit of presence versus absence of relative pronouns in the RCs.

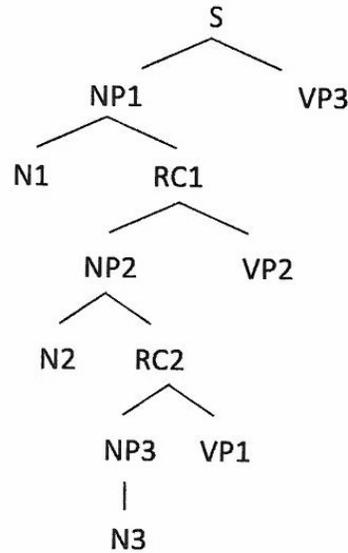


Figure 1: *Syntactic tree structure (simplified) for the 2CE-RC construction*⁵

How could the would-be speaker set about assigning a prosodic structure to this syntactic tree? The sentence is too long, even with these short constituents, to be expressed in a single prosodic phrase,⁶ so it needs to be snipped apart at natural syntactic breaks, presumably starting with the major break between the subject and predicate of the sentence. It turns out that a critical issue is how many units to divide the structure into: 2 units or 3 or 4 or more. As often noted, 2CE-RC sentences are frequently pronounced with a ‘list intonation’, which amounts to dividing the word string into 6 prosodic phrases, each NP and VP a unit to itself. This is not helpful; in fact it is a clear mark of failure to comprehend. Thus the challenge is posed: not dividing the word sequence prosodically is impossible, but dividing it into too many pieces obscures the syntactic structure.

An optimal division must satisfy two criteria: it must do as little damage as possible to the syntactic tree, while also satisfying prosodic constraints. Doing least damage to the syntactic tree structure means cutting the tree not arbitrarily but at natural syntactic joints. In other words, the prosodic units should be aligned with syntactic phrases, as far as is possible. However, the constraints that apply at the syntax-prosody interface are a heterogeneous set, and they include

⁵ Current syntactic analyses of relative clause structure differ with respect to exactly how and where the RC is embedded. For simplicity here, Figure 1 does not show DP structure dominating the NPs. Also, the RC is shown beneath NP as a sister to a lexical Noun node (which might be dominated by an N-bar node) rather than as sister to a maximal projection (Noun Phrase, NP) as in the familiar shorthand representation of 2CE-RC structure in (1) above. We will continue to use that shorthand for convenience in what follows, and hope that the variant notations create no confusion.

⁶ We use the term ‘prosodic phrase’ to denote a unit lower in the prosodic hierarchy than a full Intonational Phrase (IPh). These units are referred to in the linguistics literature in various terms: *intermediate phrase (ip)*, *major phrase (MaP)*, *phonological phrase* or *p-phrase*. Although RCs are clausal units, they do not commonly constitute IPhs, at least in English (see, e.g., Göbbel, 2013, p.136). Non-restrictive RCs do, but they require a full relative pronoun such as *which* or *who*, and are precluded in our materials which have only a *that* or null complementizer.

eurhythmic constraints on optimal phrase length and balance which may compete with alignment constraints. These are presented in Optimality Theory as 'soft' constraints, which apply except where they are out-ranked by some more prominent constraint in the language in question.

Table 2: Some constraints on prosodic phrasing

A. Relation to syntax/semantics

Edge alignment (Align_R XP): “The right edge of any XP in syntactic structure must be aligned with the right edge of a MaP in prosodic structure” (Selkirk 2000: 232)

Wrap: “Each syntactic XP must be contained in a phonological phrase” (Truckenbrodt, 1995: 10)

B. Prosodic phrase length constraints

Binary Minimum: “A major phrase must consist of at least two minor/accental phrases.” (Selkirk, 2000:244)

Binary Maximum: “A major phrase may consist of at most two minor/accental phrases.” (Selkirk, 2000:244)

Uniformity: “A string is ideally parsed into same length units.” (Ghini, 1993:56; see also the Balance principle of Gee & Grosjean 1983)

-----End of Table 2-----

Cutting the word string at the highest syntactic level, between the matrix subject and its verb phrase) yields (6). (In all examples, || indicates a prosodic phrase boundary.)

(6) The girl the man the cat scratched kicked || died.

A note on reading the examples: It is most illuminating to read them aloud, or at least to sound them out in one's head. They should be pronounced with a prosodic break everywhere where shown by || and nowhere else.

Although it fits the syntactic structure, prosodic phrasing (6) flagrantly violates the Uniformity/Balance principle. There are 9 words, divided into 8 for the first prosodic phrase and 1 for the second prosodic phrase. Counting stressed syllables is more appropriate for (English) prosody than counting words, but still there is an imbalance of 5 +1.

For the 2-phrase prosody to be successful, it needs the encouragement of balanced phrase lengths, as in (7).

(7) The girl the man I love met || died of cholera in 1960.

Balanced aligned prosody: 7+5 words; 4+4 stressed syllables

Although this example is longer than (6), remarkably the 2CE-RC construction now sounds very much like a normal sentence.

However, a prosodically balanced example like (7) is rare. The sentence has both RCs within the prosodic phrase that encompasses its matrix subject NP, which is followed and balanced by a long matrix VP. Squeezing 2 RCs into the space of a single prosodic phrase is quite an art, so it is not likely to occur often in normal language use. The stressless pronoun in the inner relative clause (RC2) in (7) provides almost the only way to achieve it.⁷ It allows the 7-word subject, containing 2 relative clauses, to be pronounced with only 4 stressed syllables. Otherwise there would have to be at least 5 stressed syllables in the subject, as in examples (8) and (9), and this is usually judged to be too much; it oversteps the maximum length limit for an (intermediate) prosodic phrase.

(8) The girl the man Jill loves met || died of cholera in 1960.
Balanced aligned prosody: 7+5 words; 5+4 stressed syllables

(9) Girls men Jill loves met || died of cholera in 1960.
Balanced aligned prosody: 5+5 words; 5+4 stressed syllables

To summarize so far: Except with a pronominal NP3, a 2-chunk prosody compatible with the syntax is hard to achieve, since an NP containing two RCs is not usually as short as a prosodic phrase needs to be (in English). For a more stable solution, therefore, we need to snip the syntactic tree structure again, creating a 3-phrase prosody.

A cut at the next level down in the syntactic tree would be between NP1 and the RC1 that modifies it (see Figure 1 above), creating a sequence of three prosodic phrases: NP1 || RC1 || VP3. This clearly should be helpful in easing the crush in the overstuffed matrix clause subject in examples (8) and (9). However, once again the constituent lengths have to cooperate. Separating off RC1 as a prosodic phrase does not by itself ameliorate syntactic processing, as can be seen in (10), where the phrase lengths are seriously imbalanced.⁸

(10) The girl || that the young man I love met in Barcelona || died.
Unbalanced aligned prosody 2+9+1 words; 1+6+1 stressed syllables

However, the same syntactic cut with cooperating phrase lengths, as in (11), does permit fairly painless processing. Note that the outer phrases (NP1 and VP3) are longer in (11) than in (10), and they balance a central RC1 which is about as short as it can be.

⁷ Non-prosodic explanations of the pronoun advantage have been proposed by Bever (1988) and Gibson and Thomas (1999). More generally, Bever (1970) noted an improvement in processing when the three NPs are varied in form.

⁸ The examples from this point onward all have an overt complementizer (sometimes termed a relative pronoun) *that* at the beginning of RC1. This is because in all of these examples there is a prosodic boundary at that position, and an overt *that* is preferred after a prosodic boundary (Fox and Thompson, 2007). This lengthens RC1 by one word but does not add to the stressed syllable count. In these examples we have not inserted *that* to introduce RC2, to avoid giving the impression that there should be a prosodic boundary there.

- (11) The elegant woman || that the man I love met || moved to Barcelona.
Balanced aligned prosody: 3+6+3 words; 2+3+3 stressed syllables

The striking difference in naturalness between (10) and (11) underscores the importance of phrase lengths in making 2CE-RC constructions pronounceable. Indeed, with encouraging phrase lengths as in (12), the 3-phrase prosody works quite well even with a non-pronominal inner subject, suggesting that this prosodic pattern is indeed more stable and realistic than the 2-phrase prosody we considered above.

- (12) The elegant woman || that the man Jill loves met || moved to Barcelona.
Balanced aligned prosody: 3+6+3 words; 2+4+3 stressed syllables

Taking stock at this point: We have found a successful recipe for creating a 2CE-RC structure that is recognizable, more or less, as a normal English sentence. The trick is to adjust the lengths of the lexical/syntactic phrases so that they are also acceptable as prosodic phrases. To the best of our knowledge this is a novel observation, though it is prefigured in large part by the Sausage Machine account of the processing difficulty of 2CE-RC sentences (Frazier & Fodor 1978:306-312); see Fodor (2013) on how the Sausage Machine's PPP (Preliminary Phrase Packager) morphed into a Prosodic Phrase Processor, as here. It is especially interesting that compatibility between syntactic phrasing and prosodic phrasing is not achieved, as might have been expected, by ensuring that all six syntactic units have the length of a typical prosodic phrase. Instead, the successful strategy packs most of the syntactic structure inside a single prosodic phrase, cramming NP2 NP3 VP1 and VP2 together without any breaks between them.

What we have arrived at so far is that 2CE-RC sentences are relatively easily parsed if their phrase lengths permit a prosodic division of the word string into weight-balanced units NP1 || RC1 || VP3, achieved by lengthening NP1 and VP3, and shortening RC1. However, there are practical limits on how short RC1 can be. In order to accommodate more typical sentences in which RC1 is more substantial than in (12), we could apply the snipping procedure once more, to break up that complex constituent. The next natural cutting point in the syntactic tree is indeed inside RC1, between its complex subject and its VP (see Figure 1).⁹

For example (12) as it stands, this is not a success; the resulting (13) is prosodically very unnatural. To satisfy the optimal length constraints on prosodic phrasing, we need to lengthen VP2, as in (14), to achieve prosodic balance inside RC1.

- (13) The elegant woman || that the man Jill loves || met || moved to Barcelona.
Unbalanced aligned prosody: 3+5+1+3 words; 2+3+1+3 stressed syllables

- (14) The elegant woman || that the man Jill loves || met on a cruise ship || moved to Barcelona.
Balanced aligned prosody: 3+5+3+3; 2+3+3+3 stressed syllables

⁹There are many other ways of creating a sequence of 4 prosodic units out of the 6 phrases of sentence (12) (e.g., *The elegant woman || that the man Jill || loves met || moved to Barcelona*), but they all align improperly with the syntax and are considered extremely unnatural; see Fodor (2013) for discussion in terms resembling Selkirk's Sense Unit Constraint.

However, though intended to appease the prosodic processor, this extra cut, dividing the sentence into a sequence of four balanced prosodic phrases, is not obviously an improvement for the syntactic processor. According to our intuitions and those of other English speakers we have consulted, sentence (14) feels as if it is beginning to break up into a list-like structure, reminiscent of the familiar unhelpful 6-phrase pronunciation of (2) and (3) above. Thus the additional prosodic break in (14), though it would have been expected to contribute by relieving the crush inside RC1, seems to be a move in the wrong direction from the perspective of syntactic processing. Dividing the word string at its joints is good but this division goes a step too far.¹⁰ Therefore the 3-phrase prosody NP1 || RC1 || VP3 may be the best truce between syntax and prosody that can be achieved. Our goal is to understand why this is so. But at least, the fact that this prosody imposes such stringent constraints on phrase lengths does explain why it is so rarely encountered.

To summarize: We have observed here a struggle in 2CE-RC sentences between balanced prosodic weight and prosody-syntax alignment. Depending on the lexical content of a particular sentence, there may or may not be a good way of reconciling these conflicting concerns. Table 3 summarizes the intuitions we have presented informally above. In section 3 we report two experiments which corroborate these intuitions. In section 4 we offer a theoretical explanation.

Table 3: Summary of intuitive judgments of processing difficulty in relation to prosodic phrasing

Division of 2CE-RC sentence structure into 2 syntactically aligned prosodic phrases (NP1 NP2 NP3 VP1 VP2 || VP3) is very difficult to achieve, but when phrase lengths permit it, it is helpful for comprehension.

Division of the sentence structure into 3 syntactically aligned prosodic phrases (NP1 || NP2 NP3 VP1 VP2 || VP3) is difficult but can be achieved if the inner constituents are short and the outer ones are long. It greatly facilitates parsing and comprehension.

Division into 4 syntactically aligned prosodic phrases, by breaking VP2 out of the upper relative clause (NP1 || NP2 NP3 VP1 || VP2 || VP3), is less acceptable prosodically and less helpful for parsing than the 3-phrase prosody. It shares some of the unnaturalness of the common but unhelpful 6-phrase ‘list intonation’ pronunciation (NP1 || NP2 || NP3 || VP1 || VP2 || VP3).

----- End of Table 3 -----

¹⁰ Some English speakers may be able to control two degrees of boundary strength (see Liberman, 2013). That could allow a 4-phrase pattern such as *The elegant woman || that the man Jill loves | met on a cruise ship || moved to Barcelona*, with the break between the subject and predicate inside RC1 weaker than the breaks surrounding RC1. We encountered this rarely in our experiments, but there may be individual variation here, such that speakers who are particularly attuned to prosody are better able to deploy this pronunciation than linguistically naive speakers. Individual differences certainly deserve attention in future research.

3. Elicited prosody experiments

We report two experiments here, each described in more detail below, to assess the predicted facilitating effect of the 3-phrase prosody. In Experiment 1 (Fodor & Nickels 2011) participants read sentences first silently for comprehension, then aloud for recording, followed by judgments of pronounceability and comprehensibility. A familiarization procedure, described below, was employed in hope that it would increase the percentage of successfully parsed items. In Experiment 2 (Schott 2012; Schott & Fodor 2013) the ‘missing-VP2 illusion’ described above was employed as a more objective measure of successful syntactic parsing. Participants read the sentences first silently, then aloud for recording, followed by a yes/no answer to the question “Is something missing from this sentence?”

In both experiments, we manipulated phrase lengths in order to compare sentence versions designed to be susceptible to the helpful 3-phrase prosody and versions which were designed to resist that prosody. We refer to the former as ENCouraging, and the latter as DISCouraging. In both cases RC1 was introduced by *that* and RC2 was not.

3.1 Experiment 1 (rating task with familiarization)

Materials

Experiment 1 manipulated both the length and the ‘weight’ of the six phrases in a sentence, and compared the 2CE-RC structure with items with a single RC embedding. Items were constructed as follows; examples of each type are in Table 4 below.

2CE-RC(length): 4 pairs of 2CE-RC sentences, with phrase length manipulation. Paired items had the same total number of words, plus or minus one. They had similar though not identical semantic content, but differed in their distribution of phrase lengths. To ENCourage the 3-phrase prosody, the outer constituents NP1 and VP3 were long and RC1 was quite short (by relative clause standards), with the result that these three constituents were more or less equal in length. In their DISCouraging counterparts, the outer constituents were too short to be phrased alone,¹¹ while the RC1 was too long to be phrased as a single unit.

2CE-RC(weight): 4 pairs of 2CE-RC sentences, with lexical ‘weight’ manipulation. In contrast to 2CE-RC(length) sentences, each of the 6 phrases were matched in word count across the ENC/DISC items in a pair.¹² Paired items had roughly similar semantic content, but they differed in the predictability (corpus frequency, default status) of their content words, to either

¹¹ Short phrases consisting of even a single word can be prosodically acceptable if heavily stressed. Such pronunciations presuppose a rich discourse context with prominent contrasts. However, this observation is of interest because, although we have not tested it yet, it suggests that the prosodic weights of the constituents are more relevant to 2CE-RC parsing than measures of lexical/syntactic length.

¹² In all 2CE-RC(weight) sentences, both ENCouraging and DISCouraging, every NP consisted of a definite determiner and a single noun (or a proper name in NP3 position), and every VP consisted of a single verb (sometimes with a particle/preposition), see example in Table 4.

ENCourage or DISCourage the 3-phrase prosody, on the hypothesis that less predictable words would be less susceptible to phonetic reduction and thus would create prosodically ‘weightier’ phrases. The mean lexical frequencies for the ENC and DISC sentences in a pair were matched.

1CE-RC: 4 pairs of sentences with the structure NP1 [NP2 VP2] VP3, in which a single-level RC modifies the subject of the main clause. Paired items had, again, the same number of words, plus or minus one, similar though not identical semantic content, but differed in the distribution of phrase lengths. To ENCourage the 3-phrase prosody the outer constituents were long and the RC was short; to DISCourage the 3-phrase prosody the outer constituents were short and the RC was long. In their overall length and their phrase length distributions these sentences were comparable to the 2CE-RC(length) items, although their syntactic structure was shallower.

2CE-RC(G&T): 4 typical 2CE-RC items from a previous study (Gibson & Thomas, 1999), with uniformly long constituents, as in sentence (3) above. We regarded these phrase lengths as DISCouraging the 3-phrase prosody.

16 assorted filler items, of 4 subtypes that differed in structure but contained multiple clauses and mild parsing challenges: the *if not because* construction, parenthetical adverbial clauses, early/late closure garden paths, NP/clausal complement garden paths.

ENC and DISC examples of each item type are shown in Table 4.

Table 4: Examples of each type of Experiment 1 materials

2CE-RC(length)

ENC: The rusty old ceiling pipes that the plumber my dad trained fixed continue to leak occasionally.

DISC: The pipes that the unlicensed plumber the new janitor reluctantly assisted tried to repair burst.

2CE-RC(weight)

ENC: The soufflé that the waitress the boss hired brought disintegrated.

DISC: The drink that the hostess the nightclub employed stirred spilled.

1CE-RC

ENC: The elderly Austrian woman that the retired boxer danced with just died in an automobile accident.

DISC: The woman that the recently retired middle-weight boxer had danced with on a South-American cruise died.

2CE-RC(G&T)

DISC: The prayer that the monk the religious fanatic had persecuted relentlessly was chanting every day was echoing in the empty church.

4 types of filler items

1. If Barbara wasn't crying because she lost her excellent exam notes, what was her problem?
2. The engineers continued, even though they knew it was hopeless, to try to repair the damaged bridge support.
3. Bertram told the physiotherapist that whenever he tries to exercise his leg muscles start to cramp.
4. Professor Thompson knew the internationally famous literary critic giving the speech was a fraud.

-----End of Table 4-----

Participants and procedure

Twenty-eight native English speaking participants (9 male) recruited at CUNY Graduate Center were tested individually. Their task was to judge the pronounceability and comprehensibility of sentences that were displayed visually on a computer screen. On the assumption that even in their most ENCouraging versions these materials would be too challenging for many people to process, we employed a familiarization procedure with the aim of increasing the overall level of comprehensibility and thus avoiding floor effects that could obscure judgment differences between item types. Each sentence (including fillers) was built up in 5 successive steps, as illustrated in (15) and (16) for the ENC and DISC versions respectively of 2CE-RC(length), and in (17) for a filler item.

(15) 2CE-RC(length), ENC version

My dad trained a plumber.

Here is the plumber my dad trained.

The plumber my dad trained fixed the rusty old ceiling pipes.

Here are the rusty old ceiling pipes that the plumber my dad trained fixed.

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

(16) 2CE-RC(length), DISC version

The new janitor reluctantly assisted an unlicensed plumber.

Here is the unlicensed plumber the new janitor reluctantly assisted.

The unlicensed plumber the new janitor reluctantly assisted tried to repair the pipes.

Here are the pipes that the unlicensed plumber the new janitor reluctantly assisted tried to repair.

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

(17) Filler sentence

The bridge support was damaged.

The engineers were trying to repair it.

The engineers were trying to repair the damaged bridge support.

They continued to try, even though they knew it was hopeless.

The engineers continued, even though they knew it was hopeless, to try to repair the damaged bridge support. <NEEDS to be on one line>

The 5 sentences in a set were displayed successively, each one on a single line, across the middle of the screen. The participant was instructed to read each sentence silently first for comprehension, then aloud for recording, and then to press an arrow key to remove that sentence and bring up the next one in the set. The first four sentences in a set were in white font against a dark background; the fifth one was in yellow font, and the participant knew s/he would have to judge the yellow sentence on two 5-point scales (5 = best) that appeared in succession on the screen: How easy was it to pronounce? How easy was it to understand?

Predictions: ENC versions were expected to be rated higher on the pronounceability scale than their DISC counterparts. With regard to the prosody with which they were pronounced, we anticipated that ENC versions would more often exhibit the optimal NP1 || RC1 || VP3 prosodic structure, while DISC versions would be divided into more chunks, creating a less natural and more 'list-like' prosody for the sentence. On the hypothesis that a more natural prosodic phrasing would facilitate construction of the correct syntactic structure, ENC versions were expected to be rated higher on the comprehensibility scale than their DISC counterparts.

Results: Pronounceability judgments and evaluation of produced prosody

Participants' ratings of pronounceability are shown in Figure 2.

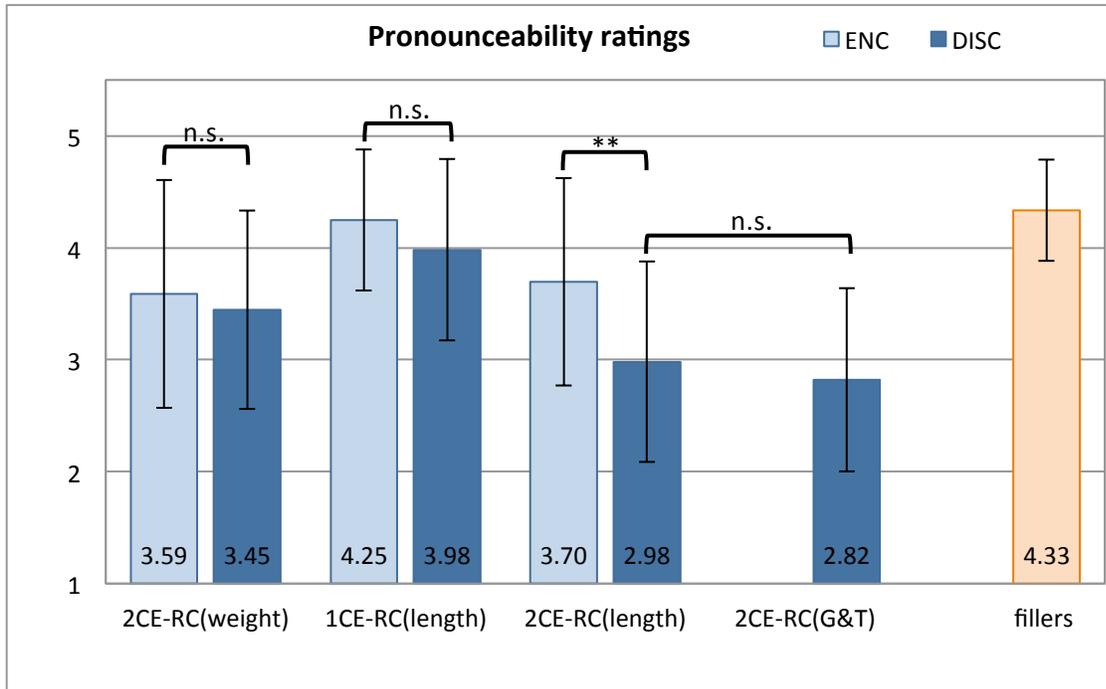


Figure 2: Mean scores on the pronounceability judgment scale, by stimulus type for $n=28$ subjects. Whiskers indicate the standard deviation. ** indicates $\alpha < .001$.

A one-way repeated-measure ANOVA including all eight different conditions revealed significant differences among them ($F(7, 189) = 26.11, p < .001$). Pairwise contrasts were computed to reveal which conditions differed specifically, only a selection of which will be reported here for reasons of space. The ENC versions of the 2RC-CE(length) items were rated as significantly easier to pronounce than their DISC versions ($F(1, 27) = 25.35, p < .001$), as predicted. The ratings for the 2CE-RC(G&T) items did not differ reliably from those for the DISC versions of the 2CE-RC(length) items ($F(1, 27) = 1.73, p = .199$); however, this may not be a fair comparison since the G&T sentences were longer, by 4.75 words on average, than the 2RC-CE(length) DISC items that we constructed. The ratings for the ENC versions of the 2CE-RC(length) sentences were significantly lower than those for the ENC versions of the 1CE-RC items ($F(1, 27) = 11.21, p < .01$) and than those for the fillers ($F(1, 27) = 20.90, p < .001$), showing that even with favorable phrase lengths there remained some difficulty in finding an appropriate pronunciation of the nested 2CE-RC structure. For the 2CE-RC(weight) items, the ENC and DISC versions differed numerically in the direction expected but the effect was small and not statistically reliable ($F(1, 27) = 0.66, p = .424$). For the 1CE-RC (single level relative clause) items, which had phrase length patterns quite similar to those of the 2CE-RC items, ENC versions were also judged to be easier to pronounce than the DISC versions, but this difference showed only a weak trend towards statistical significance ($F(1, 27) = 2.55, p = .122$).

These self-reports of pronounceability by participants are corroborated by judgments of the appropriateness of the produced prosodic contours, by two trained judges (graduate students of linguistics) who were unaware of the design of the experiment. They judged only the doubly center-embedded sentences (i.e., 2CE-RC(weight), 2CE-RC(length) and 2CE-RC(G&T); see Figure 3). Their judgments, on a scale of 1 to 5 (5 = fully natural), were very similar to the pronounceability ratings by participants. The differences between the five conditions were confirmed by an ANOVA ($F(4, 104) = 27.84, p < .001$). Follow-up pairwise comparisons showed significantly better ratings for the 2CE-RC(length) in their ENC versions than in their DISC versions ($F(1, 26) = 34.54, p < .001$). The 2CE-RC(G&T) sentences did not differ from the 2CE-RC(length) DISC items ($F(1, 26) = .58, p = .45$). In contrast to the pronounceability judgments by participants, the judges' ratings of prosodic appropriateness for the ENC and DISC versions of the 2CE-RC(weight) sentences showed a significant difference in favor of the ENC version ($F(1, 26) = 7.34, p < .02$).

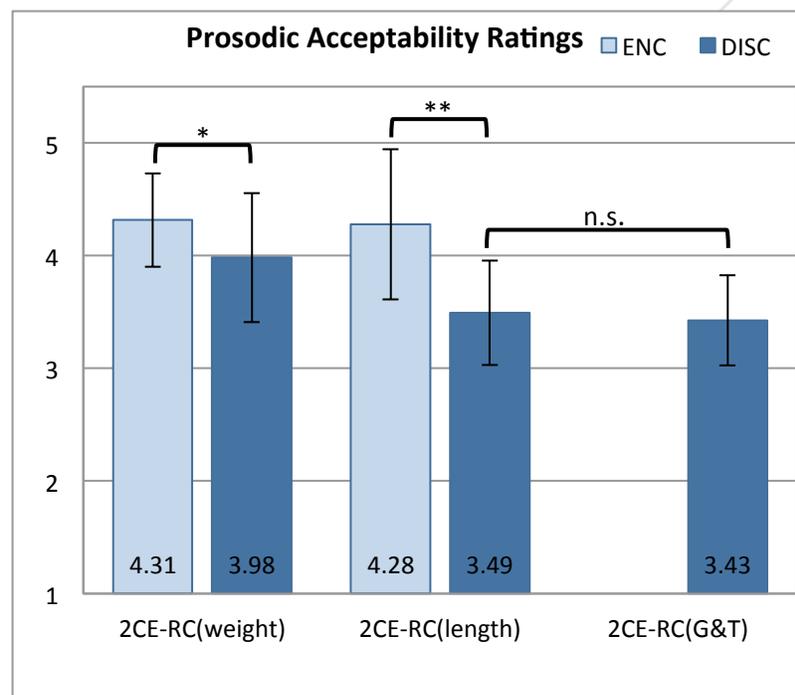


Figure 3: Mean appropriateness ratings of produced prosody, by trained judges. Whiskers indicate standard deviations. * indicates $\alpha < .05$; ** indicates $\alpha < .001$.

In short: the ENC phrase length manipulation did make the 2CE-RC(length) sentences easier for readers to pronounce, and expert judges evaluated the overall prosodic contour of the ENC versions as more appropriate than that of the DISC versions for both the length and weight manipulations.

Results: Comprehensibility judgments

The comprehensibility ratings by participants (see Figure 4) also showed significant differences among the eight tested conditions ($F(7, 189) = 33.95, p > .001$). Specifically, among the 2RC-CE(length) items the ENC versions were judged to be easier to understand than the DISC versions ($F(1, 27) = 30.98, p < .001$), as predicted. However, the ENC(weight) items showed no comprehensibility advantage over their DISC versions ($F(1, 27) = 1.75, p = .197$), possibly because the unfamiliarity of some of the words they contained was their most prominent property.

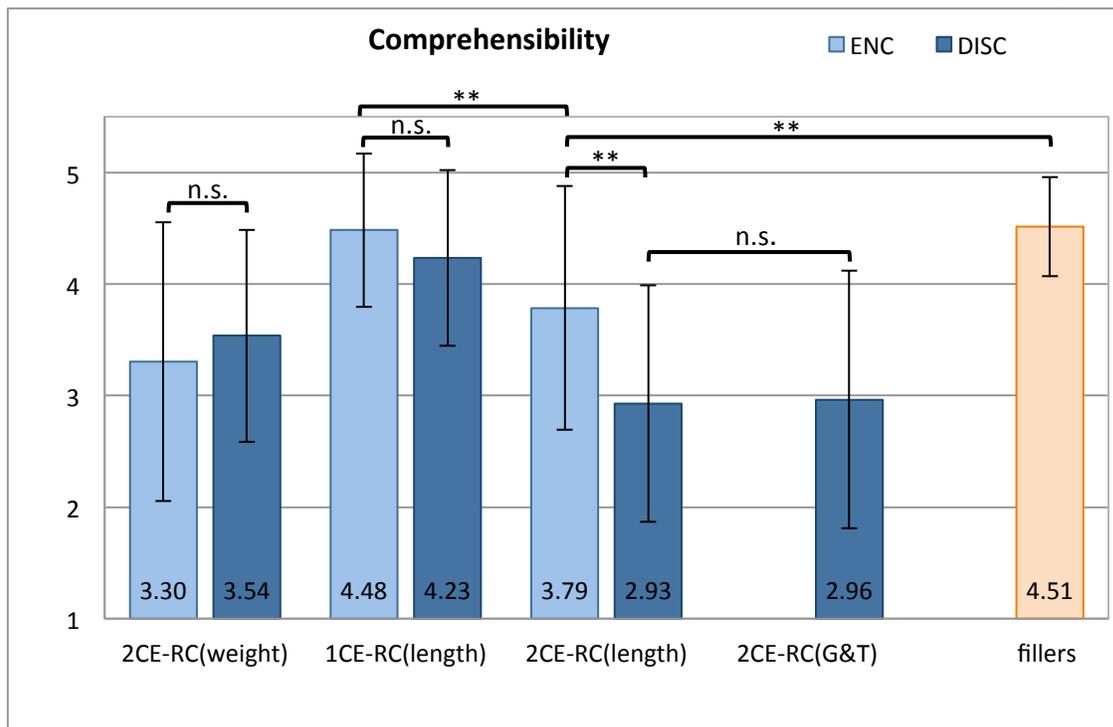


Figure 4: Mean scores on the comprehensibility judgment scale, by stimulus type. Higher scores indicate higher judged comprehensibility. Whiskers indicate standard deviations. ** indicates $\alpha < .001$.

It is noteworthy that even the ENC versions of the 2CE-RC(length) items were judged to be less comprehensible than the filler items ($F(1, 27) = 22.40, p < .001$) and than the ENC versions of the 1CE-RC(length) items ($F(1, 27) = 22.68, p < .001$). Thus, it cannot be claimed that the prosodic facilitation achieved by ENC phrase lengths eliminated all processing difficulty from the doubly center-embedded construction. As anticipated, the comprehensibility ratings of the 2CE-RC(G&T) items did not differ from those for the 2CE-DISC(length) items ($F(1, 27) = .07, p = .787$).

Results: Produced prosodic phrasing

A closer look at the produced prosody of these sentence types supports the hypothesis, originally formulated on the basis of intuitive judgments, that the prosodic phrasing most conducive to comprehensibility consists of three prosodic phrases, with the entire RC1 as the middle one: NP1 || NP2 NP3 VP1 VP2 || VP3. The two expert judges (see above) established the locations of prosodic boundaries, coding them either as typical phonological ('intermediate') phrase boundaries (level 3 in the ToBI transcription system) or as stronger (ToBI level 4). Sentence prosody patterns were then classified into 3 types, focusing on the latter part of the sentence, the sequence of three VPs, where the greatest differences were observed. The contrast of primary interest was between pronunciations in which VP2 was integrated into the RC1 prosodic phrase, and pronunciations in which VP2 was preceded by a boundary separating it from the preceding sequence of NP2 NP3 VP1. This 'separated VP2' prosody was expected to be more common for DISC than for ENC items. Since it precludes the 3-phrase grouping believed to be facilitative of syntactic parsing, it was predicted to be associated with low comprehensibility ratings. The third prosodic pattern encoded was one that lacked a prosodic boundary before VP3. This was an occasional pronunciation, which the judges had regarded as very unnatural, and which we deemed to indicate complete failure to comprehend the VP sequence.

Figure 5 shows the 2CE-RC(length) items, in their ENC and DISC versions, and also the 2CE-RC(G&T) items, which had patterned like the DISC versions of the 2CE-RC(length) items in the judgment data reported above.

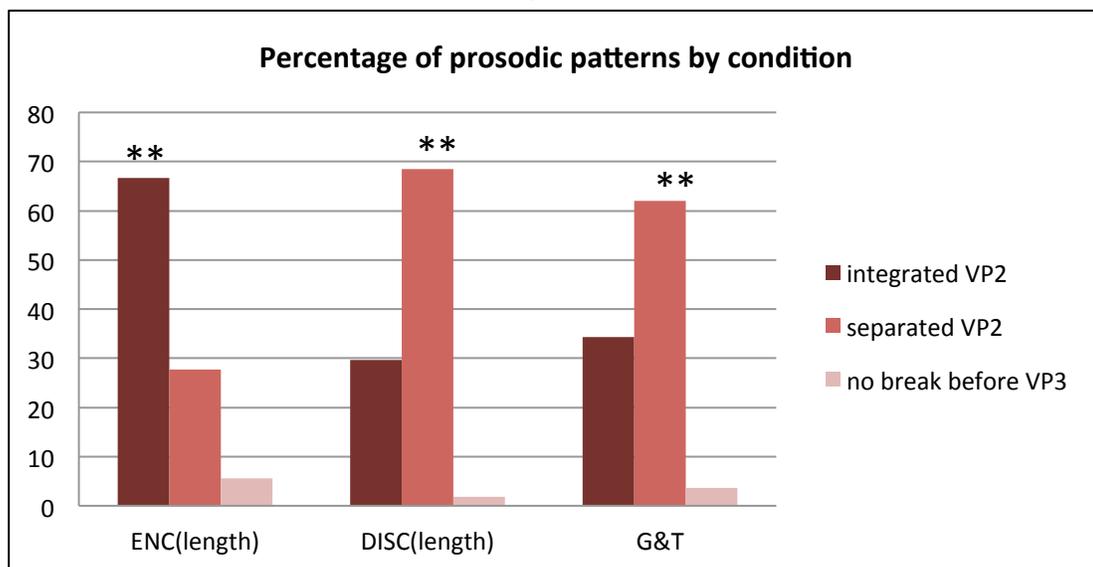


Figure 5: *Prosodic phrasing patterns in productions of 3 sentence types. Asterisks indicate significant differences from equally distributed frequencies across sentence types ($\alpha < .001$).*

A chi-square test confirmed that the frequencies of the observed phrasing patterns differed significantly between conditions ($\chi^2(4, N = 216) = 22.54, p < .001$). Specifically, an integrated VP2 (i.e., absence of any prosodic break immediately preceding VP2) was more common for ENC than for the DISC and G&T item types ($p < .05$). In fact, the DISC and G&T items exhibited a strong bias toward the separated VP2 pattern ($p < .05$). The pattern with no break before VP3 was rare in all conditions. These data provide some support for the hypothesis that the ENC pattern of phrase lengths which improves comprehensibility does so by facilitating the inclusion of VP2 within the central prosodic phrase of a 3-phrase prosodic contour.

Further evidence for this conclusion comes from the relationship between the integrated-VP2 prosody and the distribution of comprehensibility scores, as shown in Figure 6.

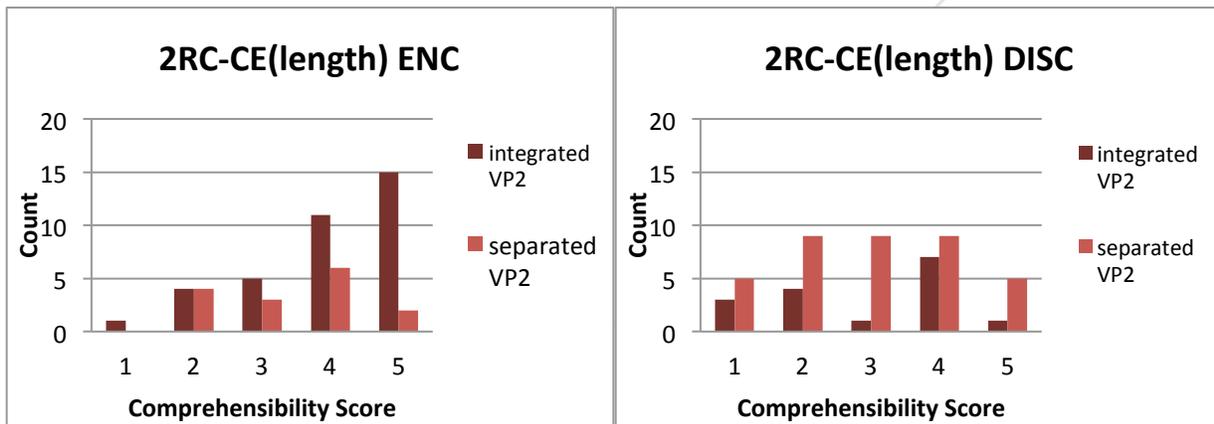


Figure 6: *Frequency of doubly center-embedded sentences phrased with integrated VP2 and separated VP2 and associated comprehensibility scores, for ENC items (left panel) and DISC items (right panel). Higher scores indicate higher judged comprehensibility.*

For the ENC versions there is an overwhelming tendency to rate the integrated VP2 items as highly comprehensible (scale positions 4 and 5). Even for the DISC versions, there is a peak at scale point 4 for the integrated VP2 items. By contrast, for both ENC and DISC the separated VP2 items mostly cluster at the mid-scale values. It is of interest that the separated VP2 items don't peak at the bottom end of the comprehensibility scale. This suggests that a list-like prosody with distinct separated phrases may improve comprehension of the individual phrases even if it obscures the syntactic/semantic relations among them.

The fact that our phrase length manipulations were not 100% successful in inducing the prosody we had intended (see Figure 5) provides the opportunity to distinguish between effects of prosody on comprehensibility and effects of phrase lengths on comprehensibility. Thus, in Figure 7 the ENC and DISC data are pooled, and the comprehensibility scores are sorted only by prosodic phrasing of VP2.

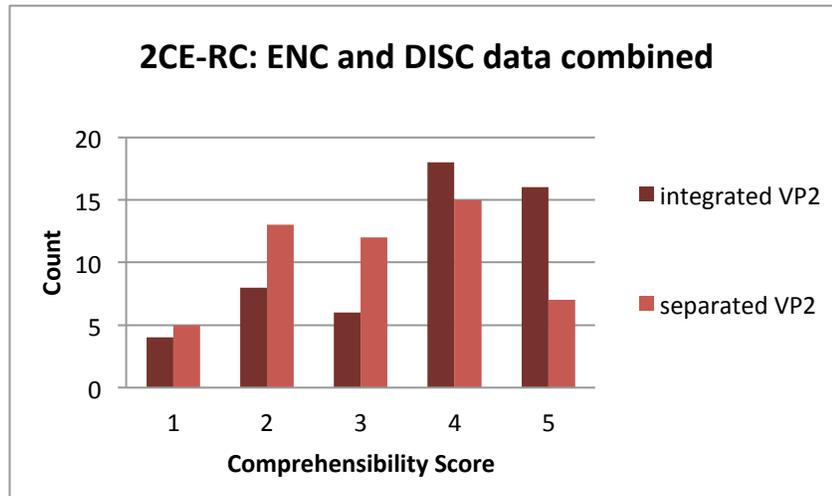


Figure 7: Distribution of comprehensibility ratings for items pronounced with integrated VP2 and separated VP2, for ENC and DISC items combined.

It is apparent, at least numerically, that a prosodic pattern with an integrated VP2 is generally conducive to comprehensibility, regardless of whether the stimuli had been intended to be prosodically ENCouraging or DISCouraging. On average, integrated VP2 phrasing was associated with higher comprehensibility scores (mean 3.65), in contrast to sentences where VP2 was phrased separately (mean 3.16). However, this numerical difference did not reach statistical significance in a mixed model regression with prosodic phrasing as the predictor, compared with an intercept-only model ($\chi^2(1) = 2.79, p = .095$). The indecisiveness of this result may be due to the very low power of this experiment (which had only 4 ENC/DISC sentence pairs in the 2CE-RC(length) category); but the trend observed here suggests that the effect of prosodic phrasing on comprehension may emerge more strongly in larger scale studies.

Overall, the findings of Experiment 1 are compatible with our general hypothesis that phrase lengths can impact prosodic contour assignment and that the appropriateness of the prosodic contour can improve comprehensibility, especially for the nested 2CE-RC structure which otherwise can be exceptionally difficult to process.

3.2 Experiment 2: In search of the ‘missing VP illusion’

Purpose

Experiment 1 provided some support for the claim that a prosodic manipulation, induced indirectly by the distribution of phrase lengths, can transform 2CE-RC sentences from their notoriously unintegratable phrase-list sequences into fairly normal-seeming sentences. Since this has not previously been noted, it deserves a more rigorous test than Experiment 1 could provide. That experiment was exploratory, designed to check a range of informal intuitions that had not been investigated before. What is needed in order to consolidate the prosody/syntax mismatch account of 2CE-RC sentence processing difficulty is an objective measure of when a perceiver has computed a complete and accurate syntactic tree – a measure more probing than the self-reports of comprehensibility in Experiment 1. At the same time, the power of the follow-up experiment should be more concentrated, focusing on the ENC/DISC phrase-length manipulation which showed robust results in Experiment 1.

As observed in section 1, an ungrammatical 2CE-RC sentence from which VP2 has been omitted is sometimes perceived as more (or no less) acceptable than the full sentence form with the required three VPs. Experimental evidence for this grammatical illusion is available in several languages (Gibson and Thomas, 1999, and Christiansen and MacDonald, 2009, for English; Gimenes, Rigalleau and Gaonac'h, 2009, for French; Bader, 2011, for German). This could provide the objective test we need, to assess accurate processing.¹³ Acceptance of a 2CE-RC sentence lacking VP2 would be a rather clear indication that a veridical syntactic tree structure had not been established. If it had, the absence of a VP2 should have been apparent (or else, if VP3 were pressed into service as VP2 in on-line processing of some examples, the absence of VP3 should have been apparent).

Our hypothesis for Experiment 2 was thus that the absence of VP2 should be easier to detect in ENC items than in DISC items, because the correct syntactic tree structure is more likely to be computed for the ENC items. The participants' task was to read the sentence first silently for comprehension, then aloud for recording, and then judge: "Is something missing from this sentence?"

Materials

There were 16 ENC/DISC sentence pairs, structured in the same way as the 2CE-RC(length) items of Experiment 1, presented either in full or with VP2 missing, counterbalanced across 4 lists. One quartet of target items is shown in Table 5. There were 25 filler sentences, which were similar to those of Experiment 1 except that an obligatory word or phrase was omitted from 12 of them. Examples of filler sentences with and without a missing element are also shown in Table 5.

¹³ We chose not to evaluate processing accuracy by means of post-sentence comprehension questions, in case reading the question might interfere with the participant's phonological memory of the target sentence.

Table 5: Examples of Experiment 2 materials

ENC Complete: The rusty old ceiling pipes that the plumber my dad trained fixed continue to leak occasionally.

ENC Missing: Admittedly, the rusty old ceiling pipes that the plumber my dad trained continue to leak occasionally.

DISC Complete: The pipes that the unlicensed plumber the new janitor kindly assisted tried to repair still leak.

DISC Missing: To no-one's surprise, the pipes that the unlicensed plumber the new janitor kindly assisted still leak.

Filler Complete: Professor Thompson knew the internationally famous literary critic giving the speech was a fraud.

Filler Missing: If Barbara wasn't crying because she lost her excellent exam notes, what was problem?

-----end of Table 5-----

Note that for the ENC and DISC target items, overall sentence lengths were matched between Complete and Missing versions by inserting at the beginning of the Missing sentences a number of words equal to those deleted from the Complete version. These words were set off from the main body of the sentence by a comma, to ensure that they did not intrude on the prosodic phrasing of the body of the sentence that followed. This was unlike the materials of Gibson and Thomas (1999) in which the Missing versions were systematically shorter than the Complete versions, which could have contributed to their positive evaluation.

Participants and Procedure

Participants were recruited through an invitation posted on websites. A web-based platform (www.limesurvey.org) was used for conducting the experiment. Participants were asked to read each visually displayed sentence silently first and then aloud while recording themselves. After the recording, they were asked to judge whether something was missing from the sentence. A practice session prior to the experiment provided examples. This was a more challenging task for participants than in Experiment 1 because we could not employ the familiarization protocol with this task. Building up a sentence stepwise despite the absence of one constituent from the final version was not feasible. However, just as for Experiment 1, we expected greater difficulty in assigning a natural-sounding prosodic contour to DISC versions than to ENC versions, and consequently a lower probability of computing the correct phrase structure of DISC versions. This could lead to greater incidence of the missing-VP illusion, i.e., more false acceptances for Missing DISC versions than for Missing ENC versions.

Results

Among the total of 49 participants who successfully completed the questionnaire there was a strongly bimodal distribution in which 23 participants (47%) rejected the majority of the Complete 2CE-RC sentences, ENC as well as DISC, as having “something missing”. This was unexpected. Standard reactions to 2CE-RC sentences such as (2) and (3) above suggest that, if anything, the Complete items might have been thought to have too many, rather than too few, constituents. These ‘CE-rejecters’ may have been responding to some property that was incidental to the purpose of the experiment, such as the absence of an overt complementizer *that* heading the inner relative clause (in all the 2CE-RC items). Or they may have regarded all doubly center-embedded items as unacceptable and used the “something is missing” response as the only means provided for rejecting them. Since our methodology was unable to assess the grounds for this across-the-board rejection of 2CE-RC sentences, we set the data from those participants aside. Balancing the participant numbers across materials lists resulted in a total of 24 ‘discriminating’ participants in the final analyses.

We computed the accuracy of detecting when a constituent was missing from the sentence. Unexpectedly, accuracy was very high in all conditions. There was no striking ‘missing VP illusion’. Table 6 summarizes the findings. Clearly, the Missing items were not being treated like the Complete items.

Table 6: Percent “something missing” judgments for Complete and Missing sentences, by ENCouraging and DISCouraging phrase lengths

	Complete	Missing
ENCouraging	10.4%	94.8%
DISCouraging	20.8%	88.5%

In other respects, the differences are in the directions predicted. For the Complete items, results from simple mixed-effects modeling with only phrase lengths as predictor show that the ENCouraging phrase lengths yielded significantly less rejection than the DISCouraging phrase lengths ($z = 2.29$, $p = .02$); this is in accord with the comprehensibility ratings of Experiment 1. The absence of VP2 (the *Missing* column in Table 6) was more detectable in 2CE-RC sentences which had phrase lengths ENCouraging helpful prosody than when phrase lengths DISCouraged that prosody, but this difference was non-significant ($z = 1.42$, $p = .15$).

The reason for this failure to robustly replicate the missing-VP2 effect remains unclear. The explicitness of the question posed to participants may have made them more alert to absent constituents than would a general judgment of grammaticality or acceptability. Also, it is possible that the web-based methodology, without ability to monitor participant behavior, allowed the ‘discriminating’ participants to study the sentences at length before responding. Another possibility of some interest is that reading the sentences aloud taps deeper syntactic processing than silent reading as in previous experiments on the missing VP effect. Replicating this study in a laboratory context, as we are currently planning, may help to disentangle these potential explanations.

Trained judges coded the prosody with which the sentences were spoken, as in Experiment 1. Phrase length manipulations predicted prosodic contours significantly, especially in the NP region. We tallied the percentage of items produced with a break after NP1 and nowhere else in the NP sequence, which we regard as optimal for syntactic parsing. We also determined the percentage of items pronounced with a break after NP2 (with or without another break in the NP sequence), which we regard as unhelpful for syntactic parsing. As shown in Figure 8, NP sequences in ENC items were pronounced with more breaks after NP1 only ($z=4.5$, $p < 0.001$), and with fewer breaks after NP2 ($z = - 6.1$, $p < 0.001$), compared to DISC items. There were also more DISC items than ENC items with breaks after *both* NP1 and NP2, in the manner of 'list intonation' (34.7% for ENC; 67.0% for DISC; $\chi^2 = 38.1$, $p < .001$). Differences in the VP region did not reach significance, contrary to Experiment 1. In discussion below we suggest that this reflects the procedural difference between the two experiments: with and without stimulus familiarization.

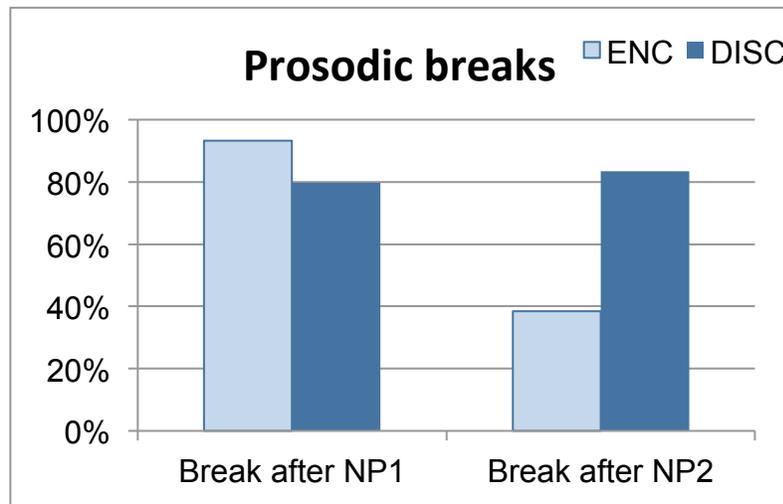


Figure 8: Percentage of prosodic breaks in the NP sequence, after NP1 only, and after NP2 (with or without a break after NP1), for ENC and DISC phrase lengths

Next, participants' judgments of sentences as complete or "missing something" were evaluated in relation to the prosody with which they had been produced, pooling ENC and DISC items; see Table 7. In this analysis, produced prosodies were grouped by presence/absence of a break after NP2, since this was the strongest factor observed.

Table 7: Relation between NP-region prosody and percent "something missing" judgments (regardless of ENC/DISC phrase length patterns)

Produced prosody	Complete	Missing
No break after NP2	6.2%	92.9%
Break after NP2	20.7%	91.7%

When participants produced a sentence with a break after NP2, their judgment of the grammatical status of the sentence was less accurate. Results from extended mixed-effects modeling, with phrase length and prosodic break location as predictors, showed that the best predictor of judgment accuracy was the produced prosody, regardless of whether the sentence was associated with phrase lengths pre-classified as ENC or DISC; see Table 8.

Table 8. Results from extended mixed-effects modeling with phrase

Contrast	Estimate	SE	z-Value	p-Value
ENC vs DISC (Complete)	1.11	0.49	1.61	.106
ENC vs DISC (Missing VP2)	0.68	0.48	0.90	.371
Break after NP1	0.10	0.53	0.19	.853
Break after NP2	-1.54	0.77	-2.00	.046 *
Break before VP3	0.23	0.55	0.43	.672

lengths and prosodic break locations as predictors of judgment accuracy

Discussion

Experiment 2 did not replicate the missing-VP effect which has been found previously (with materials that we regard as having DISCouraging phrase lengths, as in (3) above from Gibson and Thomas). But the results do confirm the link observed in Experiment 1 between phrase lengths and produced prosody, even in the absence of pre-familiarization with the materials. And its major finding bolsters the hypothesis, which was only marginally supported in the smaller-scale Experiment 1, that the produced prosody is what mediates causally between the phrase length manipulations and the ease of syntactic processing.

It is of interest that the most relevant aspect of the prosodic phrasing in Experiment 2 was the location of a boundary within the NP sequence. In Experiment 1, the most significant prosodic indicator was in the VP region: whether or not VP2 was included in a central prosodic ‘package’ consisting of NP2 NP3 VP1 VP2. But Experiment 1 employed the familiarization technique, which provided readers with an opportunity to practice pronouncing the 2CE-RC structure, so difficulties might arise only toward the end of building up the central prosodic unit. Specifically: the complete sequence of three NPs was already present in the penultimate sentence in the sequence (line 4 in (15) and (16) above), whereas the complete VP sequence was not presented until the final sentence, which was to be judged. By contrast, in Experiment 2, although items were read silently before recording there was no systematic rehearsal of the complex structure, so prosodic and parsing difficulties could arise earlier, during the NP sequence, specifically at the point of deciding whether or not to group NP1 and NP2 together prosodically. Grouping them would increase the probability of a break following NP2, which would preclude the optimal prosodic phrasing NP1 || NP2 NP3 VP1 VP2 ||VP3. To summarize: In Experiment 1 with familiarization, the prosodic grouping of the central constituents NP2 NP3 VP1 VP2 would most likely be blocked by inability to squeeze VP2 in at the end, whereas in

Experiment 2 without familiarization it would most likely be blocked by premature grouping of NP1 and NP2 together at the beginning.

4. Explanations

Our proposal has been that comprehension requires accurate building of syntactic tree structure, which must also be compatible with interface constraints. What remains to be explained is why the interface constraints apparently favor (for English) a 3-phrase prosody in which the upper RC is not divided. We will sketch here one account that lays the responsibility on constraints on prosodic phrasing, and one that invokes constraints on syntactic structure.

A possible prosodic account points to the fact that the 3-phrase prosody can provide maximum satisfaction of a cohesional constraint along the lines of Truckenbrodt's Wrap constraint (Table 2 above), which requires each syntactic phrase to be completely contained within a prosodic phrase. If this is applicable to the three clauses of the 2CE-RC structure,¹⁴ the only way to satisfy it fully would be for the whole sentence to constitute a single prosodic phrase, but that is impractical in terms of typical phrase lengths as we have observed. The 3-phrase prosody does the next best thing: it sacrifices Wrap in the main clause, but complies with it for the two embedded clauses. Alternate ways of dividing the word string would not help in this regard. For instance, dividing the string at the middle of the lower RC would violate Wrap at all three clausal levels. This line of explanation is worth pursuing. For example, the decline in judged acceptability of the 4-phrase prosody in (14) above might be attributable to the additional Wrap violation for RC1, which is split in two.

A possible syntactic explanation harks back to very early work on the syntax-prosody interface (Chomsky & Halle, 1968), in which an explanation was sought for the characteristic prosodic phrasing of recursive right-branching constructions as in (18).

(18) This is the cat || that chased the rat || that ate the cheese...

English generally aligns prosodic boundaries with the ends of syntactic phrases (Right-alignment with XP; see Table 2 above), but on the assumption that a restrictive RC modifies a noun (or N-bar) inside NP (see Note 5 above) there are no right XP brackets at the prosodic boundary locations in (18). Since the sentence must be divided into shorter chunks in some fashion, one way to account for these pre-RC boundaries would be to assume that limits on maximum phrase length outweigh the XP-alignment principle (i.e., they would rank higher in an Optimality Theory framework; see Selkirk, 2000). But Chomsky and Halle and subsequently Langendoen (1975) took a different tack. They maintained that the prosodic phrasing in (18) is fully aligned with syntactic phrasing. Where the prosody cannot match the syntax due to phrase length

¹⁴ As defined in Truckenbrodt (1995; though see also Truckenbrodt 2005), Wrap does not apply to adjuncts (which would include RCs) or to complete sentences (which would include the highest-level clause in a 2CE-RC construction), so this line of explanation may need to be based on some more general prosodic wrapping constraint.

constraints, the syntactic structure has to be brought into line with the prosody. This is effected by means of ‘readjustment rules’ which rearrange the constituents in the syntactic tree structure.

Typically, readjustment rules move a constituent A out of a larger constituent B, and adjoin A as a sister of B. In (18), for example, string-vacuous application of a relative clause extraposition rule could extract each RC from the noun phrase in which it originates and raise it to become a sister of that noun phrase. As a result, there would indeed be a syntactic right-edge bracket for the prosodic phrasing to align with preceding each RC in (18). A similar analysis of the center-embedded RC construction (Figure 1) would lift RC1 out of the noun phrase in which it originates as a modifier of NP1 and attach it as a sister to that noun phrase,¹⁵ permitting a prosodic boundary between the two constituents and thus licensing the 3-phrase prosody.

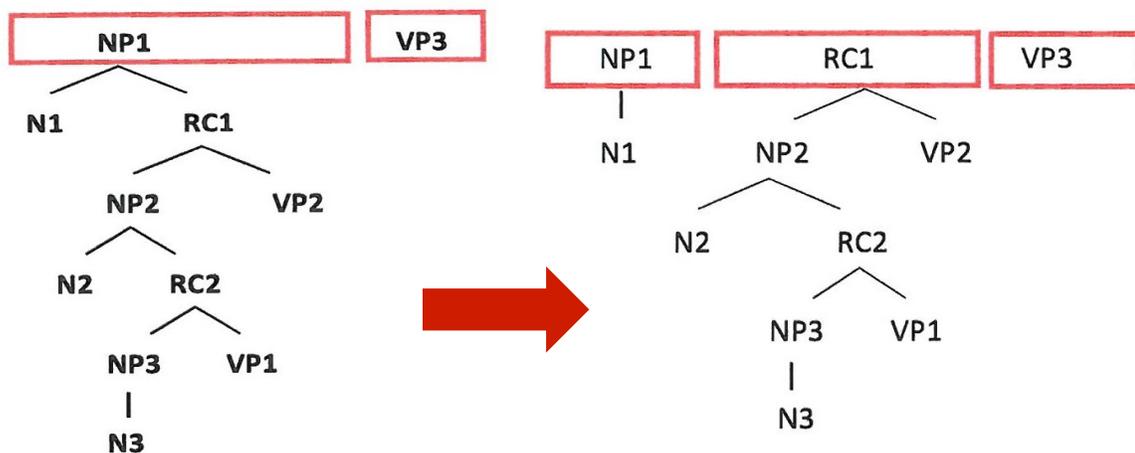


Figure 9: String-vacuous extraposition of RC1 in the 2CE-RC construction

<ADJUST LAYOUT>

Is string-vacuous RC-extraposition legitimate? Certainly, non-vacuous RC-extraposition is possible, as in examples (19) and (20) with interpretations as indicated.

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

(20) 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].

¹⁵ Where the extraposed RC1 attaches in the tree structure needs to be established. Figure 9 is not precise in this regard. Attachment to the highest syntactic node, as sister to NP1 on its left and to VP3 on its right, would be illegitimate in theories that insist on binary syntactic branching. However, if the extraposition occurs at the level of PF (phonological form), it might not be subject to a binarity constraint. See Callahan (2013) for fuller discussion of the syntax of extraposed RC constructions in subject position.

String-vacuous movement is by its nature difficult to demonstrate. It has even been argued that it should be banned from linguistic theory altogether (den Dikken and Lahne, 2013). However, Wagner (2010) has developed a positive argument in favor of vacuous extraposition of RCs, noting that when the head noun is an idiom chunk (e.g., *Mary praised the headway that John made*), extraposing the RC away from the noun is unacceptable and so is inserting a prosodic boundary between the head noun and the adjacent RC. This parallelism may suggest that a boundary following a head noun is a sign that vacuous RC-extraposition has occurred.

A benefit of the vacuous extraposition analysis is that it offers a straightforward syntactic explanation for the degraded acceptability of the 4-phrase prosody for 2CE-RC sentences. The 4-phrase prosody, with its separated VP2 (i.e., a boundary between the subject and predicate of the upper RC, as in (14) above), would have to come about by vacuous extraposition of VP2 out of the relative clause RC1 in which it originated. But if we may judge on the basis of non-vacuous extraposition, that would be syntactically illegitimate: a finite VP cannot be extraposed. Example (21), with the interpretation as indicated, is strongly ungrammatical.

(21) *The children_{RC}[who the principal e_i yesterday morning] VP_i[had scolded] were weeping.

As a result, there is no legitimate readjusted syntactic structure for the 4-phrase prosodic analysis to align with. No matter how tempting the parser may find that prosodic phrasing on-line in order to break up over-heavy prosodic phrases, it does not aid sentence comprehension because it implies an illicit syntactic structure.¹⁶

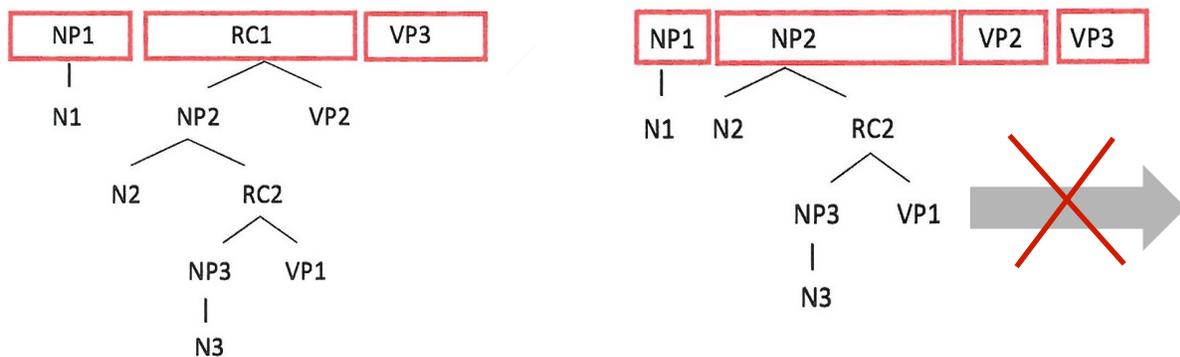


Figure 10: Illegitimate string-vacuous extraposition of VP2 in the 2CE-RC construction

<<Adjust font size etc., as for Figure 9.>>

¹⁶ Given that VP2 cannot extrapose, would it be permissible to skip down further and perform readjustment within NP2 by vacuously extraposing RC2? (In our materials that would be impossible due to absence of an overt relative pronoun/complementizer in RC2, but the question can be raised more generally.) RC2 would still be trapped inside RC1, so this would yield a multi-level phrasing NP1 || NP2 | RC2 | VP2 || VP3. This has been reported for German (Féry and Schubö, 2010), and it will be worth probing for in English, though it may be blocked if English tolerates a break between a restrictive RC and the head it modifies (i.e., vacuous RC extraposition if Wagner's argument is correct) only when it is motivated by length considerations or focus (unlike German, which favors a break before restrictive RCs generally; Augurzky, 2006).

5. Conclusion

The experiments we have reported here are just a beginning. They need to be followed up by more substantial studies that can provide insight into the linguistic and psycholinguistic mechanisms at work. But there are already promising indications that a significant factor in the near-unparsability of many doubly center-embedded sentences (at least in English) is the radical misfit between their strongly hierarchical syntactic structure and the prosodic phrasing induced by typical phrase lengths. It can be argued that this prosodic approach to syntactic center-embedding contributes to the explanation of all three well-known peculiarities of 2CE-RC processing:

- I. Unusually difficult comprehension. The heavily nested 2CE-RC construction conflicts more radically with the ‘flattening’ tendency of prosodic phrasing (Myrberg 2013) than other multi-clausal syntactic constructions do.
- II. Improvement if NP3 is a pronoun. A pronoun is short and usually unstressed, so it helps to slim down the RC1, leaving more room to include NP2 and VP2 within the middle prosodic package.
- III. Apparent improvement if VP2 is absent (the “missing-VP effect”). This is well-attested generally, though our Experiment 2 did not observe it. When the central phrase of the optimal 3-phrase prosody is too tightly packed to be able to include VP2, or if the NP sequence is mis-phrased so that there is no attachment point for VP2, the parser is better off ignoring it.

This does not mean that other factors such as working memory overload make no contribution to the difficulty of these sentences. Clearly, they contain a great deal of material to be remembered, and unlike right-branching constructions their word order does not allow subjects and predicates that belong together semantically to be composed together without delay. However, the current findings suggest that in addition to this complexity of semantic relations there is a purely syntactic disability in building a coherent tree structure for a 2CE-RC sentence when phrase lengths and hence prosodic structure are fighting against it.

Center-embedded constructions may thus once again become a tool for psycholinguistic research, providing a rich source of data in this case for elucidating details of the syntax-prosody interface, in English and also cross-linguistically. We are particularly interested in exploring the idea that when put under extreme pressure by this construction, the interface negotiations favor different solutions in different languages. There is a growing body of work in which the various interface constraints formulated by phonologists are being tested in psycholinguistic performance. It has become evident that phrase length constraints can be at least as powerful as alignment constraints and sometimes outrank them. It is well-established, beginning with Lehiste (1972) and by many studies since, that the overt prosody of spoken sentences has an impact on syntactic parsing: it can facilitate the comprehension of unambiguous sentences, and bias the interpretation of ambiguous ones. That similar effects can be induced by manipulating the lengths of phrases in visually presented sentences read silently (or aloud) has also been amply

demonstrated in recent years, by Hirose (2003) for Japanese, Lovrić (2003) for Croatian, Fernández et al. (2003) for Spanish and English, among others. Thus the present finding that the comprehensibility of doubly center-embedded sentences can be improved by such means fits very naturally into this broader research paradigm.

References

Augurzky, Petra (2006) *Attaching Relative Clauses in German - The Role of Implicit and Explicit Prosody in Sentence Processing*. MPI Series in Human Cognitive and Brain Sciences, 77, Leipzig.

Bever, Thomas G. (1970) The cognitive basis for linguistic structures. In R. Hayes (Ed.), *Cognition and language development* (pp. 277-360). New York: Wiley & Sons, Inc.

Bever, Thomas G. (1988) The psychological reality of grammar: a student's eye view of cognitive science. In W. Hirst (Ed.), *The making of cognitive science: A festschrift for George A. Miller*. Cambridge University Press.

Blumenthal, Arthur (1966) Observations with self-embedded sentences. *Psychonomic Science* 6.10, 203-6.

Callahan, Taletha (2013) "Within-DP Relative Clause Extraposition" Unpublished m.s., The Graduate Center, City University of New York.

Chomsky, Noam and Miller, George A. (1963) Introduction to the formal analysis of natural languages, in *Handbook of Mathematical Psychology* Volume II, R. Duncan Luce, Robert R. Bush and Eugene Galanter (eds.) John Wiley and Sons. Pp.269-321.

Christiansen, Morten H. and MacDonald, Maryellen C. A Usage-Based Approach to Recursion in Sentence Processing, *Language Learning* 59:Suppl. 1, December 2009, pp. 126–161.

den Dikken, Marcel and Lahne, Antje (2013) The locality of syntactic dependencies. Chapter 11 of *The Cambridge Handbook of Generative Syntax*. Cambridge University Press, Cambridge, UK.

Fernández, Eva M., Bradley, Dianne, Igoa, José Manuel and Teira, Celia (2003) Prosodic phrasing in the RC-attachment ambiguity: Effects of language, RC-length, and position. Paper presented at the 9th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Glasgow, UK.

Féry, Caroline and Schubö, Fabian (2010) Hierarchical prosodic structures in the intonation of center-embedded relative clauses. *The Linguistic Review* 27, 293-317.

Fodor, Jerry A., Bever, Thomas G and Garrett, Merrill F. (1974) *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. McGraw-Hill.

Fodor, Jerry A. and Garrett, Merrill F. (1967) Some syntactic determinants of sentential complexity. *Perception & Psychophysics*, 1967, Vol. 2 (7), 289-296.

Fodor, Janet Dean 2013 Pronouncing and comprehending center-embedded sentences, Chapter 9 of *Language Down the Garden Path: The Cognitive and Biological Basis of Linguistic Structures (Oxford Studies in Bilingualism)*. Montserrat Sanz, Itziar Laka and Michael K. Tanenhaus (eds.). Oxford University Press.

Fodor, Janet Dean and Nickels, Stefanie (2011) Center-embedded Sentences: Phrase length, prosody and comprehension, poster presented at AMLaP, September 2007, Paris, France.

Fodor, Janet Dean and Schott, Esther (2013) Prosody induced by phrase lengths facilitates syntactic processing in reading, poster presented at AMLaP September 2013, Marseille, France.

Fox, Barbara A. and Thompson, Sandra A. (2007) Relative clauses in English conversation: relativizers, frequency and the notion of construction. *Studies in Language* 31.2: 293-326.

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

Frazier, Lyn (1985) Syntactic complexity, Chapter 4 of *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, David R. Dowty, Lauri Karttunen and Arnold M. Zwicky (eds.). Cambridge, UK: Cambridge University Press. (Pp.129–189).

Gee, James P. and Grosjean, François (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.

Ghini, Mirco (1993) Phi-formation in Italian: A new proposal. *Toronto Working Papers in Linguistics*, Volume 12 Number 2, pp.41-78.

Gibson, Edward and Thomas, James (1999) Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical (1999) *Language and Cognitive Processes* 14 (3), 225–248.

Gimenes, Manuel, Rigalleau, François, & Gaonac'h, Daniel (2009). When a missing verb makes a French sentence more acceptable. *Language and Cognitive Processes*, 24, 440–449.

Hudson, Richard (1996) The difficulty of (so-called) self-embedded structures, in *University College London Working Papers in Linguistics* Vol 8, 283-314.

Hirose, Yuki (2003) Recycling prosodic boundaries. *Journal of Psycholinguistic Research*, Vol. 32, No. 2, 167-195.

Lieberman, Mark (2013) "English prosodic phrasing", *Language Log* September 21, 2013.

Lovrić, Nenad (2003) Implicit prosody in silent reading: Relative clause attachment in Croatian, unpublished Ph.D. dissertation, The Graduate Center, CUNY.

Miller, George A. and Chomsky, Noam (1963) Finitary models of language users, in *Handbook of Mathematical Psychology* Volume II, R. Duncan Luce, Robert R. Bush and Eugene Galanter (eds.) John Wiley and Sons. Pp.419-491.

Miller, George A. and Isard, Stephen (1964) Free recall of self-embedded English sentences. *Information and Control* Volume 7, Issue 3, Pp. 292-303.

Myrberg, Sara (2013) Sisterhood in prosodic branching *Phonology* 30,73-124.

Schott, Esther (2012) The Influence of Prosody on Comprehending Complex English Sentences, unpublished Diplomarbeit, Ludwig-Maximilians University, Munich, Germany

Selkirk, Elisabeth O. (1981) On the nature of phonological representation, in J. Anderson, J. Laver and T. Meyers, eds., *The Cognitive Representation of Speech*, Amsterdam: North Holland.

Selkirk, Elisabeth (1995) The prosodic structure of function words. In *University of Massachusetts occasional papers 18: Papers in Optimality Theory*, 439–469. GLSA, University of Massachusetts, Amherst.

Selkirk, Elisabeth (2000) The interaction of constraints on prosodic phrasing. In M. Horne (ed.) *Prosody: Theory and Experiment*, Kluwer Academic Publishers, pp.231-261.

Truckenbrodt, Hubert (1995) On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30, 219-255.

Truckenbrodt, Hubert (2005) A short report on intonation phrase boundaries in German, *Linguistische Berichte* 203, 273-296.