

# John Benjamins Publishing Company



This is a contribution from *Studies in Language* 31:2  
© 2007. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

Tables of Contents, abstracts and guidelines are available at [www.benjamins.com](http://www.benjamins.com)

# Relative Clauses in English conversation

## Relativizers, frequency, and the notion of construction\*

Barbara A. Fox and Sandra A. Thompson

University of Colorado, Boulder / University of California, Santa Barbara

This paper is a usage-based study of the grammar of that set of English Relative Clauses with which a relativizer has been described as optional. We argue that the regularities in the use of relativizers in English can be seen as systematically arising from pragmatic-prosodic factors, creating frequency effects, resulting in some cases highly grammaticized formats: the more the Main Clause and the Relative Clause are integrated with each other, that is, approach MONOCLAUSAL status, the more likely we are to find no relativizer used; conversely, the more separate the two clauses are, the more likely we are to find an overt relativizer. These findings have led us to suggest that the more monoclausal combinations have become unitary storage and processing chunks. We thus see these findings as a contribution not only to our understanding of Relative Clauses, but to our understanding of syntactic organization in general and of the nature of the grammatical practices in which speakers engage in everyday interactions.

### 1. Introduction

In this paper, our goal is to show that the use of relativizers in English conversation follows patterns that are sensitive to a range of pragmatic and prosodic factors, and that the regularities underlying these patterns provide a window into the organization of Relative Clauses in the heads of English speakers, contributing to a fresh understanding of the nature of the grammatical practices in which speakers engage in everyday interactions.

The phenomenon we explore in this paper is illustrated by examples (1)–(4) from our database (we put the head NP in square brackets, underline the Relative Clause and boldface the relativizer, if present):

- (1) I have [one] that you can use
- (2) there was [something] we needed
- (3) this is [the first time] we've seen this house
- (4) do you remember exactly [the road] I'm talking about?

In example (1), the NP-Rel (co-referential with the head noun *one*) participates in the Relative Clause as the object of the verb *use*; object relatives are noted for the apparent optionality of the relativizer (in contrast to, for example, subject relatives and genitive relatives). In this example, the relativizer *that* is used. In example (2), on the other hand, the object relative is not introduced with a relativizer. Example (3) illustrates an adverbial Relative Clause without a relativizer, while (4) provides an instance of an oblique Relative Clause with no relativizer.

This pattern of presence or absence of the relativizer in object, certain oblique, and adverbial Relative Clauses (hereafter ORCs) is often described as indicating the optionality of the relativizer in these types of clauses (e.g., Berk 1999; Givón 1993; Quirk et al. 1985). Bolinger suggests that “avoidance of ambiguity” may be a factor: “the reason why *that* ... cannot normally be omitted is ... because without it the constituents are hard to identify” (1972: 11).

There has been attention in the literature to this variability. A number of corpus studies, pioneered by Quirk 1957, have investigated the choice among relativizers, presenting data on such social variables as genre, written vs. spoken data, and British vs. American English, and such internal variables as position relativized on and features of the head NP and of the Relative Clause (Biber et al. 1999; Guy and Bayley 1995; Quirk et al. 1985; Tottie 1995, 1997; Tottie and Lehmann 1999, 2002). Of particular relevance to our study of ORCs in conversational English is Tottie 1995; using a VARBRUL analysis of the factors that favor zero relatives in ORCs in present-day written British and American English, she finds, as we did, that (1) definiteness of the Head NP can be dismissed as an important factor, and that (2) personal pronouns as subjects of the ORC are well correlated with a zero relativizer. We will return to this second point below.

Ariel 1999, discussing resumptive pronouns in Hebrew, suggests that accessibility theory might provide an account of the choice in English among *who*, *which*, *that*, and  $\emptyset$ . Jaeger and Wasow (to appear), working with a large spoken English corpus, provide additional evidence for the role of accessibility in the use vs. non-use of relativizers. Those of our findings which relate to NPs in Relative Clause constructions converge with theirs, though our interest in this study is with what we will call MONOCLAUSALITY.

A recent work exploring usage-based explanations for the use vs. non-use of a relativizer in clauses such as those in (1)–(4) above is Temperley 2003,

who does not mention either Bolinger 1972 or Biber et al. 1999, but finds both “ambiguity avoidance” and “anaphoricity”<sup>1</sup> to be factors influencing the use of a relativizer in written English texts. We will see below that his findings are compatible with the patterns we have discovered in the use of the relativizer in conversational English.

Intriguingly, and essentially unnoticed in the literature, Jespersen (1933: 359) proposes a classification of Relative Clauses which assumes that clauses with and without a relativizer should be analyzed (and presumably understood to be cognitively represented), not in terms of variation or optionality, but as having been categorized by speakers into different categories:

- (a) clauses with one of the two *wh*-pronouns, *who* and *which*;
- (b) clauses without any connecting word: contact clauses,

and

- (c) clauses with one of the connecting words, *that*, *as*, *but*.

It is Jespersen’s “contact clauses” that are exemplified in (b) and (c) above, without relativizers. He offers this characterization of them:

Relative clauses without any connecting word are here called *contact clauses*, because what characterizes them is the close contact in sound and sense between the clause and what precedes it: in sentences like “this is the boy we spoke of”, and “he falls in love with all the girls he sees”, the words “the boy” and “all the girls” are felt to be just as intimately connected with what follows as what precedes them. (Jespersen 1933: 360)

In this paper, we will show that this insight is beautifully borne out by the findings that emerge from our data. Thus, in the present paper we suggest, as Bolinger 1972, Thompson 2002, and Thompson and Mulac 1991a and 1991b do for the English complementizer *that*, that it is more revealing to view these patterns of use of the relativizers in English as systematically arising from pragmatic-prosodic factors (Jespersen’s “sound and sense”), creating in some cases highly grammaticized formats. Our results support this shift from accounting for English Relative Clause usage in terms of optionality to providing an explanation in terms of different categories of constructional patterns.

Specifically, we will demonstrate how much these patterns can show us about the nature of syntactic organization. That is, when we examine the distribution of relativizers in English ORCs, we find a striking regularity, as Jespersen’s notion of “contact clauses” suggests: the more the Main Clause and the ORC are integrated with each other, that is, approach MONOCLAUSAL status,

the more likely we are to find no relativizer used; conversely, the more separate the two clauses are, the more likely we are to find an overt relativizer. These findings have led us to suggest that the more monoclausal combinations may have become unitary storage and processing chunks. We problematize the notion of a **RELATIVE CLAUSE CONSTRUCTION** by suggesting that there is a continuum of organization of utterances by clausality and lexical specificity, a finding which provides a new and more subtle understanding of the role of usage on the speakers' cognitive representations of Relative Clauses. This approach focuses on lexical specificity, low-level formats, and more abstract patterns, as well as the networks of resemblances among these families of related and divergent types. We thus see these findings as a contribution not only to our understanding of Relative Clauses, but to our understanding of syntactic organization in general.

The remainder of the paper is organized as follows. Section 2 describes our data and methods. Section 3 presents the findings of our quantitative study of pragmatic-prosodic factors associated with the presence or absence of a relativizer. Section 4 provides an account for our quantitative findings, and Section 5 explores the implications of our findings for the notion of **CONSTRUCTION**. Section 6 summarizes the highlights of the paper.

## 2. Data and methods

Our data were culled by examining 36 audio- and video-taped American English conversations among people who were friends and family members of each other, ranging in length from 5 minutes to 1 hour. The conversations exhibit diversity in the age of the participants, region of the country, and date of recording.

Within this corpus we identified 300 utterances that both authors agreed could be non-controversially considered restrictive Relative Clauses.<sup>2</sup> From our collection of Relative Clauses we found 195 ORCs, that is examples of object, oblique, and adverbial Relative Clauses in which we as native speakers felt that the relativizer could have been present or absent (that is, in some sense "optional"). We excluded all instances of subject Relative Clauses, on the grounds that relativizer usage is quite different for subject relatives (and in fact we had no instances of subject relatives without the relativizer; but see Lambrecht 1988 and Lehmann 2002 for useful discussion); we also excluded all genitive Relative Clauses, and various subtypes of oblique relatives in which the relativizer appears to be required.<sup>3</sup> We included instances in which the Head NP did not participate in any Main Clause (e.g., *according to [one of the signs] I saw.*)<sup>4</sup>

These 195 ORCs were then coded for a wide range of lexical, syntactic, pragmatic and prosodic characteristics. We found that 7 of these characteristics strongly correlate with the presence/absence of a relativizer. In the next section we report on these correlations.

### 3. Distribution of relativizers

In this section we provide our findings for the distribution of relativizers (including *that*, *where*, *when*, and *who*)<sup>5</sup> in the 195 ORCs in our collection.

In our collection of 195 ORCs, 117 (60%) of the utterances showed no overt relativizer. This confirms what earlier research has established for most English data, and is an important observation: in our quantitative study we have found that  $\emptyset$ -relativizer is the more preferred choice for American English conversation.

What are the factors which correlate with the occurrence or non-occurrence of a relativizer in our ORCs? Here we provide evidence to show that the instances without the relativizer have a number of special properties.

We found that 7 characteristics correlate with the presence/absence of relativizers in our ORCs. And as we discuss in Section 4, all of these variables are part of a larger parameter: whether the Main Clause and ORC tend towards being highly integrated, to the point of being nearly monoclausal. The 7 variables are:

- Emptiness of Head NP
- Complexity of Main Clause
  - length of Head NP
  - Copular Main Clause
- Unique Head
- Length of ORC verbal expression
- Subject of the ORC
- Breaks between Head NP and ORC
- Rhythmic relationship between Head NP and ORC

We now turn to a discussion of each variable.

#### *Variable 1: Emptiness of Head NP*

We define Empty Head NPs to be those which are not lexically specific and/or which index generic groups or sets of individuals or objects. Examples of Empty Head NPs include *time*, *way*, *thing*, and *all*. We have found a correlation

between Empty Head NPs and  $\emptyset$ -relativizer. Examples (5)–(7) illustrate Empty Head NPs with  $\emptyset$ -relativizer:

- (5) [all] she wants to do is sleep  
 (6) [the next time] I have a baby, ... Trace will probably be=, ... about three?  
 (7) is there [any way] he could like, meet us in Great Falls or something?

Table 1 gives the numbers for Empty Head NP and presence or absence of relativizer. Empty Head NPs favor  $\emptyset$ -relativizer 70% to 30%, whereas for non-Empty Heads, the ratio is 50%/50%.<sup>6</sup> This pattern is significant ( $\chi^2= 8.15$ ,  $df=1$ ,  $p = .004$ ).<sup>7</sup>

**Table 1.** Empty/Non-Empty Heads and Relativizer

	Empty Head	Non-Empty Head	Total
+ Relativizer	20 ( 27%)	58 ( 48%)	78
– Relativizer	55 (73%)	62 (52%)	117
Total	75	120	195

*Variable 2: Complexity of Main Clause*

The complexity of the Main Clause is a strong predictor of the occurrence of a relativizer. We took two measures of Main Clause complexity.<sup>8</sup> First, we measured the length of the Head NP in number of words. For example, if we measure the length of the head NP for utterances (8) and (9) below, we find that the Head NP in example (8) has 5 words in it; in example (9), on the other hand, the Head NP has only two words in it.

- (8) [this pair of suede pants] that I got,  
 (9) [the weight] I should be at

Table 2 shows that there is a substantial difference between the average number of words in the Head NPs with and without a relativizer, such that longer Head NPs tend to co-occur with a relativizer:

**Table 2.** Average words in Head and Presence/Absence of Relativizer

	Average Words in Head
+ Relativizer	2.4
– Relativizer	1.1

The second measure we took of Main Clause complexity was whether or not the Main Clause was a copular clause, that is, a clause with a copula verb in it. This meant that we counted Predicate Nominal clauses, as well as Predicate Adjective, Predicate Oblique clauses, and Existential clauses as copular clauses. The Head NP, then, could be the:

- Predicate Nominal itself
- Subject of a Predicate Nominal, Predicate Adjective, or Predicate Oblique clause
- Post-copular Subject of an Existential clause

Copular clauses are highly frequent in English conversation (Thompson and Hopper 2001), and are structurally the most minimal possible Main Clauses in English: they have the lowest possible Transitivity (Hopper and Thompson 1980; Thompson and Hopper 2001), and the least possible semantic verbal content. We will see that these factors figure importantly in the inferences we can make about the way ORCs are stored and used by English speakers, but for now, the point is that copular Main Clauses favor  $\emptyset$ -relativizer. Let's consider each of the types of copular Main Clauses in turn, to be summarized in Table 6.

First, we illustrate Main Clauses in which the Head NP is a Predicate Nominal with these examples:

- (10) .. that's [something] I'm not used to.  
 (11) that's just [a bed] they move around  
 (12) it's [the one] they're working on,

Table 3 shows that Predicate Nominal Heads prefer  $\emptyset$ -relativizer:

**Table 3.** PN Heads and Relativizer<sup>a</sup>

	Predicate Nominal Head
+ Relativizer	12 (23%)
– Relativizer	41 (77%)
Total	53

<sup>a</sup> The Chi-Square results for Tables 3-5 are given in the summary Table 6.

Second, we show that Head NPs that are the subjects of copular Main Clauses also tend to occur with  $\emptyset$ -relativizer. Examples (13)–(15) illustrate Head NPs that are the subjects of copular clauses:



- (13) [all] she did was sleep  
 (14) [the only thing] you can do is be the best you can.  
 (15) [everything] we do is like that<sup>9</sup>

Table 4 shows that subject Heads of copular clauses prefer  $\emptyset$ -relativizer:

**Table 4.** Subject Heads of copular clauses and Relativizer

	Subject Heads of copular clause
+ Relativizer	5 (23%)
– Relativizer	17 (77%)
Total	22

Third, we considered Heads that are the post-copular subjects of Existential Main Clauses, as illustrated by (16)–(17):

- (16) there's not [much] they can do  
 (17) is there [any way] he could meet us?

Table 5 suggests that Heads that are the subjects of Existential Main Clauses have a slight preference for an overt relativizer. However, if we examine the individual cases, all but two of the eight instances of overt relativizer are triggered by other factors: five of them have breaks between the Head NP and the ORC, and one has another ORC preceding the ORC we're concerned with (see discussion of these variables below). The apparent preference for an overt relativizer is thus produced by other factors.

**Table 5.** Existential Heads and Relativizer

	Existential Head
+ Relativizer	8 (57%)
– Relativizer	6 (43%)
Total	14

Table 6 presents a summary of the facts about Head NPs in copular Main Clauses in our database. What Table 6 shows is that the percentage of Head NPs in copular Main Clauses with  $\emptyset$ -relativizer is much higher than for Head NPs in non-copular Main Clauses (including instances without Main Clauses); these differences are significant ( $\chi^2 = 8.04$ ,  $df = 1$ ,  $p = .004$ ):

**Table 6.** PN/Subj. of Cop. Heads compared to Other Heads and Relativizer

	PN /Subj. of Cop. /Exist Head	Other Heads	Total
+ Relativizer	25 (28%)	52 (49%)	77
– Relativizer	64 (72%)	54 (51%)	118
Total	89	106	195

In sum, then, the complexity of the Main Clause, measured both by the sheer number of words in the Head NP and by whether or not the Head NP occurs in a copular Main Clause, can be seen to play a central role in whether or not a speaker will use a relativizer.

*Variable 3: Unique Head NP*

Unique Head NPs include superlatives and nouns with *only* and *first*. This category is illustrated in (18)–(20):

- (18) that's [the only place] you can go at night  
 (19) that was [the .. u=glieſt set of shoes] I ever saw in my life  
 (20) that's [the first compliment] I've got in a long time

In our data, Unique Heads categorically take  $\emptyset$ -relativizer, as shown in Table 7:

**Table 7.** Unique Heads and Relativizer

	Unique Head
+ Relativizer	0 (0%)
– Relativizer	23 (100%)
Total	23

We've now considered three variables relating to the Main Clause in an ORC : the emptiness of the Head NP, the complexity of the Main Clause, and whether or not the Head NP is Unique. Next we turn to two variables that relate to the ORC itself.

*Variable 4: Length of verbal expression of the ORC*

Under Variable 2 above, we considered the complexity of the Main Clause. Now we consider one aspect of the complexity of the ORC itself: the length of the verbal expression in the ORC. As with the length of the Head NP, we

counted number of words, this time the number of words in the verbal expression, that is, excluding the subject but including adjuncts.

Thus, for example, we can see that in (21) the verbal expression consists of only one word, *got*, and the speaker used no relativizer, while in example (22), the verbal expression consists of 5 words, and the speaker did use a relativizer:

- (21) that's [the one] you got  
 (22) [all the stuff] that Vicki's told me that she pulls

We found that this pattern was consistent in the data: ORCs with  $\emptyset$ -relativizer have shorter verbal expressions, as shown in Table 8:

**Table 8.** Average number of words in ORC 'verbal expression' and Relativizer

	Average Words in ORC 'VE'
+Relativizer	3.77
- Relativizer	2.91

#### *Variable 5: Subject of ORCs*

The other variable relating to the ORC itself that we considered was the nature of the subject of the ORC. As Biber et al. 1999 and Tottie 1995 have found, the type of NP in the ORC subject plays a role in the use of the relativizer.

In Tottie's (1995) study of ORCs in written English, she found a continuum: 77% of her ORCs with subjects that were personal pronouns had a  $\emptyset$ -relativizer. When the relative subject was a definite NP, only 52% had a  $\emptyset$ -relativizer, and when the subject was an indefinite NP or a proper noun, only 30% had a  $\emptyset$ -relativizer.

Biber et al. 1999 comment on the relationship between subject type and use of relativizer in the ORCs in their database, which included four registers: academic writing, news, fiction, and conversation:

When the subject of the Relative Clause is a pronoun, 60–70% of Relative Clauses have the relativizer omitted. When the subject of the Relative Clause is a full NP, 80–95% of Relative Clauses retain the relative pronoun. Surprisingly, these grammatical constraints hold across all four registers. (Biber et al. 1999: 620)<sup>10</sup>

We illustrate this point from our database. In (23) and (24), the ORC subject is *I*, the most frequent subject in English conversation; note that neither utterance has a relativizer:

- (23) I have [two cats] I'd like to turn in to the Humane Society  
 (24) do you remember exactly [the road] I'm talking about?

In (25), in contrast, the subject of the ORC is *Tom*, a lexical NP, and here we do find a relativizer.

- (25) oh and you know [another thing] that Tom had the audacity to bitch about?

In fact, whereas Biber et al. 1999 considered only the distinction between pronouns and lexical NPs, we found a continuum, though a different one from Tottie's. But all these results find an explanation in terms of frequency: the more frequent the subject is in English conversation in general, the less likely the speaker is to use a relativizer. In other words, the data show that, although the distinction between pronouns and Lexical NPs is important, we can make a further distinction for conversational English between *I* and other pronouns. So the frequency of use of the relativizer increases according to whether the subject of the ORC is *I* vs. other pronoun vs. Lexical NP, as shown in Table 9:

Table 9. Subjects of Relatives and Relativizer<sup>a</sup>

	I	Other Pro	Lex NP	Total
+Relativizer	27 (34%)	42 (41%)	6 (55%)	75
- Relativizer	52 (66%)	61 (59%)	5 (45%)	118
TOTAL	79	103	11	193 <sup>b</sup>

<sup>a</sup> The results of a Cochran-Mantel-Haenszel test failed to show significance, so we acknowledge that this variable does not seem to be as critical in the use of the relativizer as the others; nevertheless the trend is visible and suggestive

<sup>b</sup> Total = 193 because 2 RCs had uncodable Heads. Table 9 indicates that if the subject of the ORC is *I*,  $\emptyset$ -relativizer is preferred (66%), while with other pronoun ORC subjects, speakers still prefer  $\emptyset$ -relativizer, but not as strongly (59%). With lexical ORC subjects, speakers display a slight preference for an overt relativizer (55%).

Analogous distributions are taken by Temperley 2003 as evidence in favor of “ambiguity avoidance” as a factor in the use of relativizers in written English, since pronouns could not be misparsed as continuations of the same NP as the Head NP, but lexical nouns could. Our preliminary study of prosodic factors in ORCs suggests to us that “ambiguity avoidance” may not be a strong factor in spoken English, since stress and intonational patterns will typically prevent combinations which could conceivably be misparsed in writing from being misparsed in spoken interactions. For Temperley, these findings also suggest “anaphoricity” as a factor in the use of relativizers, since relativizers tend to occur less frequently with RC subjects that are anaphoric with some referent earlier in the discourse. As we will see below, our analysis, which focuses on

the integration of Main and Relative Clause, provides an alternative — and we believe more compelling — account of these properties.

*Variable 6: Breaks between Head NP and ORC*

The next variable that we will consider in our quest to understand the use of relativizers in our ORCs in conversation has to do with the production of instantiations of the entire ORC. In our database, there are about twenty instances of a prosodic break, a repair, or another ORC or an adverb between the Head NP and the ORC. In every case, a relativizer categorically appears. Let's examine each of these briefly.

a. Prosodic breaks

Prosodic breaks between the Head NP and the ORC can be seen in (26) and (27), where commas indicate intonation unit breaks (Du Bois et al. 1993):

- (26) th- there's [a white line],  
that you go by,  
(27) .. (H) So we have [this frozen horse hoof],  
that we have to start out on,

b. Repairs

If speakers repair the Head NP in an instantiation of a ORC in our data, they always use a relativizer:

- (28) a hassock is [a- an item] that you could sit on or put your feet on  
(29) you also went through [the one], [the number] that Jerry gave me

c. Another ORC between Head NP and ORC

When speakers use two ORCs together, the second one will show a relativizer:

- (30) there was [a boy] that played the trombone that he kind of knew

d. Adverbs between Head NP and ORC

If a speaker uses an adverb between the Head NP and the ORC, a relativizer always appears:

- (31) you have [a home] here that you could rent  
(32) this guy has not done [anything] yet that I understand  
(33) there's [three courses] already that I'm not going to do well in

Thus, in our data, if there is any kind of separation between the Head NP and the ORC, a relativizer always occurs.

*Variable 7: Rhythmic relationship between Head NP and ORCs*

It was our intuition that the presence or absence of a relativizer might be sensitive to prosodic factors. In particular, we wondered if the rhythmic structure of the utterance might not play a role in whether the speaker chose to use a relativizer or not. In order to explore this possibility, we created digital sound files for 37 randomly selected instances from our 195 ORCs, and conducted careful auditory and acoustic analysis of the rhythmic properties of these utterances (using the program PRAAT). Of the 37 instances we listened to, 14 had an overt relativizer and 23 had no relativizer. Because of the small size of this subcollection, we take our findings as preliminary, but nevertheless relevant, support for our hypothesis.

After repeated listening to these examples, an interesting pattern emerged. We found that the more rhythmically integrated the Head NP and ORC were, the more likely we were to find  $\emptyset$ -relativizer. Making use of the concepts and methods in Couper-Kuhlen 1993, we found two ways in which this integration could manifest itself: isochrony between Head NP and ORC; and very simple and regular rhythmic relationship between Head NP and ORC. Many examples display both types of relationships; a few examples, however, show one but not the other, as we'll see below.

a. Isochrony<sup>11</sup>

It is commonly believed that the principle of isochrony operates in conversational English; that is, intervals between full or accented syllables tend to be perceived to be of roughly equal length (Auer, Couper-Kuhlen and Muller 1999; Bolinger 1986; Couper-Kuhlen 1993).<sup>12</sup> Couper-Kuhlen 1993 found that isochrony is not perceived in every utterance in conversation (or across utterances), but it does appear to be perceived in many utterances. We found in our 37 ORCs that those ORCs with an isochronous relationship between the Head NP and the ORC were more likely to exhibit  $\emptyset$ -relativizer than were instances in which Head NP and ORC were non-isochronous. Isochrony was listened for perceptually, and was checked by measuring the distance between either the full vowels in the Head NP and ORC or the accented vowels<sup>13</sup> in the Head NP and ORC (as Couper-Kuhlen has found, isochrony can exist at different levels of rhythmic structure).<sup>14</sup>

Consider example (34) below.

(34) help him (1.2) look for [something] he'd lost

In example (34) we have 4 full vowels in the Head NP and ORC, and they are evenly spaced, as can be seen in the following rhythmic structure chart (adapt-

ed from Couper-Kuhlen 1993). The stars mark the presence of a full vowel (not necessarily an accented vowel), and the numbers between the stars indicate time from vowel onset to beginning of next full vowel, in milliseconds:

(35) \* 239 \* 269 \* 218 \*  
 something he'd lost

This utterance displays a highly isochronous relationship between the Head NP (*something*) and the ORC (*he'd lost*), in that each interval is roughly the same length: the distance between the /ʌ/ in *some* and the /ɪ/ in *thing* is 239 ms, the distance between the /ɪ/ in *thing* and the /i/ in *he'd* is 269 ms, and the distance between the /i/ in *he'd* and the /ɔ/ in *lost* is 218 ms. This part of the utterance is perceptually isochronous, and the acoustic measurements certainly support that hearing (the difference between 218 ms and 269 ms is not large enough to disqualify the intervals as isochronous; see Couper-Kuhlen 1993). And we can see that no relativizer is present. Compare this pattern with example (36) below, which is actually the continuation of the utterance started by (34):

(36) \* 296 \* 633 \*  
 an important paper that he'd lost

In this example we have 3 full vowels, indicated by the stars, and the durations between them are not equivalent: the distance between the /o/ in *important* and the /e/ in *paper* is 296 ms, while the distance between the /e/ in *paper* and the /ɔ/ in *lost* is 633 ms. In this case the relationship is not isochronous, and we see that a relativizer is used.

In an exploratory count of our small digitized database, we found that there is indeed a correlation between isochrony and presence/absence of a relativizer. Consider Table 10:

Table 10. Isochrony and the use of relativizers

	Isochronous	Non-isochronous	Total
-Rel	13 (68%)	6 (32%)	19 (100%)
	(81%)	(38%)	
+Rel	3 (23%)	10 (77%)	13 (100%)
	(19%)	(63%)	
Total	16 (100%)	16 (100%)	32 <sup>a</sup>

<sup>a</sup> A total of five instances were excluded. Although isochronous and non-isochronous instances are evenly distributed in this small subcollection, there is a noticeable skewing in their association with

the relativizer: 81% of the isochronous instances (13/16) showed no relativizer, while only 32% (6/16) of the non-isochronous instances showed no relativizer; 77% (10/13) of the instances of overt relativizer were non-isochronous, and 68% (13/19) of the instances of  $\emptyset$ -relativizer were isochronous. These figures are significant ( $\chi^2 = 4.66, p=0.03$ ).

It seems likely, as noted below, that the pattern noted here is affected by other variables. For example, 4 of the 10 non-isochronous instances with overt relativizer also show an intonation boundary between the Head NP and the ORC; as noted above, all such cases in our larger database exhibit an overt relativizer. And many of the instances of isochronous ORCs with  $\emptyset$ -relativizer also exhibit short Head NPs and short ORCs, both of which are associated in general with  $\emptyset$ -relativizer. This fact does not appear to be a problem; rather, it suggests that many of our factors interpenetrate one another, as might be expected with a study of language in use.

Although this is a very preliminary exploration of the relationship between isochrony and relativizer choice, an extremely striking correlation seems to have emerged.

#### b. Simple and regular rhythmic patterns

We also found that Head NPs and ORCs that together showed simple and regular rhythmic patterns tended to have  $\emptyset$ -relativizer, while Head NPs and ORCs with more complex and irregular rhythmic patterns tended to exhibit an overt relativizer. For this pattern, only the clause nucleus (verb plus core arguments) of the ORC seems to play a role (thus, for instance, in (36) and (37) below, *the other night* and *in the article* do not seem to participate in this regular rhythmic pattern). For example, Head NP+ORC combinations that showed a pattern of [/'.. '], that is, a strong (accented) syllable followed by two weak syllables and then a strong syllable, tended to have no relativizer, as can be seen in examples (37) and (38) below:

- (37) you heard about [the orgy] we had the other night {órgy we hád}  
 (38) [one thing] they said in the article that was really intriguing was {óne thing they saíd}

Compare the pattern exhibited by (37) and (38) with that in (39) — a very complex rhythmic pattern:

- (39) y'know she ho- ho- held onto [all th- (. ) jewelry boxes] that everybody made for her  
 {JÉWl ry BÓxes that ev ry bo dy MÁDE for her} / . / ... / ..

Instances in which the Head NP and the ORC show the same rhythmic pattern also tend to be produced without a relativizer. Consider example (40):



- (40) he said it's just [something new] they're discovering, y'know  
{something NÉW they disCÓvering}

In (40), the Head NP shows two weaker syllables (*some*, *thing*) preceding a strong syllable (*new*); the ORC shows the same pattern — two weaker syllables (*they're*, *dis-*) preceding a strong syllable (*co*). Instances from our corpus that show this kind of parallelism between Head NP and ORC are typically produced without a relativizer. (Note that in this case, we have both isochrony and a simple, regular relationship.)

Our findings on rhythm seem to be consistent with some of the syntactic variables listed above, especially variables 2 (Complexity of Main Clause), 4 (length of ORC verbal expression), 5 (subject of the ORC), and 6 (breaks between Head NP and ORC). Variables 2, 4 and 5 all focus on the length and complexity of the Head NP and ORC, and it is quite possible that long and complex Head NP and ORCs would be more rhythmically complex, thus tending to favor an overt relativizer. And a break of any kind, especially intonational, between the Head NP and the ORC clearly disrupts the rhythmic integration of the two clauses, which should, if our hypothesis is correct, favor an overt relativizer. Thus the syntactic factors and the prosodic factors correlate well with one another.<sup>15</sup>

While these findings are exploratory, they point to the possibility that rhythm, and perhaps prosody more generally, plays a role in what are typically thought to be syntactic processes. This possibility clearly deserves further study.<sup>16</sup>

In summary, we've considered seven variables that are involved in speakers' use of the relativizer in our ORCs. In particular, we've shown that the apparent alternation between an ORC with and without a relativizer is strongly correlated with the interaction of the variables we've just considered. As we will now suggest, all of these variables point to a larger pattern: the closer to monoclausal the [MainClause-ORC] combination is, the more that combination is treated by speakers as a single processing unit, and the more likely we are to find Ø-relativizer.

We characterize MONOCLAUSALITY in terms of both semantic and formal properties. Thus, following Lambrecht and Diessel and Tomasello (see just below), an apparently biclausal form is more monoclausal the less semantic content its "Main Clause" has. Similarly, if an apparently biclausal form can be shown to be systematically shorter, more unitary, and/or arguably less burdensome to process than a counterpart apparently biclausal form, we consider it more monoclausal.

#### 4. ORCs, frequency, and monoclausality

Now that we've outlined the variables that correlate with the use of relativizers in ORCs, we turn to the issue of the implications of these correlations. The findings of Section 3, accounting for the use of the relativizer, can be seen in terms of an overall pattern: the more monoclausal an ORC utterance is, the more likely it is to display  $\emptyset$ -relativizer. The more biclausal an ORC utterance is, the more likely it is to display an overt relativizer. Let us briefly review just how this is the case, by considering those six variables which directly affect monoclausality. Variables 1 and 2 involve the semantic emptiness of the head noun and of the verb in the "Main Clause"; we argue that the less semantic content there is in the Main Clause, the less the Main Clause is taken as syntactically separate from the Relative Clause. Variable 3 has to do with superlatives. Recall that superlatives categorically occur in our database without a relativizer; we take this as evidence that superlatives are represented for speakers as monoclausal constructions whose qualifying phrase, even if it has a predicate, is not stored and retrieved as a Relative Clause at all.<sup>17</sup> Variables 4 and 5 involve the complexity of the Relative Clause, Variable 4 having to do with the length of the ORC and Variable 5 with its subject. Clearly, the longer the Relative Clause is, the less likely it is to be taken as part of the "Main Clause", while the subject of the Relative Clause affects monoclausality in that the continuum [*I* > other pronoun > lexical noun] reflects a subject weight that is both shorter and could be hypothesized to impede processing the entire utterance as a single unit. Variables 6 and 7 deal with the prosody of the two parts of an ORC utterance; again, we hypothesize a break between the two parts to impede monoclausality and rhythmic integration to enhance it.

The obvious highly iconic nature of this result supports the views of Haiman 1983 and 1985, who argues that "grammatical closeness" reflects "conceptual closeness", that is, unitization as a single chunk. In other words, the data suggest that speakers' everyday practices involving Main Clauses and ORCs can be modeled in terms of a continuum. At one end of the continuum are clearly biclausal, presumably more compositionally constructed, MainClause-ORC formats, such as (41), in which the use of the relativizer is favored.

(41) I ran into [this other person] who we were talking about

At the other end of the continuum are simpler formats with more hallmarks of single monoclausal packages, where the distinction between Main and Relative Clause is blurred, and in which the use of the relativizer is disfavored, as illustrated in (42)–(44):

- (42) is that [the one] I bought?  
 (43) there was [something] we needed  
 (44) that's [the first compliment] I've got in a long time

In proposing that speakers' conversational practices can be modeled in terms of a continuum, we are, following Bybee (1998, 2001, 2002a, 2002b), and Bybee and Thompson (1997), making a claim about the relationship between the storage and retrieval of linguistic units and the frequency with which they occur: the more frequent a group of linguistic items is, the more it will tend to be stored and used as a single unit. In the case of Relative Clauses, the more frequent a Main-Clause-ORC combination is, the more likely it is to be stored and used as a single-clause-like unit without a relativizer, with the effect that it may have rather little in common with an instantiation of a heavier Main-Clause-ORC combination which we suggest is stored at the "bi-clausal" end of the continuum.

As support for the claim that these utterances can be insightfully viewed as monoclausal for speakers, we'd like to take a brief look at the acquisition of Relative Clauses by English-speaking children. Diessel and Tomasello 2000 show that children's earliest Relative Clauses are "presentational relatives" (with or without a relativizer) (Lambrecht 1988), as in (45)–(46):

- (45) Here's [a tiger] that's gonna scare him  
 (46) Here's [a mouse] go sleep

Following Lambrecht's lead, they show that "92% of the earliest Relative Clauses occur in constructions that express a single proposition" (138; emphasis added). They argue that "presentational relatives form a specific grammatical construction ... (with) semantic and pragmatic properties that cannot be entirely explained in terms of the properties of their components" (137). Taking a constructional perspective, they suggest that the constructional schema for these earliest Relative Clause utterances takes the form:

[[PRO-COPULA-NP] [RC]]

and is paired with a specific meaning, that of asserting new information.<sup>18</sup> Lambrecht and Diessel and Tomasello also note that the Main Clause in instantiations like (45)–(46) is minimal, hardly a Main Clause at all, and can be seen as similar to French *voici/voilà* or *ya (il y'a)* Main Clauses in low verbiness, as in these constructed examples (Lambrecht 1986):

- (47) **voilà** mon frère qui arrive  
*here's my brother who's arriving*

ya mon frère qui arrive  
*there's my brother who's arriving*

In fact, these “presentational” Main Clauses with children’s earliest Relative Clauses are a type of Copular Main Clause.<sup>19</sup>

Comparing the Diessel and Tomasello results to our own findings, we see a very similar monoclausal tendency in the utterances in our data. But, as we have noted, what is particularly striking about what the data reveal is that this tendency is strongly related to frequency: in particular, the single most frequent pattern for our ORC utterances encompasses a set of patterns at the most integrated (what we are calling the MONOCLAUSAL) end of the continuum. That is, copular Main Clauses, which represent the least complex Main Clause, account by themselves for 46% of all of our examples, and they account for 53% of all instances of ORCs with a Main Clause (recall that a small number of our ORCs occurred with no Main Clause).<sup>20,21</sup>

a. Predicate Nominal Heads (see Table 3)

(48) .. that’s [something] I’m not used to.

(49) that’s just [a bed] they move around

b. Subject Heads in Copular Main Clauses (see Table 4)

(50) [all] she did was sleep

(51) [the only thing] you can do is be the best you can.

c. Existential Heads (see Table 5)

(52) there’s not [much] they can do

(53) is there [any way] he could meet us?

As we’ve suggested, these copular Main Clauses are the least substantive Main Clauses possible; just as with the “presentative relatives”, they are almost not Main Clauses. And, as noted just above, they are also the most frequent single pattern in the data, as compared with all other Main Clause types.

Further, as we have seen, these ORC instantiations tend to have *I*, *you*, or *they* as ORC subject; these pronouns contribute to greater monoclausality of [MainClause-ORC] because they offer no lexical content to process, and because they almost universally represent “given information” (Chafe 1994). They also are the 3 most common pronouns in English conversation.<sup>22</sup> The most frequent Copular Main Clause ORC patterns tend to be simpler overall as well, with Ø-relativizer, shorter Main Clauses, and shorter ORC verbal

expressions. They are thus as close to monoclausal as a [MainClause-ORC] combination can get.

Interestingly, in these monoclausal ORC utterances, the Main Clauses also don't have much semantic substance, so they stand by themselves rather poorly, and often 'need' something with them to provide semantic substance. The ORC provides this 'something', as shown in these examples:

- (54) is there *any way* ...?  
 (cf. is there [any way] he could meet us?  
*all* ... was sleep  
 (cf. [all] she wants to do is sleep)  
 she thought that would be *something* ...  
 (cf. she thought that would be [something] I'd like)

This is especially true for Unique Heads; in utterances like (55), the ORC is almost 'obligatory'.<sup>23</sup>

- (55) that's *the only place*...  
 (cf. that's [the only place] you can go at night)

Our most frequent subtype of ORC is thus one in which the Main Clause is hardly clausal at all; these instances are as close to monoclausal as they can be. And, importantly for our claim, they strongly tend to exhibit  $\emptyset$ -relativizer (73%).

In addition to being monoclausal, they also tend to contain lexical items that are extremely frequent in English conversation: the Main Clause verb is *be*, which is the most frequent verb in English (and indeed with its inflected forms is the second most frequent word in the language); the head NP tends to be (empty) items like *thing*, *something*, *way*, *time*, *one*, which are highly frequent in conversation; the subject of the ORC is typically one of the 3 most common pronouns (*I*, *you* and *they*),<sup>24</sup> which are themselves the most common subjects in English conversation; and even the verb of the ORC tends to be one of the very frequent ones (*have*, *do*, *make*, *know*, *see*).<sup>25</sup>

Almost everything about these instances, then, suggests that they exhibit highly frequent collocations and could very well represent sets of pre-stored patterns. Consider example (56):

- (56) That's [the way] it is

This utterance, along with others like it — e.g. *that's the way he is*, *that's the way it goes*, *that's the way the ball bounces*, *that's the way the cookie crumbles* — are stock phrases which seem to be stored as entire units. Intriguingly, Tottie (1997: 5) makes the same point for the British National Corpus:

Many of the most frequent constructions border on grammaticalization (cf. *the way Ø you look tonight, the time Ø he left, the reason Ø he did it*)

Tottie and Lehmann's 1999 findings also show that *way* and *time* account for fully 82% of all the Relative Clauses with Ø-relativizer in their BNC data, and that these two head nouns account for most of the "zero constructions" in their sample.

A crucial finding reveals itself in these examples: these stock phrases — all without relativizer — are at the monoclausal end of the continuum. This fact suggests that the properties of being pre-stored and being highly integrated are closely related in the case of ORCs. It may be frequency which unites these phenomena; that is, it could be that utterance-types which are highly frequent tend to become reduced and also tend to become stored as a unit (see, for example, Bybee 2002a and 2002b). As we suggest in Section 6, however, it is possible that other factors may be at work in these cases. Consider, for example, the apparently stock format *that's the way [ORC]*: although there are five instances of *that's the way [ORC]*, the form of the ORC is different in each case; and it is not clear to us if 5 cases out of 195 instances of ORCs (and out of many many hours of conversation) ought to count as "very frequent" or not. We have thus been led to wonder if factors other than, or in addition to, frequency might play a role in the pre-storing and mergedness of this clause-type. We explore this intriguing question further in Section 6.

Other utterance-types on this end of the continuum might not be stored as entire units but still seem to be made of highly recurrent collocations: for example, *that's the way I think, you're the best son I have, next time I write*. In this regard it is interesting to note that all 4 examples we found of Head NP *all* or *everything* have the verb *do* in the ORC:

[all] *she did* was sleep  
 [everything] *we do* is like that  
 [all] *I could do* is say...  
 pretty much [all] *we're gonna do* is go out

On this extreme end of the monoclausal portion of continuum, then, there seems to be a high level of lexical specificity and pre-storage (and notice that none of these examples show an overt relativizer).

Still other utterance-types at the monoclausal end of the continuum show somewhat less lexical specificity and typical collocations, but nonetheless have a familiar higher-level pattern to them and some lexical specificity. Consider example (57):

(57) that was [the ugliest set of shoes] I ever saw in my life

Although the Head NP is quite complex (and Unique in our database), this utterance clearly instantiates a recurrent Unique template of the form: *the unique NP I've ever seen/ever saw in my MOD life*. In fact, we have another example that fits this pattern: *that was [the most weird day] I've ever seen in my entire adult life*.<sup>26</sup> Such examples are almost certainly not pre-stored as entire utterances, but they may represent instantiations of common formats that exhibit some lexical specificity and some generality.

Thus the more monoclausal end of the continuum is itself made up of varying degrees of integration, ranging from memorized stock phrases to somewhat higher-level formats with some degree of lexical specificity and some degree of lexical freedom; notably they all tend to lack an overt relativizer.<sup>27</sup>

On the other end of the continuum, we would like to suggest, are highly biclausal ORC utterances, illustrated by (58):

(58) she ho- ho- held onto all [those jewelry boxes] that everybody made for her when we were kids

Example (58) displays a main-clause verb that is unique in our database, with a complex full noun phrase Head NP (of which it is the only example), and the ORC itself shows another level of embedding with the *when*-clause. We thus consider such an example to illustrate the highly biclausal end of the continuum, composed of items that are not common collocations, and thus are almost certainly less likely to be pre-stored (Bybee 1998; Bybee and Hopper 2001a, 2001b). And, as we have suggested, it is no coincidence that such examples tend to contain an overt relativizer: the relativizer serves to mark the boundary between the two quite distinct clauses, and we might surmise, as Bolinger 1972 hinted, to aid in distinguishing the parts of the utterance.

As might be expected, even this end of the continuum shows variability, however. Some of the utterances on the biclausal end are extremely complex and obviously composed on the fly, such as (58) above. Other instances may reveal a bit more pre-formatting. Consider (59):

(59) I ran into [this other person] who we were talking about

The Main Clause verb in (59) is not common, but we do have one other instance of it in our collection; the Head NP is complex, and we have no other instances of it in our collection, but the noun *person* is frequent enough to show up on frequency lists; the ORC shows a less common pronoun (only 19 instances in our collection have *we* as subject of the ORC, and in frequency lists

*we* lags behind *it, I, you, he, and they*), but nonetheless it is a pronoun and not a full NP and is thus vastly more common than any full NP; finally, the ORC verb, while not highly frequent, still shows 6 instances in our collection. It is possible that this kind of biclausal utterance could have some pre-stored or formatted elements to it. Thus both ends of the continuum exhibit variability in their organization.

In this section we have argued for the view that the instances of ORC utterances in our database form themselves along of continuum of monoclausality to biclausality, the more monoclausal end having an almost empty Main Clause, simple head NP and short ORC, with highly integrated rhythm, and displaying a range from extreme lexical specificity (memorized chunk) to a somewhat more general pattern with still a fair amount of lexical specificity; and we have also argued that the monoclausal pattern correlates with a lack of relativizer in a great percentage of our cases. The biclausal pattern tends to be made up of less frequent lexical items, and is less likely to be composed of pre-stored chunks or low-level formats with lexically specific components. This end of the continuum shows a greater preference for an overt relativizer.

These findings regarding clausality, frequency and presence/absence of relativizer raise questions for the view that all Relative Clauses are members of a single grammatical construction. In the next section we explore the implications of our findings for the notion of “construction”.

## 5. ORCs and “constructions”

Traditionally, noun phrases containing Relative Clauses have been assumed to be “constructions” with their Main Clauses. The term “construction” is used by most functionalists in talking about Relative Clause patterns, as can be seen in such typological works as Comrie 1989, Croft 1990, Keenan 1985, Payne 1997, and Song 2001.

Our findings suggest, however, that the notion of a single “construction” for Relative Clauses is problematic. First, considering the subconstruction that we have labeled ORC, as we have shown, the apparent alternation between the use vs. non-use of a relativizer is highly correlated with lexical, grammatical, semantic, prosodic and pragmatic factors (as discussed in Fox 1994), and, as we noted in Section 4 above, the practices underlying the fluent conversational use of utterances with ORCs seem to be organized in the form of a networked family of patterns (cf. Bybee 1998, 2002a, 2002b), which can be described in terms of a continuum, from highly biclausal to close-to-monoclausal. In other



words, our data suggest that the practices involved in producing what have been called Relative Clauses involve a wide range of formats, with more, or less, biclausality: some of those at the monoclausal end of the continuum exhibit a great deal of lexical specificity, while others of them exhibit less lexical specificity and greater low-level generality. We can model speakers' fluent use of the full, compositional Relative Clause, as we know it from traditional linguistic studies, by placing it at the biclausal end of the continuum, although there may be some degree of common collocations occurring even there. There is thus a great deal of scalable diversity in patterning among what have been traditionally lumped together under one "construction".

Second, some members of this "family" — most notably the Unique Head format and the *all/everything* format — seem to have basically grammaticized without the relativizer. This fact suggests that there is a great deal of small, local patterning, with varying degrees of lexical specificity, within the large domain we think of as Relative Clauses, and it would be difficult to capture this small domain of patterning if we think of these formats as a single unified "construction".

Paradoxically, we started out with a criterial attribute for inclusion as an ORC in the database, namely "optionality" of the relativizer, but in so doing we have found that in actual usage, the members at the extreme monoclausal end of our continuum in fact do not take a relativizer. Thus we have a cluster of utterance-types presumably related historically which now share no single attribute, but which seem to exhibit local, small-domain patterning. This is precisely what Wittgenstein meant by the notion of "family resemblance".<sup>28</sup>

We propose, then, that it is more illuminating to describe the patterning in the use of ORCs that we have found in terms of a family of dynamic formattings, each of which experiences its own discourse pressures and its own grammaticization tendencies, such that certain members of the family may not share any features with other members of the family. For example, the utterance we considered in (56), repeated here:

(56) That's [the way] it is

shares very few features with the utterance we looked at in (58), repeated here:

(58) she ho- ho- held onto all [those jewelry boxes] that everybody made for her when we were kids

(56) is almost certainly stored as a unit and has very few of the properties typically associated with Relative Clauses: the Relative Clause does not help to pick

out the referent of the Head NP, and although one could imagine considering *the way* to be gapped from the Relative Clause, the English utterance *it is the way* bears no semantic or pragmatic relationship to the utterance in (56). Considering this example to exhibit “gapping” is thus perhaps stretching our understanding of that concept somewhat. (58), on the other hand, is clearly not stored as an entire unit, and is a classic instance of a Relative Clause Construction: a full and complex Main Clause with a Unique and complex Head NP, a Relative Clause with a lexical subject, and the referent of the Head NP obviously has a grammatical role in the Relative Clause. Without analytically abandoning the notion of Relative Clause Construction, we note that our analysis embraces this continuum, and is compatible with a cognitive representation appealing to exemplars and networks suggested by many current researchers (Bybee 1998, 2002a, 2002b; Pierrehumbert 2001 and *inter alia*).<sup>29</sup> As we suggested in Section 4 above, utterances like *that’s the way it is*, or *that’s the way it goes*, are entirely formulaic, stock phrases, often offered in conversation as a bid to close a sequence (see Schegloff 1995). On the other hand, utterances like (57) above, repeated here:

(57) that was [the ugliest set of shoes] I ever saw in my life

exhibit some lexical specificity but also instantiate low-level formats which allow for some lexical variability. And utterances like:

(60) where were [those ones] Jill was talking about?

are yet another step removed, exhibiting fewer stock phrases, much less lexical specificity and possibly instantiating no low-level formats.

This great diversity, even among apparently similar utterances, is something we believe needs to be foregrounded and highlighted as we develop our syntactic theories. Our current syntactic models tend to gloss over such diversity in the name of capturing generalizations; we believe it is time to reverse the trend in syntactic studies, to focus on everyday spoken language use which reveals lexical specificity and low-level formatting, in addition to more general formatting, and to recognize the interesting and subtle differences among items previously treated as “the same.”

## 6. Conclusions

In this paper, we have explored the organization of Relative Clauses in particular, and grammar more generally, through the lens of the presence/absence

of the relativizer. This exploration suggests the following shifts in our understanding.

First, as most discourse and variationist linguists have suspected, the choice between relativizer and  $\emptyset$ -relativizer for ORCs is strongly correlated with the pragmatic-prosodic properties of the utterance; the relativizer is thus not “optional”.

Second, the knowledge that enables speakers to do what they do in their use of utterances with ORCs can be seen in terms of a continuum. At one end are highly biclausal ORC utterances, which can be understood as compositional. At the other end are nearly monoclausal ORC utterances, which are characterized by:

- $\emptyset$ -relativizer
- maximally simple and relatively short “Main” and “Relative” clauses
- an integrated syntax, semantics and prosody, with blurred distinction between Main and Relative clauses
- high frequency

Third, and most importantly, we have suggested that the concept of grammatical organization underlying the notion “Relative Clause Construction” might be usefully replaced by a view of grammatical organization that has small-domain, even sometimes lexically specific, formats which exist in a dynamic, family-resemblance relationship to one another that can be modeled in terms of a continuum. This view of grammatical organization is in alignment with recent studies of other constructions in conversation (see for example Akatsuka and Clancy 1993; Bybee 1998, 2002a, 2002b; Couper-Kuhlen 1999; Duranti and Ochs 1979; Fox 1994; Geluykens 1992; Hopper 1998, 2000; Manes and Wolfson 1981; Nicita 2002; Scheibman 2000; Thompson 2002; Thompson and Mulac 1991a, 1991b; Thompson and Hopper 2001), and appears to represent a promising new direction for usage-based research in syntax.

As noted above, our study, however, also raises an exciting question with regard to this growing literature on frequency, storage and grammatical patterning. While our data confirm earlier studies in finding that highly frequent patterns tend to be reduced, stored and retrieved as fragments or chunks, and to show high lexical specificity, our data also show that frequency is not the only factor involved. We are referring here to the finding, mentioned earlier, that some of the formats that have grammaticized without the relativizer might be considered less than highly frequent: for example, we have only 4 instances of the *all/everything* format, and yet 100% of these instances show  $\emptyset$ -relativizer. If these figures do represent relatively high frequency, as has been suggested

to us (Bybee, p.c.), then frequency remains a primary explanatory factor. If, however, these patterns might be considered to be less than highly frequent, then factors other than, or perhaps in addition to, frequency ought to be considered.<sup>30</sup> We have explored the possibility here that it is integration, or degree of “mergedness”, of the two clauses that induces the  $\emptyset$ -relativizer, and not frequency alone (although frequency clearly plays a role in what becomes merged and monoclausal). Clearly, however, the answer to this question regarding the explanatory universe of frequency, and to the smaller question of “What counts as a frequent pattern?” remain for future research.

In addition, the current study expands our understanding of grammatical organization to include a range of syntactic practices. In much of the earlier work in this area, scholars have found that grammatical formats tend in conversation to become grammaticized as something like discourse markers. Thompson and Mulac 1991a and 1991b and Thompson 2002 argue that *I think* and *I guess* have become grammaticized as epistemic/evidential/evaluative markers in English,<sup>31</sup> and Kärkkäinen 2004 has shown a range of stance-marking functions for *I think*; in a similar vein, Hopper 2000 suggests that pseudo-clefts in English conversation tend to display one of 3 highly frequent verbs and basically function as discourse markers. In the current study, however, we have found something more complex than this. Our ORC patterns reveal not only what appear to be pre-stored chunks; they also display a large range of utterance types that vary (1) in their lexical specificity — from highly lexically specific to some degree of lexical specificity and some low-level formatting, to no lexical specificity at all; (2) in the degree of integration between “Main” and “Relative” clauses — from highly integrated to much less integrated; (3) and in their complexity — from simple structures to quite complex syntactic structures.

What is important about our findings is that they suggest a wide range of patternings, not just one highly frequent, probably pre-stored chunk. Our findings suggest that speakers make use of a wide range of practices — some entirely pre-stored, others partially pre-stored and partially composed based on low-level formats, others not at all pre-stored — and this diversity must be acknowledged and described in our syntactic theories. Focusing too heavily on one practice — whether it is the pre-stored, monoclausal end or the entirely compositional end misses the diversity that underlies the practices by which speakers use ORCs in conversation.

With these findings, we hope to have contributed to a deeper understanding of the nature of grammar, that is, of the grammatical practices in which speakers engage in their everyday talk.

## Notes

\* We are grateful to the following people, who have provided valuable input to our thinking on constructions and Relative Clauses in conversation: Mira Ariel, Joan Bybee, Elizabeth Couper-Kuhlen, William Croft, Susanna Cumming, Carol Genetti, Adele Goldberg, Marja-Liisa Helasvuo, T. Florian Jaeger, Dan Jurafsky, Tsuyoshi Ono, Stephanie Schulze-Wenck, Hongyin Tao, Michael Tomasello, Gunnel Tottie, Thomas Wasow, and one anonymous reviewer. We are also indebted to Kristine Hildebrandt and Joanne Schroeder for their database help, and to Terran Brown and Rebecca Zwick for their expert statistical help. For any errors that may remain, we alone are responsible.

1. By ‘anaphoricity’, Temperley means “the use of RP/comp [i.e., relative pronoun or complementizer — BAF/SAT] should be determined by whether the subject of the Relative Clause is anaphoric with items presented earlier in the discourse” (2003: 473).

2. These were distributed evenly across the speakers and the conversations in the corpus, with no apparent skewing to any particular speaker or conversation.

3. For example, in the following utterance, *where* appears to be required (the speaker is relating the plot of the movie *Austin Powers*):

And then there’s [this point] where he’s got like this spacesuit on?

4. where the ‘Main Clause’ is in fact a synchronic preposition, and only diachronically a clause.

5. No instances of the relativizer *whom* occurred in our collection.

6. A similar pattern is found by Tottie 1995 and 1997 and Tottie and Lehmann 1999; they show that in their corpora of written British English, Relative Clauses with empty heads and those playing an adverbial role in the Main Clause, including those with head nouns such as *day*, *time* and *way*, show much higher rates of  $\emptyset$ -relativizer than the average.

7. We are aware that some researchers insist on independence of data points for chi-square results to be meaningful. Although our data points are not independent in a strict sense, since many Relative Clauses come from the same speaker or the same conversation, we take our chi-square results to be valuable indicators of significance, since our considerable familiarity with the interactional contexts in which these utterances were embedded provides no basis for believing that occurrences of Relative Clauses influence each other in any way that would be likely to invalidate a chi-square test.

8. There are obviously many ways of measuring complexity; our argument here is that with two plausible measures, complexity can be shown to play a role in the use of a relativizer.

9. Examples (13)–(15) illustrate the fact that the variables we are considering are not independent of each other: these examples show, e.g., that Heads that are subjects of copular clauses tend to be empty. We will continue to see this overlap as we examine the other variables.

10. Biber et al. 1999 give their percentages this way rather than as an average to indicate the percentage range.
11. Many thanks to Elizabeth Couper-Kuhlen for helping us think about isochrony.
12. Couper-Kuhlen 1993 makes it clear that isochrony is a perceptual phenomenon and does not correlate exactly with actual duration.
13. We actually found a few cases of isochrony at the level of the syllable, that is, where the intervals between syllables –both full and reduced – were perceptually the same. (We are grateful to Elizabeth Couper-Kuhlen for her help in identifying these examples.)
14. As Couper-Kuhlen 1993 suggests, it takes 3 beats for a pattern of isochrony to be recognizable: the interval between the first two beats sets up the pattern; the interval between the second and third beats is needed to confirm if the pattern is being continued. We thus excluded cases from our counts in which there were not at least 3 beats encompassing the Head NP and the ORC. We did include cases like the following, though, in which the pattern was established between something earlier in the Main Clause and the Head NP:

\*                    \*    \*

that was the one I wanted

The interval between the Main Clause subject, *that*, and the Head NP, *one*, sets up the rhythmic pattern (it is 190 ms). The pattern is continued in the next interval between *one* and *wanted* (which is 261 ms). This RC thus exhibits isochrony, according to our criteria (the difference between 190 ms and 261 ms is still within the range of what is hearable as isochronous; see Couper-Kuhlen 1993). Utterances that started with no full or accented syllables before the Head NP were generally excluded. Consider the example below:

\*    \*                \*    \*

there was a tape Taryn was talking about

In this utterance, there are no full or accented syllables in *There was a* and therefore no rhythmic pattern has been established before the interval between *tape* and *Taryn*. Even though there is another foot later in the Relative Clause, we excluded cases like this from our counts, on the grounds that the crucial foot — the one that could have contained a relativizer — cannot be evaluated for isochrony at the moment of its production (although it could establish a rhythm which becomes isochronous).

15. It is not possible at this stage of our understanding to know if the syntactic factors are primary and the prosodic factors are derivative, or if the prosodic factors are primary and the syntactic factors derivative, or if both are primary.
16. As John Local has said (p.c): “There is no grammar without phonetics.” The current study provides evidence in support of this claim.
17. Note in this connection that the ‘qualifying phrase’ indeed need not be a clause, as in this example from our materials: *these are the shittiest speakers on earth*.

18. An analogous constructional claim is made for children's acquisition of complement constructions in Diessel and Tomasello 2001.

19. We note that these 'presentational relatives' differ from those we are considering in this study in that they are Subject relatives, while ours are Object, Oblique and Adverbial relatives. There are no occurrences of such presentational relatives in our adult database, but a larger database might well reveal some instantiations. In this case, if our analysis were to be applied to Subject relatives, these would readily be included at the more 'monoclausal' end of our continuum.

20. Using a comparable database, Thompson and Hopper 2001 showed that about 26% of clauses in English conversation are copular clauses; the fact that 46% of the ORCs have copular Main Clauses is thus worthy of note.

21. The next most frequent Main Clause verb is *have*; with only 20 instances, it accounts for just 10% of the data.

22. Our English frequency figures come from (<http://www.eecs.umich.edu/~qstout/586/bncfreq.html>).

23. We are grateful to Adele Goldberg for this observation.

24. Interestingly, in our data the pronoun *he* (together with *his* and *him*), which in some frequency lists occurs as the second most common pronoun in English (after *I*, *me* and *my*), occurs quite infrequently: it is the subject of the ORC only 13 times. In 8 (62%) of those cases, no relativizer occurs.

25. All of these verbs except *make* are in the top 50 most commonly used words in English.

26. This format represents one type of what scholars in Conversation Analysis call "extreme case formulations" (Pomerantz 1986).

27. It could be argued that if ORCs utterances with fairly empty Main Clauses favor  $\emptyset$ -relativizer, then ORCs with no Main Clauses at all should categorically show  $\emptyset$ -relativizer. Our data suggest that this is not the case, however. The rate of relativizer use in ORCs with no Main Clause is almost identical to the rate for the collection as a whole (60%). We take this as support for our claim: what the data show is that it is not 'monoclausality' per se that correlates with lower use of relativizer; rather, what seems to be the key factor is the INTEGRATION of erstwhile 'Main' and 'Relative' clause.

28. We appreciate Hongyin Tao's sharing of an example from his corpus of an utterance at the extreme monoclausal end of our continuum:

- (i) next thing you know Dobbin will sue himself

This example is reminiscent of [ay 5 no], the extreme fused end of the *I don't know* continuum discussed by Bybee and Scheibman 1999 and Scheibman 2000. Bybee and Scheibman also used a criterial attribute, namely the occurrence of *don't*, to collect their data, but found that the examples at the extreme end of the fusion continuum actually contain no perceptible occurrences of *don't*.

29. In fact, some readers of this paper have suggested that utterances like *that's the way it is* are not Relative Clauses at all. This comment supports exactly the point we are trying to make: even trained linguists disagree as to what belongs within a single construction. Recall that our criterion for inclusion in our database was the possibility of an utterance being used with or without a relativizer. According to this criterion, all the examples in our database can be seen as forming a 'natural class'. In fact, if we excluded utterances like *that's the way it is* as instances of 'Relative Clause', we might have missed some of the richness of the phenomenon, precisely because what linguists tend to think of as Relative Clauses — from our history of basing analyses on invented data — are precisely the highly biclausal cases. To come closer to an understanding of how the grammar of Relative Clauses must be organized in the heads of speakers, we have thus felt it crucial to include as wide a range of actually occurring cases as possible.
30. Nicita 2002 uncovers a similar pattern in Spanish cognitive verbs: one of the least frequent verbs, *imaginarse* displays the highest degree of grammaticization of any of the verbs examined. For further discussion of this issue, see also Tomasello and Stahl 2004. For an account based on predictability, see Jaeger and Wasow (to appear).
31. For similar findings for Spanish, see Nicita 2002.

## References

- Ariel, Mira. 1999. Cognitive universals and linguistic conventions: The case of resumptive pronouns. *Studies in Language* 23(2): 217-269.
- Akatsuka, Noriko and Patricia M. Clancy. 1993. Conditionality and deontic modality in Japanese and Korean: Evidence from the emergence of conditionals. *Japanese/Korean Linguistics* 2: 177-192.
- Auer, Peter; Elizabeth Couper-Kuhlen and Frank Muller. 1999. *Language in time*. Oxford: OUP.
- Berk, Lynn M. 1999. *English syntax: From word to discourse*. New York NY: OUP.
- Biber, Douglas; Stig Johansson; Geoffrey Leech; Edward Finegan; Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Bolinger, Dwight. 1972. 'That's 'that'. The Hague: Mouton.
- Bolinger, Dwight. 1986. *Intonation and its parts*. Stanford CA: Stanford University Press.
- Bybee, Joan. 1998. The emergent lexicon. *Chicago Linguistic Society* 34: 421-435.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: CUP.
- Bybee, Joan. 2002a. Mechanisms of change in grammaticization: The role of repetition. *Handbook of historical linguistics*, Richard Janda and Brian Joseph (eds.), 602-623. Oxford: Blackwell.
- Bybee, Joan. 2002b. Sequentiality as the basis of constituent structure. The rise of language out of pre-language, Bertram Malle and T. Givon (eds.), 109-132. Amsterdam: John Benjamins.
- Bybee, Joan and Paul Hopper (eds.). 2001a. Frequency and the emergence of linguistic structure. Amsterdam: John Benjamins.



- Bybee, Joan and Paul Hopper (eds.). 2001b. Introduction. Frequency and the emergence of linguistic structure, Joan Bybee and Paul Hopper (eds.), 1-24. Amsterdam: John Benjamins.
- Bybee, Joan and Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics* 37(3): 575-596.
- Bybee, Joan and Sandra A. Thompson. 1997. Three frequency effects in syntax. *Berkeley Linguistics Society*, 378-388. Berkeley CA: Berkeley Linguistics Society.
- Chafe, Wallace. 1994. Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing. Chicago IL: University of Chicago Press.
- Comrie, Bernard. 1989. *Language universals and linguistic typology* (2nd ed.). Chicago IL: University of Chicago Press.
- Couper-Kuhlen, Elizabeth. 1993. *English speech rhythm*. Amsterdam: John Benjamins.
- Couper-Kuhlen, Elizabeth. 1999. Varieties of conditionals and their emergence in discourse. *Arbeitspapier 99: Issues in interdisciplinary research on the lexicon*, Aditi Lahiri; Alexander Pasterovsky and Christophe Strauss (eds.), 89-130. Konstanz: University of Konstanz.
- Croft, William. 1990. *Typology and universals*. Cambridge: CUP.
- Diessel, Holger and Michael Tomasello. 2000. The development of relative clauses in spontaneous child speech. *Cognitive linguistics* 11(1/2): 131-151.
- Diessel, Holger and Michael Tomasello. 2001. The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics* 12(2): 97-141.
- Du Bois, John; Stephan Schuetze-Coburn; Danae Paolino and Susanna Cumming. 1993. Outline of discourse transcription. *Talking data: Transcription and coding methods for language research*, Jane A. Edwards and Martin D. Lampert (eds.), 45-89. Hillsdale NJ: Lawrence Erlbaum.
- Duranti, Alessandro and Elinor Ochs. 1979. Left-dislocation in Italian conversation. *Discourse and syntax*, Talmy Givón (ed.), 377-416. New York NY: Academic Press.
- Fox, Barbara A. 1994. Contextualization, indexicality, and the distributed nature of grammar. *Language Sciences* 16(1):1-37.
- Geluykens, Ronald. 1992. From discourse process to grammatical construction: On left dislocation in English. Amsterdam: John Benjamins.
- Givón, Talmy. 1993. *English grammar*. Amsterdam: John Benjamins.
- Guy, Gregory R. and Robert Bayley. 1995. On the choice of relative pronouns in English. *American Speech* 70(2): 148-162.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59: 781-819.
- Haiman, John. 1985. *Natural syntax*. Cambridge: CUP.
- Hopper, Paul, J. 1998. Emergent grammar. *The new psychology of language: Cognitive and functional approaches to language structure*, Michael Tomasello (ed.), Mahwah NJ: Lawrence Erlbaum.
- Hopper, Paul. 2000. Grammatical constructions and their discourse origins: Prototype or family resemblance? *Applied cognitive linguistics: Theory, acquisition and language pedagogy*, Martin Pütz and Susanne Niemeier (eds.), Berlin: Mouton de Gruyter.
- Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56(2): 251-299.

- Jaeger, T. Florian and Thomas Wasow. To appear. Processing as a source of accessibility effects on variation. *Berkeley Linguistics Society* 41.
- Jespersen, Otto. 1933. *Essentials of English grammar*. London: Allen & Unwin.
- Kärkkäinen, Elise. 2004. Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think. Amsterdam: John Benjamins.
- Keenan, Edward L. 1985. Relative clauses. *Language typology and syntactic description*, Timothy Shopen (ed.), 141-170. Cambridge: CUP.
- Lambrecht, Knud. 1986. Pragmatically motivated syntax: Presentational cleft constructions in spoken French. *Chicago Linguistic Society Parasession 22*: 115-126.
- Lambrecht, Knud. 1988. There was a farmer had a dog: Syntactic amalgams revisited. *Berkeley Linguistics Society* 14: 319-339.
- Lehmann, Hans Martin. 2002. Zero subject relative constructions in American and British English. *New frontiers of corpus research*, Pam Peters (ed.), 163-177. Amsterdam: Rodopi.
- Manes, Joan and Nessa Wolfson. 1981. The compliment formula. *Conversational Routines*. Florian Coulmas (ed.), 115-132. The Hague: Mouton.
- Nicita, Linda. 2002. *Cognitive verbs in Spanish*. PhD dissertation, University of Colorado.
- Payne, Thomas. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: CUP.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, Joan L. Bybee and Paul J. Hopper (eds.), 137-144. Amsterdam: John Benjamins.
- Pomerantz, Anita M. 1986. Extreme case formulations: A way of legitimizing claims. *Human Studies* 9: 219-30.
- Quirk, Randolph. 1957. Relative clauses in educated spoken English. *English Studies* 38: 97-109.
- Quirk, Randolph S; Geoffrey Leech; Sidney Greenbaum and Jan Svartvik. 1985. *A grammar of contemporary English*. London: Longman.
- Schegloff, Emanuel. 1995. *Sequence organization*. UCLA ms.
- Scheibman, Joanne. 2000. I dunno: A usage-based account of the phonological reduction of don't in American English conversation. *Journal of Pragmatics* 32: 105-124.
- Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax*. London: Longman.
- Temperley, David. 2003. Ambiguity avoidance in English relative clauses. *Language* 79(3): 464-484.
- Thompson, Sandra A. 2002. 'Object complements' and conversation: towards a realistic account. *Studies in Language* 26(1): 125-164.
- Thompson, Sandra A. and Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. *Frequency and the emergence of linguistic structure*, Joan L. Bybee and Paul J. Hopper (eds.), 27-60. Amsterdam: John Benjamins.
- Thompson, Sandra A. and Anthony Mulac. 1991a. A Quantitative perspective on the grammaticization of epistemic parentheticals in English. *Grammaticalization II*, Elizabeth Traugott and Bernd Heine (eds.), 313-339. Amsterdam: John Benjamins.
- Thompson, Sandra A. and Anthony Mulac. 1991b. The discourse conditions for the use of complementizer that in conversational English. *Journal of Pragmatics* 15: 237- 51.

- Tomasello, Michael and Daniel Stahl. 2004. Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* 31(1): 101-121.
- Tottie, Gunnel. 1995. The man Ø I love: An analysis of factors favouring zero relatives in written British and American English. *Studies in Anglistics*, Gunnel Melchers and Beatrice Warren (eds.), 201-215. Stockholm: Almqvist and Wiksell.
- Tottie, Gunnel. 1997. Relatively speaking: relativizer marker usage in the British National Corpus. *To explain the present: studies in the changing English language in honour of Matti Rissanen, Terttu Nevalainen and Leena Kahlas-Tarkka* (eds.), 465-481. Helsinki: Société Néophilologique.
- Tottie, Gunnel and Hans Martin Lehmann. 1999. The realization of relative adverbials in present-day British English. *Out of corpora: Studies in honour of Stig Johansson, Hilde Hasselgård and Signe Oksefjell* (eds.), 137-157. Amsterdam: Rodopi.
- Tottie, Gunnel and Hans Martin Lehmann. 2002. As and other relativizers after same in present-day standard English'. *From the COLT's mouth, and other places: Studies in honour of Anna-Brita Stenström, Leiv Egil Breivik and Angela Hasselgren* (eds.), 231-245. Amsterdam: Rodopi.

*Authors' addresses:*

Barbara A. Fox  
University of Colorado, Boulder  
bfox@spot.colorado.edu

Sandra A. Thompson  
Department of Linguistics  
University of California, Santa Barbara  
3607 South Hall  
Santa Barbara, CA 93106-3100  
sathomps@linguistics.ucsb.edu