Corpus studies of multiple center-embedding are rare despite its importance within syntactic theory and psycholinguistics. In the most comprehensive corpus study, Karlsson (2007) analyzed 132 doubly center-embedded clauses from seven European languages, postulating several (quasi-)grammatical constraints specifically tailored to the needs of multiple center-embedded clauses.

Based on the deWaC corpus (Baroni, Bernardini, Ferraresi & Zanchetta, 2009), this poster presents a corpus study of multiply center-embedded relative clauses (RCs) in German which goes beyond Karlsson (2007) not only by looking at a larger number of examples but also by taking into account close variants involving extraposition. This makes it possible for the first time to determine empirically whether multiply center-embedded RCs have unique properties requiring specific grammatical constraints.

Sentence (1) is an original corpus example with a doubly center-embedded RC. In (2), RC-high is center-embedded but RC-low has been extraposed. In (3), RC-high has been extraposed but RC-low is center-embedded. In (4) finally, RC-high and RC-low have both been extraposed. A search of the deWaC corpus (1,278,177,539 tokens of text) revealed 343 instances of doubly center-embedded RCs as in (1). In accordance with Karlsson (2007), sentences with more deeply embedded RCs were practically absent.

(1) **RC-high center-embedded, RC-low center-embedded**

Probleme gab es, weil einige Kurse, die in der schönen Broschüre, die man vorher zugeschickt bekammt, aufgelistet waren, die in der schönen Broschüre, die man vorher zugeschickt bekammt, gestrichen worden waren.

‘There were problems because some courses which were listed in the nice brochure which was sent out in advance were canceled.’

(2) **RC-high center-embedded, RC-low extraposed**

Probleme gab es, weil einige Kurse, die in der schönen Broschüre aufgelistet waren, die man vorher zugeschickt bekammt, gestrichen worden waren.

(3) **RC-high extraposed, RC-low center-embedded**

Probleme gab es, weil einige Kurse gestrichen worden waren.

(4) **RC-high extraposed, RC-low extraposed**

Probleme gab es, weil einige Kurse gestrichen worden waren.

Thus, doubly center-embedded RCs do occur, but they are rare. This raises two major questions:

(i) Why are doubly center-embedded RCs not avoided completely by means of extraposition?

A logistic-regression model confirms the null-hypothesis that the same factors are at work as
in extraposition of simple RCs (e.g., Hawkins, 1994): rate of extraposition ((1) versus (3)) can be predicted with high accuracy (Somers’ C about 0.93) from the amount of material crossed by extraposition.

(ii) Why do doubly center-embedded RCs occur so rarely? The null-hypothesis is that the lengthy dependency between the antecedent NP and the clause-final verb makes complex intraposed relative clauses difficult to process (e.g., Gibson, 2000), and not center-embedding as such. The null-hypothesis is confirmed again. In intraposed position, sentences with degree of center-embedding = 1 occurred about five times as often as sentences with degree of center-embedding = 2 ((2) versus (1)). Importantly, a similar ratio turned up in extraposed position (sentences with degree of center-embedding = 0 (cf. (4)) versus sentences with degree of center-embedding = 1 (cf. (3)).

The present results argue that constraints on multiple center-embedding follow completely from processing limitations; grammatical constraints are superfluous. This conclusion is confirmed by two further pieces of evidence. First, 15% of all corpus instances of type (4) had the VP of RC-high missing. Thus, memory limitations causing the missing-VP effect (e.g., Gibson & Thomas, 1999) also apply to language production. Second, a grammaticality-judgments experiment investigating the structures in (1)–(4) and sentences with the VP of RC-high missing showed that such structure are perceived as basically grammatical and that the missing VP effect also occurs in language comprehension (see Table 1).

<table>
<thead>
<tr>
<th></th>
<th>RC-low center embedded</th>
<th>RC-low extraposed</th>
<th>Missing VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC-high extraposed</td>
<td>85 (2.7)</td>
<td>83 (4.0)</td>
<td>13 (3.9)</td>
</tr>
<tr>
<td>RC-high center-embedded</td>
<td>77 (4.4)</td>
<td>73 (4.1)</td>
<td>58 (4.8)</td>
</tr>
</tbody>
</table>

Table 1: Mean percentages of grammatical judgments. Standard error (by participants) is given in parentheses.

References: