Condition C in German A′-movement: Tackling challenges in experimental research on reconstruction[1]

MARTIN SALZMANN

University of Pennsylvania

MARTA WIERZBA

Universität Potsdam

DOREEN GEORGI

Universität Potsdam

In recent experimental work, arguments for or against Condition C reconstruction in A′-movement have been based on low/high availability of coreference in sentences with and without A′-movement. We argue that this reasoning is problematic: It involves arbitrary thresholds, and the results are potentially confounded by the different surface orders of the compared structures and non-syntactic factors. We present three experiments with designs that do not require defining thresholds of 'low' or 'high' coreference values. Instead, we focus on grammatical contrasts (*wh*-movement vs. relativization, subject vs. object *wh*-movement) and aim to identify and reduce confounds. The results show that reconstruction for A′-movement of DPs is not very robust in German, contra previous findings. Our results are compatible with the view that the surface order and non-syntactic factors (e.g. plausibility, referential accessibility of an R-expression) heavily influence coreference possibilities. Thus, the data argue against a theory that includes both reconstruction and a hard Condition C constraint. There is a residual contrast between sentences with subject/object movement,

---

which is compatible with an account without reconstruction (and an additional non-syntactic factor) or an account with reconstruction (and a soft Condition C constraint).

KEYWORDS: A′-movement, binding, Condition C, experimental syntax, German, reconstruction, relative clauses, *wh*-questions

1. BACKGROUND: RECONSTRUCTION IN A′-MOVEMENT

Reconstruction for Condition C as in (1) has played an important role in linguistic theory as a diagnostic for movement:

(1)   *[Which picture of John$_i$]$_1$ do you think he$_i$ likes__$_1$?

The ungrammaticality of (1) follows if the *wh*-phrase containing the R-expression *John* is interpreted in its pre-movement position as in (2). Since *John* is c-commanded by the coreferential pronoun *he*, a Condition C violation obtains. Example (1) is thus ungrammatical for the same reason as **He$_i$ likes this picture of John$_i$.*

(2)   *[Which x]$_1$ do you think he$_i$ likes [x picture of John$_i$]$_1$?

Two aspects of Condition C reconstruction have played a prominent role in the literature. First, they have been claimed to display argument/adjunct asymmetries: only R-expressions inside arguments trigger Condition C effects, while R-expressions inside adjuncts do not, see, e.g. Lebeaux (1991: 211–212):

(3)   (a)   *[Whose claim that John$_i$ likes Mary]$_1$ did he$_i$ later deny __$_1$?
      (b)   [Which claim that John$_i$ made]$_1$ did he$_i$ later deny __$_1$?

(4)   (a)   *[Which pictures of John$_i$]$_1$ did he$_i$ like __$_1$?
      (b)   [Which pictures near John$_i$]$_1$ did he$_i$ look at __$_1$?

The asymmetry has been linked to theta theory. While arguments have to be merged cyclically with their predicates to ensure that they receive the proper thematic interpretation, adjuncts can be introduced after movement, i.e. undergo so-called late merger and thus bleed Condition C.

Second, it has been claimed that only R-expressions contained inside predicates obligatorily lead to Condition C violations, while those inside arguments do not always. This has been linked to the fact that only the former reconstruct obligatorily (see, e.g, Huang 1993; Heycock 1995: 558–561). The contrast in (5) thus partially conflicts with the baseline data in (1), see Huang (1993: 110):

(5)   (a)   ?[How many pictures of John$_i$]$_1$ do you think that he$_i$ will like __$_1$?
      (b)   *[How proud of John$_i$]$_1$ do you think he$_i$ should be __$_1$?

Another contrast that has been discussed prominently concerns movement types. Relative clauses are sometimes claimed to display weaker Condition C effects than *wh*-movement or no Condition C effects whatsoever (Citko 2001: 136):

(6)   The picture of John$_i$ which he$_i$ saw ___ in the paper is very flattering.

The absence of Condition C effects in relatives has been accounted for by means of the matching analysis, where either the RC-internal representation of the external head can be deleted without violating recoverability, see (7a) (Citko 2001), or where vehicle change relates the R-expression inside the head of the relative clause to a pronoun inside the relative clause, see (7b) (Sauerland 2003):

(7)   (a)   The [picture of John$_i$] [$_{CP}$ which [~~picture of John$_i$~~] he$_i$ saw [x ~~picture of John$_i$~~] in the paper] is very flattering.
       (b)   The [picture of John$_i$] [$_{CP}$ which [~~picture of him$_i$~~] he$_i$ saw [x picture of **him$_i$**] in the paper] is very flattering.

While these facts are often cited in the literature, several aspects of Condition C reconstruction are contested. In the following subsections, we will briefly summarize some of the major empirical issues, for both English and German.

1.1 *Contested facts about English*

Apart from the studies to be discussed in the next section, virtually all of the literature on Condition C reconstruction is based on introspective judgments. Against this background, it is unsurprising that there is disagreement on both the basic facts and their theoretical interpretation.

The major issue concerns the general robustness of Condition C reconstruction. While often taken for granted under A′-movement, there is a sizable list of dissenting voices, see, e.g. Heycock (1995) and Fischer (2002, 2004). The examples in (8) are a small selection from data presented in Safir (1999: 609) that are supposed to show the absence of Condition C effects:

(8)   (a)   [Whose criticism of Lee$_i$]$_1$ did he$_i$ choose to ignore ___$_1$?
       (b)   [Which picture of John$_i$]$_1$ does he$_i$ like best ___$_1$?
       (c)   [Most articles about Mary$_i$]$_1$ I am sure she$_i$ hates ___$_1$.
       (d)   [That John$_i$ had seen the movie]$_1$ he$_i$ never admitted ___$_1$.

The second controversial issue concerns the argument–adjunct asymmetry. On the one hand, the empirical contrast has been called into question, on the other hand, putative contrasts have been linked to other factors (see Heycock 1995; Lasnik 1998; Fischer 2004: 161–162). Moreover, analytically, it is not really clear what qualifies as an argument in the nominal domain (see, e.g. Fischer 2002, 2004; Donati & Cecchetto 2011; and the references in Bruening & Al Khalaf 2019: 248).

Huang (1993: 110) observed another empirical complication, namely that the strength of Condition C effects with arguments decreases with increasing distance between R-expression and pronoun. Thus, while in the minimal pair in (5), there was an argument/predicate contrast in that the effect was rather weak with

arguments, the contrast disappears once the coreferential pronoun is in the matrix clause (and thus closer to the R-expression) as shown in (9); in that configuration both arguments and predicates seem to display a Condition C effect:

(9) (a) ?*[How many pictures of John$_i$]$_1$ does he$_i$ think that I like __$_1$?
     (b) ?*[How proud of John$_i$]$_1$ does he$_i$ think I should be __$_1$?

1.2 *Condition C reconstruction in German*

While the issue is somewhat less prominent in the literature, the robustness of Condition C reconstruction seems to be equally contested in German. In the first systematic discussion of Condition C reconstruction, Frey (1993: 143–153) presents evidence for reconstruction of topicalized arguments on the basis of (10):

(10) (a) *Sie hat ihm$_i$ [Peter$_i$ Buch] zurückgegeben.
          she has he.DAT Peter's book   returned
          'She returned him$_i$ Peter$_i$'s book.'
      (b) *[Peter$_i$ Buch]$_1$ hat sie ihm$_i$ __$_1$ zurückgegeben.
          Peter's book   has she he.DAT   returned
          'Peter$_i$'s book, she returned him$_i$.'

The fronted direct object is reconstructed to its base position below the indirect object and therefore causes a Condition C violation. Similarly, in Salzmann (2017: 137) it is argued that Condition C effects are robust in *wh*-movement and topicalization but weak/absent in relativization:

(11) (a) *[Welche Nachforschungen über   Peter$_i$]$_1$ hätte   er$_i$
          which   investigations   about Peter] had.SBJV.3SG he
          dir   lieber __$_1$ verschwiegen?
          you.DAT rather   conceal.PTCP
          'Which investigations about Peter$_i$ would he$_i$ have rather concealed from you?'
      (b) die [Nachforschungen über   Peter$_i$], [die   er$_i$ mir lieber__
          the investigations   about Peter   which he me rather
          verschwiegen hätte]
          conceal.PTCP have.SBJV.3SG
          'the investigations about Peter$_i$ that he$_i$ would have rather concealed from me'

However, as discussed in Fischer (2002: 70–71, 79; 2004; 161–164, 175–177), many of the types of examples that are controversial in English – recall (8) – also don't seem to display strong Condition C effects in German:

(12) [Dass Hans$_i$ verloren hat], hat er$_i$ mir   natürlich __$_1$ verschwiegen.
     that John lost   has has he me.DAT of.course   concealed
     'That John$_i$ had lost he$_i$ didn't tell me, of course.'

(13) [Marias mutwillige Zerstörung von Peters$_i$ Sachen]$_1$ konnte er$_i$ nicht
      Mary's willful   destruction of   Peter's belongings could he   not
      einfach __$_1$ hinnehmen.
      simply   accept
      'Mary's willful destruction of Peter$_i$'s belongings, he$_i$ couldn't accept.'

1.3 *Intermediate summary and objectives*

As this section has shown, while Condition C reconstruction has played a prominent
role in syntactic theory, its force is rather unclear given the empirical controversies
surrounding it. The goal of this paper is thus to provide an empirically more solid
base by investigating Condition C reconstruction from an experimental perspective.
Our focus will be on German because German is less studied in this respect, both
theoretically and experimentally.

   Our paper is structured as follows: In Section 2, we will summarize previous
experimental work on Condition C reconstruction in English and German and
point out shortcomings related to the use of thresholds and concerning the failure
to properly take non-syntactic factors into account. In Sections 3–6, we report our
experiments on reconstruction in German A′-movement. These experiments
are based on grammatical contrasts and successively neutralize possible non-
syntactic factors. Overall, the case for Condition C reconstruction weakens. In the
general discussion in Section 7, we conclude that the results argue against a
theory that includes both reconstruction and a hard/inviolable Condition C
constraint.[2]

2. PREVIOUS EXPERIMENTAL WORK

Despite the empirical controversies, Condition C reconstruction has only recently
been subjected to experimental scrutiny.[3] There are three studies on English,
namely Adger et al. (2017), Bruening & Al Khalaf (2019), and Stockwell,
Meltzer-Asscher & Sportiche (2021) and one study on German of our own
(Wierzba, Salzmann & Georgi 2021). Apart from Stockwell et al. (2021), which
we will address in Section 5 because it involves a different reasoning, we will

---

[2] We will, in what follows, often loosely speak of 'reconstruction' by which we mean that PP
    modifiers of DPs/APs are present in the bottom copy of an A′-moved phrase and that this copy is
    interpreted. Conversely, 'absence of reconstruction' means that no PP modifiers are present in the
    bottom copy. Cyclic versus late merger are established means to capture this difference, but other
    possibilities are conceivable as well.

[3] For an influential comprehensive study on Condition C in sentences without syntactic movement,
    see, e.g. Gordon & Hendrick (1997). For a comparison of different methods in cataphoric
    configurations, see Patterson & Felser (2019) and references cited there. For a study on the
    acquisition of Condition C, see Crain & McKee (1985). For experiments on the role of linear order
    and c-command in Condition C in German, see Bader & Webelhuth (2020).

581

summarize these studies and their major results as they have influenced the questions and methodologies we pursue in what follows.

The arguments in favor or against Condition C reconstruction in these works are based on experiments testing the following constellation:

(14) *Experimental design in previous research*
    (a)   pronoun$_i$ [… [R-expression$_i$ …]]   (in situ)
    (b)   [… [R-expression$_i$] …]$_1$ … pronoun$_i$ __$_1$ (moved)

The basic idea is that Condition C is violated in (14a) (in situ condition) and therefore coreference between the pronoun and the R-expression should be unavailable. If a different pattern – i.e. higher availability of coreference – is observed in (14b) (moved condition), this has been interpreted as evidence against reconstruction; and if a similar pattern is observed in (14b) (low availability of coreference), this has been interpreted as evidence in favor of reconstruction.

2.1 *English*

Adger et al. (2017) report on three experiments that investigate different aspects of Condition C reconstruction, including the difference between predicates (APs) and arguments (DPs), the difference between arguments and adjuncts of nouns (complement clauses versus relative clauses), and the effect of (linear and structural) distance. The participants were presented with matrix *wh*-questions as in (15): R-expression and pronoun were highlighted and the participants were asked in a forced-choice task (yes/no) whether the two could refer to the same individual.

(15)   How proud of Elisabeth is she?

The AP versus DP-contrast was tested by comparing local *wh*-movement of APs with local *wh*-movement of DPs, with the R-expression contained in a complement:

(16) *APs versus DPs*
    (a)   How proud of Elisabeth is she?   (AP with PP complement)
    (b)   Which side of Elizabeth does she prefer?   (DP with PP complement)

The argument–adjunct asymmetry was tested by contrasting *wh*-movement of DPs with the R-expression contained in either a complement clause or a relative clause:

(17) *Complement versus adjunct clauses*
    (a)   Whose claim that Elizabeth is too old did she overhear? (complement)
    (b)   Which allegation that shocked Elizabeth did she deny? (adjunct)

The factor distance contained three levels. SHORT refers to a monoclausal *wh*-question. EMBEDDED 1 refers to a long-distance question with the coreferential

pronoun as the matrix subject. EMBEDDED 2 refers to a long-distance question with the coreferential pronoun as the embedded subject. They are illustrated in (18):

(18) *Distance conditions*
    (a)   Which side of Elizabeth does she prefer?   (short)
    (b)   Which side of Elizabeth does she say Philip prefers?   (embedded 1)
    (c)   Which side of Elizabeth did Philip say she prefers?   (embedded 2)

Finally, the experiments contained control items without movement, which also varied the distance between coreferential pronoun and R-expression:

(19) *In situ conditions*
    (a)   He$_i$ saw that enemy of Superman$_i$'s partner.   (short)
    (b)   He$_i$ thinks Lois saw that enemy of Superman$_i$ (embedded 1)

The major results of the experiments can be summarized as follows: There was a clear contrast between predicates and arguments in that non-coreference was robust in the former, while, in the latter, coreference was available to varying degrees. No clear evidence for an argument–adjunct asymmetry (in the sense that only the former reconstruct) was found, even though coreference was more available with adjuncts. Finally, in all experiments, there was an effect of linear distance – coreference becomes more available the larger the distance between R-expression and coreferential pronoun. The results are illustrated in Table 1 (we omit the results of experiment 2, where the availability of coreference with *wh*-moved DP-PP short was significantly higher, viz., 58.7%; RC = relative clause; CC = complement clause; PP = PP complement).

Exp.	Category	Movement	Distance	Result
1	AP-PP	*wh*-moved	short	8.6%
1	AP-PP	*wh*-moved	embedded 1	22.4%
1	AP-PP	*wh*-moved	embedded 2	36.4%
1	DP-PP	*wh*-moved	short	30.0%
1	DP-PP	*wh*-moved	embedded 1	53.1%
1	DP-PP	*wh*-moved	embedded 2	64.3%
1	DP-PP	in situ	short	11.1%
1	DP-PP	in situ	embedded 1	18.5%
3	AP-CC	*wh*-moved	short	67.5%
3	AP-CC	*wh*-moved	embedded 2	81.7%
3	DP-RC	*wh*-moved	short	88.2%
3	DP-RC	*wh*-moved	embedded 2	87.3%
3	DP-CC	*wh*-moved	short	64.7%
3	DP-CC	*wh*-moved	embedded 2	80.7%

Table 1
Summary of two of the experiments reported in Adger et al. (2017). Percentages in the result column indicate the proportion of cases in which participants responded that coreference between pronoun and R-expression is possible.

583

The authors conclude from these results that only predicates and their (PP) complements reconstruct. With respect to reconstruction of DPs, they argue that all modifiers inside DPs can be deleted in the bottom copy so that DP arguments generally do not cause Condition C effects. This conclusion is based on the high availability of coreference in the conditions with DP movement. The asymmetry between DPs and APs is argued to follow from independent differences in the interpretation of LF-structures between predicates and arguments. The distance effect is finally linked to non-syntactic factors.

Bruening & Al Khalaf (2019) present two experiments on Condition C reconstruction in questions. They focus on the argument–adjunct asymmetry and investigate PP modifiers (experiment 1) as well as CP-modifiers (experiment 2), i.e. relative clauses versus complement clauses. The conditions are tested with *wh*-movement and without. The authors criticize the method of Adger et al. (2017) as inviting the subjects to engage in metalinguistic reasoning. They instead propose a different approach which involves embedded questions with two possible referents for the pronoun: the R-expression in the matrix clause (the subject) and the one within the *wh*-phrase. The participants then had to answer a forced-choice question and had to decide to which of the two R-expressions the pronoun referred (in the in situ condition, the embedded clause was a simple declarative clause):

(20) The chambermaid told me which portrait of the countess she considered to be the most valuable.
Who considers the portrait valuable? ☐ the chambermaid ☐ the countess

The major results of their experiments can be summarized as follows: The authors found a significant contrast between the in situ and moved conditions in that coreference was much more available in the latter. There was no significant contrast between arguments and adjuncts; coreference was more available with CP-modifiers than with PP modifiers. The results are summarized in Table 2.

Exp.	Category	Movement	Distance	Result
1	DP (with CP argument)	in situ	short	4.7%
1	DP (with CP argument)	*wh*-moved	short	42.7%
1	DP (with CP adjunct)	in situ	short	2.7%
1	DP (with CP adjunct)	*wh*-moved	short	56.0%
2	DP (with PP argument)	in situ	short	2.7%
2	DP (with PP argument)	*wh*-moved	short	22.0%
2	DP (with PP adjunct)	in situ	short	1.3%
2	DP (with PP adjunct)	*wh*-moved	short	30.7%

Table 2
Summary of experiments 1–2 reported in Bruening & Al Khalaf (2019). Percentages in the result column indicate the proportion of cases in which response 'B' (i.e. the response that violates Condition C in the conditions without movement) was chosen.

584

In their interpretation of the results, the authors primarily capitalize on the difference between the moved and the in situ conditions. They argue that if Condition C is a hard grammatical constraint, one expects there to be no difference between movement and in situ. However, they do find a substantial difference. In experiment 1, coreference is chosen at a rate close to chance (50%) under movement, while in the in situ condition, coreference was chosen at rates close to zero. This is interpreted as evidence against Condition C reconstruction. In experiment 2, the rate of coreference is below chance level, but since it is significantly higher than in the in situ condition, it is argued that this shows that no grammatical constraint on coreference is at play but rather non-syntactic factors. To capture the pattern, the authors argue that dependents of N should be uniformly treated as adjuncts, which is why they need not be present in the bottom copy.

In addition, Bruening & Al Khalaf (2019) present an experiment on English PP fronting. They observe that a fronted adjunct containing an R-expression does not easily allow coreference with a pronominal subject, as in *The policeman said that near Dan, he saw a snake*, where coreference between *Dan* and *he* was only chosen at a rate of 8.6%. The 8.6% is taken to be close enough to zero to indicate a Condition C violation and the authors therefore conclude that these PPs, including their nominal complement, reconstruct.

2.2 *German*

In Wierzba et al. (2021) we present four experiments on reconstruction for Condition C in German *wh*-questions. Experiment 1 investigates reconstruction of R-expressions contained in predicates. In experiment 2, reconstruction of R-expressions contained in either PP arguments or PP adjuncts of nouns is investigated. Experiments 3 and 4 investigate the effect of distance by testing reconstruction in long-distance movement. The methodology in these experiments was inspired by that used in Bruening & Al Khalaf (2019) and thus also involved embedded questions with two possible antecedents for the pronoun. The major difference was that instead of asking a forced-choice question, participants were asked two questions after each item and had to decide for each R-expression whether it was a possible antecedent of the pronoun (for discussion of this methodological choice, see Section 3.2). A (translated) sample item with questions is given in (21) (here and in what follows Q1 refers to the question about coreference with the matrix R-expression, while Q2 refers to the question about coreference with the embedded R-expression, which is the one within the *wh*-phrase in the moved condition):

(21) Boris told us which report on David he ignored.
      Can this sentence be interpreted such that …
      Q1: … Boris ignored a report?   ☐ yes   ☐ no
      Q2: … David ignored a report?   ☐ yes   ☐ no

585

Exp.	Category	Movement	Distance	Result Q1	Result Q2
1	AP	in situ	short	98.4%	1.6%
1	AP	*wh*-moved	short	98.4%	2.4%
2	DP (with PP arg.)	in situ	short	97.9%	7.6%
2	DP (with PP arg.)	*wh*-moved	short	97.2%	6.9%
2	DP (with PP adj.)	in situ	short	99.3%	4.9%
2	DP (with PP adj.)	*wh*-moved	short	97.9%	11.8%
3	AP	*wh*-moved	short	99.1%	1.9%
3	AP	*wh*-moved	embedded 1	88.8%	16.8%
3	AP	*wh*-moved	embedded 2	76.9%	23.1%
4	DP (with PP arg.)	*wh*-moved	short	100.0%	11.1%
4	DP (with PP arg.)	*wh*-moved	embedded 1	89.8%	26.9%
4	DP (with PP arg.)	*wh*-moved	embedded 2	73.1%	45.4%

Table 3
Summary of the four experiments reported in Wierzba et al. (2021), omitting conditions with coordination. Percentages in the result columns indicate the proportion of cases in which participants responded that coreference is possible. Question Q2 targets the interpretation that violates Condition C in the conditions without movement.

The major results of this study are the following: Coreference was disfavored with both APs and DPs. This is particularly obvious in the short conditions, where the difference between in situ and moved is rather small. Coreference was more available with adjuncts than with arguments, but the difference is numerically very small. As in the experiments by Adger et al. (2017), there was a distance effect in that coreference with the embedded R-expression becomes more available with increasing distance between *wh*-phrase and coreferential pronoun. The results are illustrated in Table 3.

We concluded in Wierzba et al. (2021) that these results are compatible with the view that A′-moved constituents (including the PP modifiers) reconstruct, with the caveat that the result is based on a null effect (the lack of a difference between the in situ and moved conditions). Numerically, DPs/arguments showed more positive responses to Q2 than APs/predicates. However, since they were not tested within the same experiment, no firm conclusions could be drawn. The results did not provide conclusive evidence for an argument–adjunct asymmetry and, thus, a late merger approach. There was a small difference in the short condition in the predicted direction, but, even with adjuncts, coreference was available only to a very limited extent. The higher availability of coreference in the long-distance conditions was attributed to processing difficulties.

In a later, up to now unpublished, follow-up experiment, we tested reconstruction of APs and DPs within a single design to find out whether the predicate–argument asymmetry could be confirmed. We will report on this experiment (henceforth: AP/DP experiment) very briefly here as this asymmetry is not the main focus of this paper; more details can be found in our data repository (see the link in Section 3.3).

Category	Movement	Distance	Q1	Q2
AP	in situ	short	97.9%	3.5%
AP	moved	short	97.2%	5.6%
AP	in situ	embedded 1	97.2%	13.9%
AP	moved	embedded 1	87.5%	27.1%
AP	in situ	embedded 2	94.4%	4.2%
AP	moved	embedded 2	76.4%	20.1%
DP	in situ	short	97.9%	4.9%
DP	moved	short	95.1%	20.8%
DP	in situ	embedded 1	98.6%	13.2%
DP	moved	embedded 1	92.4%	29.9%
DP	in situ	embedded 2	95.8%	10.4%
DP	moved	embedded 2	72.2%	45.1%

Table 4
Proportion of positive answers to the coreference questions in the AP/DP experiment.

36 native speakers of German, recruited via prolific.co, took part by way of the platform L-Rex (Starschenko & Wierzba 2020); the same two-question method as in Wierzba et al. (2021) was used. The experiment had a $2 \times 2 \times 3$ design with the factors MOVEMENT (moved vs. in situ), CATEGORY (AP vs. DP; sum-coded) and DISTANCE (short, embedded 1, embedded 2). With DPs, the R-expression was contained in a PP argument. The main results are the following: As in Wierzba et al. (2021), coreference was disfavored with both APs and DPs, but it was significantly more available with DPs.[4] Also, as in Wierzba et al. (2021), there was a distance effect in that coreference increases in the embedded 1/2 conditions. The results are listed in Table 4.

The results are in line with the predicate-argument hypothesis in that coreference is less available with APs than with DPs. The distance effect is most likely unrelated to reconstruction since it affects APs and DPs, even some of the in situ conditions and there is a concomitant decrease in the availability of coreference with the matrix subject in the embedded 1 and 2 conditions. The responses to both Q1 and Q2 are closer to chance level in embedded 1/2; a possible interpretation is that participants found it more difficult to judge the interpretation possibilities in

---

[4] The statistical results reported here are based on generalized linear mixed models. The AP/DP asymmetry emerges as an interaction of MOVEMENT and CATEGORY within the baseline level (short) of the treatment-coded factor distance ($z = 1.973$, $p = 0.049$). The difference is larger in the embedded 2 structures than in the short baseline (significant interaction between DISTANCE and MOVEMENT: $z = -2.287$, $p = 0.022$). For embedded 1, an overall higher proportion of positive responses to Q2 was found (interaction between DISTANCE and CATEGORY: $z = 2.357$, $p = 0.018$), but this equally affected in situ and moved structures (no significant interaction between DISTANCE and MOVEMENT at this level of DISTANCE: $z = -0.136$, $p = 0.891$). No significant three-way interactions were found. For all relevant statistical data, see the link in Section 3.3.

587

these more complex cases.[5] It is less clear what the results imply for the reconstruction of DPs. Coreference is clearly higher in the moved condition than in the in situ condition, which may suggest the absence of reconstruction (viz., late merger of modifiers). Compared to the first four experiments by Wierzba et al. (2021), coreference is more available, even though the follow-up experiment was based on the same materials. We will come back to the interpretation of the DP reconstruction data at several points in this paper.

2.3 *Problems of the threshold-based reasoning and non-syntactic factors*

As described in (14), experimental research on Condition C reconstruction has relied on a comparison of the availability of coreference in structures with and without movement. Depending on the result, different conclusions are drawn:

(22) *Types of conclusions in previous research*
    (a) Low availability of coreference in situ, high availability with movement → no reconstruction (Adger et al. 2017: Bruening & Al Khalaf 2019 on DPs)
    (b) Low availability of coreference in situ, low availability with movement → reconstruction (Bruening & Al Khalaf 2019 on PPs; Wierzba et al. 2021 on DPs)

In what follows, we want to point out potential problems with both types of conclusions, having to do with the interpretation of the terms 'low' and 'high'.

We will start with the first type of reasoning based on an asymmetry between movement and in situ. Given the significantly higher availability of coreference under movement, Bruening & Al Khalaf (2019) conclude that there is no reconstruction for Condition C (conclusion type (22a)). The difference that Adger et al. (2017) report between in situ and moved DPs is interpreted the same way.

There is, however, a question that remains open with respect to these findings. Bruening & Al Khalaf (2019) predict that the responses in the conditions without a surface violation of Condition C should allow for both interpretations and, thus, responses at chance level (around 50% for each interpretation) are expected in their forced-choice paradigm; however, for DP movement with PP arguments/adjuncts, the observed values are between 22% and 31%. In Adger et al.'s (2017) study, in which there was only one R-expression and participants judged whether coreference is possible, one might expect the proportion of positive responses to approach 100% in the absence of a grammatical violation, but in their first experiment on DPs, the observed values are between 30% and 64%. Bruening & Al Khalaf (2019: 257)

---

[5] Adger et al. (2017) and Wierzba et al. (2021) discuss an explanation of the distance effect in terms of vehicle change (where the R-expression is replaced by a pronoun in the lower copies) and conclude that the facts are not compatible with it as it wrongly predicts coreference with APs in embedded 2 to be less available than in embedded 1 (because of a Condition B effect in the former).

argue that the reason for the discrepancy between the expected and observed responses cannot be core-syntactic, because values close to zero would be expected if the violation of a hard grammatical constraint is involved. They suggest that the discrepancy might instead have to do with linear distance: coreference between R-expression and pronoun might be dispreferred because they are very close to each other in these conditions (recall also the effect of linear distance observed in Adger et al. 2017 and Wierzba et al. 2021).

Thus, the argumentation rests on the assumption that coreference values that are not close to zero but also not as high as the expected chance level are compatible with a scenario in which there is no reconstruction and extra-syntactic factors cause a decrease in positive responses (reducing it from ~50% to ~20–30%), but incompatible with a scenario in which there is reconstruction and other factors cause an increase in positive responses (raising it from ~0% to ~20–30%).

In our view, however, there are several conceivable scenarios in which reconstruction does play a role, but nevertheless we do not find complete unavailability of coreference in the experiments. One possibility is that there is inter-speaker variation, with some participants generally employing reconstruction, while others do not. A second possibility is that the argument/adjunct status and, thus, reconstruction behavior of PP modifiers may vary between items and/or participants. A third possibility is that non-syntactic factors interact with the binding principles and in some cases even override them (see below). This could distort the interpretation of the results, even if the in situ and moved versions of the same type of sentence are directly compared – these versions inevitably differ not only with respect to the surface syntactic relations, but also with respect to linear order (anaphoric vs. cataphoric relation – the latter typically being dispreferred; see also Yoshida, Potter & Hunter 2019: 1535–1539) and distance between R-expression and pronoun. Thus, it is conceivable that asymmetries between the two versions are not (necessarily) due to a difference in the syntax but may be caused by surface-oriented extra-syntactic factors.[6]

We now turn to problems with reasoning that is based on symmetry between movement and in situ: In Bruening & Al Khalaf (2019) on PP fronting in English, in experiments 1–4 in Wierzba et al. (2021) on *wh*-movement of DPs and APs in German as well as the AP/DP experiment, low availability of coreference ('close to zero') in sentences with movement is interpreted as evidence for reconstruction for Condition C (conclusion type (22b)). There are at least two problems with this reasoning. First, as mentioned above (and discussed in more detail below), other – non-syntactic – factors might disfavor the relevant reading (viz., coreference with the embedded R-expression) independently and lead to low coreference values even

---

[6] Note that the scenario where non-syntactic factors lead to an increase of coreference despite reconstruction and a Condition C violation need not be interpreted as a grammatical illusion; if Condition C is interpreted as a pragmatic condition or as a soft/violable syntactic constraint, as e.g. argued for in Gor & Syrett (2019), such a scenario is, in principle, just as plausible as one where non-syntactic factors lead to a decrease of coreference in the absence of reconstruction.

589

in the absence of reconstruction. Second, a certain amount of random noise is always expected in behavioral data, and it is difficult to define a systematic threshold at which a value is or is not close enough to zero in absolute terms. Thus, in the experiments by Bruening & Al Khalaf (2019), 8.6% (in the PP-fronting experiment) is indeed closer to zero than 22% (PP arguments inside DPs), but what about the 11.8% found for the PP adjuncts inside DPs in Wierzba et al. (2021) reported above, and what if the values were around 15%? Note also that experiments 1–4 in Wierzba et al. (2021) and the AP/DP experiment are based on the same materials, but the coreference values for the reconstruction of DPs (with PP arguments) in the short condition vary between 6.9% (experiment 2), 11.1% (experiment 4), and 20.8% (AP/DP experiment). Given the threshold logic, one would have to conclude that there is reconstruction of DPs in experiment 2 (where the values are similar to the PP cases in Bruening & Al Khalaf 2019) but probably not in the AP/DP experiment (where the values are close to the DP-cases in Bruening & Al Khalaf 2019). It should be clear that this will quickly lead to contradictions. To a large extent, then, setting a threshold at a certain value will be arbitrary. In fact, Adger et al. (2017) use a different criterion in the interpretation of their first experiment, namely whether coreference is accepted in the majority of cases, viz., above 50% (interpreted as evidence against reconstruction) or less (interpreted as compatible with reconstruction). This criterion would imply reconstruction for experiments 1–4 in Wierzba et al. (2021) and the AP/DP experiment as well as for DP movement with arguments in Bruening & Al Khalaf (2019), in partial conflict with their conclusions.[7]

The issue with thresholds is related to a more general problem that arises in the experimental investigation of reconstruction: when sentences with/without movement are compared, the hypothesis that there is no reconstruction predicts a difference, but the hypothesis that there IS reconstruction basically predicts the lack of a difference, i.e. a null effect, which is more difficult to interpret.

In addition to the issues raised by threshold-based reasoning, another shortcoming of previous work is that it does not sufficiently take into account the influence of non-syntactic factors, especially factors that generally govern pronoun resolution and can increase or decrease the availability of coreference (see also Gor 2020). In what follows we will list the factors that have received most attention in the literature and discuss their implications for the current debate.[8]

---

[7] We would like to stress that our objections to the reasoning in Bruening & Al Khalaf (2019) primarily target the threshold logic and not so much their conclusion that the data argue against Condition C reconstruction with DPs. In our view, the main problem arises with their PP experiment which, given the problems with thresholds, cannot clearly be interpreted as showing that there is Condition C reconstruction. But once one can no longer be certain about the status of seemingly uncontroversial cases of Condition C reconstruction, the interpretation of the controversial cases (viz., reconstruction of DPs) becomes difficult as well.

[8] Bruening & Al Khalaf (2019) do discuss non-syntactic factors at two points. As mentioned above, to account for the reduced availability of coreference in the moved condition, they refer to linear distance, implying that coreference is less likely if R-expression and pronoun are very close to each other. We will see in Experiments 2 (S/O) and 3 (non-syn) that such a factor is unlikely to play any role in these experimental settings. It also clashes with the literature that has shown that

590

First, it has been shown that the more prominent an expression is in a certain hierarchy, the more likely it is to act as an antecedent. This can involve prominence with regard to thematic role (agent > patient > other), grammatical function (subject > object > other) or information structure (topics are preferred antecedents), see, e.g. Grosz & Sidner (1986), Brennan (1995), Cowles, Walenski & Kluender (2007), Kaiser (2011), and Schuhmacher, Dangl & Uzun (2016) for German. Second, there is work showing that if there are two similarly salient antecedents for a personal pronoun, there is a preference for coreference with the linearly closer antecedent, see, e.g. Cunnings, Patterson & Felser (2014). A third factor is plausibility. As Gor & Syrett (2019) and Gor (2020) demonstrate, it can even override Condition C in backward anaphora.

It is likely that these factors affect the judgments in some of the experimental settings. With respect to the design in Bruening & Al Khalaf (2019), Wierzba et al. (2021), the AP/DP experiment, as well as Experiments 1 and 2 below, preferences of pronoun resolution may lead to a preference for the embedded subject pronoun to corefer with the matrix subject. Thus, the availability of coreference with the embedded R-expression is likely to be decreased. Proximity may have the opposite effect in these designs. Given that the R-expression within the *wh*-phrase is the closer antecedent, this factor could increase the availability of coreference with the embedded R-expression. The factor plausibility can play a role in all experimental designs discussed in this paper. It can either lead to an increase or a decrease in the availability of coreference, depending on the item.[9]

Given the problems with threshold-based reasoning and the probable influence of non-syntactic factors on coreference judgments, we believe that we currently do not fully understand what the data tell us. The aim of this paper is thus to develop an experimental design allowing us to determine more precisely to what extent the observed patterns actually reflect reconstruction for Condition C. We will investigate grammatical contrasts that inform us about Condition C without requiring reference to absolute coreference values. We will also investigate the influence of some non-syntactic factors on coreference judgments. We will focus on those that the syntactic literature and the literature on pronoun resolution have shown to be most influential.

---

proximity/recency in fact facilitates coreference. In their Section 6, they discuss (but do not investigate experimentally) a pragmatic bias that affects coreference that we will come back to in our discussion of Experiment 3.

[9] To illustrate the potential effect of plausibility, compare, e.g. the following items from experiment 2 by Bruening & Al Khalaf (2019):

    (i)   (a)   A literature professor explained which unauthorized biography of Putin he was most angry about.

           (b)   The assistant didn't know which evaluation from the department head's office he should submit as part of a periodic review.

Intuitively, coreference between the pronoun and the lower R-expression (*Putin*) seems to yield a more plausible interpretation in (ia), whereas coreference with the higher R-expression (*the assistant*) seems to be the more likely reading in (ib).

3. EXPERIMENTS: PRELIMINARY REMARKS

3.1 *Motivation and outlook*

In the following sections, we present three experiments that investigate Condition C reconstruction in German A′-movement. They are based on grammatical contrasts and do not require reference to absolute coreference values. We attempt to successively neutralize possible non-syntactic factors. We will see that the case for Condition C reconstruction weakens once factors like plausibility and referential accessibility are taken into account. The results clearly argue against a theory that includes both reconstruction and a hard/inviolable Condition C constraint. There remains some residual evidence that the base position matters, the implications of which are explored in the general discussion.

3.2 *Methodological remarks*

In our experiments, we used the method introduced in Wierzba et al. (2021). Participants were told they were going to see one sentence and two questions on each page of the questionnaire. They were instructed that the sentence might have more than one interpretation and that they were going to be asked whether certain interpretations of the sentence are possible or not. The task was illustrated using the example *Maria hat Anna besucht, weil sie nett ist* 'Mary visited Anna because she is nice'. Participants were explicitly told that this sentence has two interpretations (even if one might be more readily available), and that in an example like this they should answer 'yes' to both presented questions ('Can the sentence be interpreted such that … (i) Mary is nice (ii) Anna is nice'). The instructions also stated that both potential interpretations should be carefully considered and that sometimes one, both, or neither of them might be available. Each following page of the questionnaire looked as follows:

(23)   *Target sentence*
     Kann man den Satz so interpretieren, dass …
     'Can this sentence be interpreted such that…'
     … Q1?   ☐ Ja 'yes'   ☐ Nein 'no'
     … Q2?   ☐ Ja 'yes'   ☐ Nein 'no'

The target sentence is a construction with two possible antecedents for the pronoun, the R-expression in the matrix clause (the matrix subject) and the second R-expression, which is either in the embedded clause (within the *wh*-phrase in the case of embedded questions) or inside the external head of a relative clause, e.g. as in *Boris told us which report on David he ignored/Boris mentioned every report on David that he ignored*. In Experiment 3, the design is slightly modified in that the R-expression inside the *wh*-phrase is already introduced together with the other R-expression in a sentence preceding the indirect question. The participants then have to decide for each referent whether it is a possible antecedent or not by

592

answering two yes/no questions, e.g. Q1 *Can this sentence be interpreted such that Boris ignored a report?* / Q2 *Can this sentence be understood such that David ignored a report?* The order of presentation of the two questions was balanced: in half of the stimuli, Q1 appeared above Q2, and in the other half it was the other way round. The R-expressions we used were exclusively common first names.[10]

Our design is a modification of that used in Bruening & Al Khalaf (2019), where a forced-choice question was asked and speakers had to choose between the matrix and the embedded R-expression. The major reason for our modification is to ensure we can determine which coreference options are available and which are not, even in the presence of non-syntactic factors favoring one of the readings.

In the forced-choice task used by Bruening & Al Khalaf (2019), participants have to pick one of the readings, even if neither of the options violates Condition C and thus both should be available from a syntactic point of view. However, non-syntactic factors (which will be discussed in more detail in connection with Experiments 2 and 3) can favor one of the readings: one would thus not necessarily expect a 50% : 50% distribution of answers even in the absence of a Condition C violation, but there might be a preference toward one of the options for independent reasons. The benefit of the forced-choice approach is that if participants do choose the other option in many cases (contra the expected preference) – as found by Bruening & Al Khalaf (2019) in the sentences with *wh*-movement – a strong argument can be made in favor of the view that both options are in fact available and neither of them violates a grammatical principle. The disadvantage of the method, however, is that if a preference of almost 100% for one of the options was found, it would be difficult to determine whether this is just due to the fact that this is the preferred reading (while the other is grammatical, but dispreferred), or whether the other option is really completely excluded on syntactic grounds. In this case, asking two questions – whether reading A is possible, and whether reading B is possible – can provide the crucial information that would be missing in a forced-choice task. In our view, the two-question method is thus better suited for our purposes: in the German sentences that we aim to investigate, we suspect that non-syntactic factors might play a major role, and it is thus particularly important to choose a method that allows us to determine possible rather than preferred readings.

This reasoning is supported by an experiment that we conducted in order to compare the two methods. We replicated the AP/DP experiment with the forced-choice method (see the Appendix A.1 for details). In the replication, we found that coreference with the embedded R-expression was extremely close to zero with both

---

[10] In Bruening & Al Khalaf (2019), definiteness and/or prominence were not systematically controlled for: the R-expressions inside the *wh*-phrase were always definite, while the R-expressions in the matrix clause were sometimes indefinite. Additionally, the R-expressions inside the *wh*-phrase often referred to more prominent individuals than the matrix R-expressions (e.g. Hillary Clinton, Putin, president, Queen vs. reporter, secret service agent, literature professor, female aide). Using only first names avoids these potential confounds.

593

APs/DPs, viz., 0% (APs) and 0.7% (DPs). Under the threshold-based logic, this would imply a Condition C violation and the lack of a predicate–argument asymmetry. The comparison with the AP/DP experiment with the two-question method, where the corresponding value was at 20.8% with DPs, suggests that the forced-choice method is indeed too coarse when strong non-syntactic factors are present. The limits of the forced-choice method become even more visible once subject questions are used where Condition C is not at stake (*John wonders which picture about Bill pleased him*). In an exploratory experiment (see the Appendix A.3 for details), the two-question method showed a preference for coreference with the matrix subject, but coreference with the embedded R-expression was also highly available (58%) (as in Experiment 2 below). Under the forced-choice method, however, coreference with the embedded R-expression was chosen only in 5.6% of the cases. Interpreting the low value as a Condition C effect would be an obviously wrong conclusion in this case. In our Experiment 3 below we have reduced the bias in favor of the matrix R-expression by introducing the embedded/lower R-expression in the prior linguistic context.[11]

As we will see in the results of the experiments discussed below, the two-question method worked as intended in that participants were willing to give two 'yes' or two 'no' responses in this type of task (depending on the item/filler); they were thus not biased toward giving exactly one positive response, suggesting that they were evaluating both options and did not end up interpreting the instructions as a forced-choice task after all. We will also see that there can be some interaction between the two questions in that a strong preference for coreference with the matrix subject (high percentage of yes-answers to Q1) can decrease the amount of *yes*-answers to Q2, even if that corresponds to a perfectly grammatical option. But in all cases, the combined percentages clearly exceed 100%. The same pattern can be found in the results for the fillers. We will report on two groups of fillers that served as controls in this respect. The first group involves ambiguous relative clauses (*Leyla erzählt, dass die Verwandte, die sie besucht hat, in Budapest wohnt* 'Leyla tells us that the relative [who she visited/who visited her] lives in Budapest' with the question whether it can be understood such that Leyla/the relative was visited), for which we expected two positive responses. The results are in line with this expectation: in the AP/DP experiment we found 83.3% positive responses to Q1 and 89.8% for Q2. The second selected group are sentences for which we expected two negative responses (*Gustav erwähnte, dass Karl und Jonas ihn Bücher einscannen ließen* 'Gustav mentionend that Karl and Jonas had him scan books' with the question whether Karl/Jonas did the scanning). The results in the AP/DP experiment showed the same proportion of positive responses (7.6%) for both Q1

---

[11] An alternative way to avoid the bias for coreference with the matrix subject is to construct items without a second referent as in Adger et al. (2017) or Stockwell et al. (2021). In the latter, two possible interpretations have to be judged on a 7-point Likert scale, one indicating that pronoun and R-expression are coreferential (without referring to coreferentiality) and one indicating that the pronoun refers to someone else.

594

and Q2. This shows that the task worked as intended and did not induce a bias to give exactly one positive response.[12] The method's reliability is also supported by the fact that the proportions observed in our critical items are similar in Wierzba et al. (2021), the AP/DP experiment and in Experiment 1 for the same conditions.

3.3 *Supplementary materials*

The materials, raw results files, and analysis scripts for all experiments reported here can be found on OSF under the following link: https://osf.io/24xh3

4. EXPERIMENT 1: *WH*-QUESTIONS VERSUS RELATIVE CLAUSES

As mentioned above, the major goal of this paper is to investigate Condition C reconstruction without having to refer to absolute coreference values. We instead develop designs that are based on grammatical contrasts from which we can draw conclusions about the reconstruction behavior of A′-movement. In Experiment 1, we compared two constructions involving A′-dependencies: *wh*-questions (with DP movement) and relative clauses. As discussed in Section 1, it has been proposed in the literature that they differ in their reconstruction behavior. Our aim was to test whether reconstruction for Condition C is indeed less robust in relative clauses in comparison to *wh*-questions. If this is the case, it can provide an argument in favor of the view that coreference under A′-movement is (also) constrained by grammatical factors, such as by movement type.

4.1 *Participants and procedure*

Participants were recruited via prolific.co and 32 native speakers of German took part. A web-based questionnaire was set up using SoSciSurvey (Leiner 2018). The basic procedure was as in Wierzba et al. (2021) and the AP/DP experiment where participants were asked to answer two questions with regard to coreference possibilities.

In addition to the coreference questions, we asked participants to rate the sentence on a 1–7 scale (as in experiments 3–4 in Wierzba et al. 2021). The ratings were collected to check whether any problems were introduced by potentially low acceptability of some of the tested conditions: long-distance movement and, in particular, long relativization are often perceived as degraded in German.

A total of 76 stimuli were presented to each participant (32 critical items, 32 fillers, and 12 exploratory items for additional research questions). For the critical items, 128 data points were collected per condition/question (four from each participant). On average, the questionnaire took about 25 minutes to complete.

---

[12] In Experiment 1, the percentages were 76.6%/92.2% (first type of controls) and 8.6%/10.9% (second type); in Experiment 2 83.3%/90.6% (first type) and 11.9%/12.5% (second type of controls).

595

4.2 *Design and materials*

Experiment 1 had a 2 × 4 design. The first manipulated factor was DEPENDENCY (*wh*-question vs. relative clause). The second factor was DISTANCE. We tested the same levels as in Wierzba et al's (2021) experiments 3–4 (short, embedded 1, embedded 2, and coordination). We are omitting the level coordination here for presentational reasons – to facilitate visual comparison of the AP/DP experiment and Experiment 1 and to avoid discussing effects that are tangential to the main research questions of this paper and would require digressing exposition. All data, including all levels of DISTANCE, were included in the statistical analyses reported below. The remaining six conditions are illustrated in the sample item in (24).

(24)   *Matrix clause*: **Mark** erwähnte … 'Mark mentioned …'
    (a) *Wh-question, short*
      … [welche Bemerkung über **Ben**] er __ mitbekommen hat.
        which   comment   about Ben he   overheard   has
      '… which comment about Ben he overheard.'
    (b) *Wh-question, embedded 1*
      … [welche Bem. über **Ben**] er denkt, dass du __ mitbekommen hast.
      '… which comment about Ben he thinks that you overheard.'
    (c) *Wh-question, embedded 2*
      … [welche Bem. über **Ben**] du denkst, dass er __ mitbekommen hat.
      '… which comment about Ben you think that he overheard.'
    (d) *Relative clause, short*
      jede Bemerkung über **Ben**, die   er __ mitbekommen hat.
      every comment   about Ben which he   overheard   has
      '… every comment about Ben that he overheard.'
    (e) *Relative clause, embedded 1*
      … jede Bem. über **Ben**, die er denkt, dass du __ mitbekommen hast.
      '… every comment about Ben which he thinks that you overheard.'
    (f) *Relative clause, embedded 2*
      … jede Bem. über **Ben**, die du denkst, dass er __ mitbekommen hat.
      '… every comment about Ben which you think that he overheard.'

The corresponding questions Q1 and Q2 were *Kann man den Satz so verstehen, dass Mark/Ben eine Bemerkung mitbekommen hat?* 'Can this sentence be interpreted such that Mark/Ben overheard a comment?' (questions) and *Kann man den Satz so verstehen, dass Mark/Ben die Bemerkungen mitbekommen hat?* 'Can this sentence be interpreted such that Mark/Ben overheard the comments?' (RCs).

    All items involved either a *wh*-question or a relative clause; no in situ versions were included. The relative clause heads were preceded by the universal quantifier *jede/jeder* 'every' to ensure a restrictive reading of the relative clauses. The

596

R-expression was always included in a PP argument to the noun.[13] We adopted most of the materials from Wierzba et al's (2021) experiment 4, but replaced the matrix verb *erzählen* 'tell' with *erwähnen* 'mention' so that it would be compatible with both CP-complements (embedded questions) and DP-complements (relative clauses). We also changed some of the proper names and nouns to ensure that the interpretation of the relative pronoun was unambiguous (with respect to number and gender), i.e. it was only compatible with the head noun.

4.3 *Hypotheses and predictions*

The matching analysis of relative clauses predicts the absence of Condition C effects since no R-expression is present in the RC-internal copy that is c-commanded by the pronoun (recall (7)). If this hypothesis is correct, we expect a significant effect of the factor DEPENDENCY with respect to Q2 (the question asking about coreference between the embedded R-expression and the pronoun): the proportion of positive responses to Q2 should be higher for relative clauses than for *wh*-questions. If a different derivation underlies relative clauses, viz., the raising analysis, where there is a full copy of the external head inside the RC, we expect no asymmetry between questions and relative clauses.

As in the AP/DP experiment, the embedded 1 and embedded 2 conditions were included in order to make the design parallel to the previous studies discussed in Section 2 and to reassess Wierzba et al's (2021) proposal that the effect of structural distance can be attributed to non-syntactic factors.

4.4 *Results*

The results are summarized in Figure 1 and Table 5. The statistical results reported in this paper are based on generalized linear mixed models (GLMMs).[14]

For the analysis of Experiment 1, the factor DEPENDENCY was sum-coded, while DISTANCE was treatment-coded with short as the baseline. This means that with respect to dependency, we will treat both levels – relative (rel) versus *wh* – symmetrically (comparing each of them to the overall mean). For distance, we will be making the following comparisons: short versus embedded 1, short versus embedded 2. This type of contrast coding means that the model output for

---

[13] About half of the nouns were event nominals (*ung*-derivations), while the other half was underived (e.g. 'statue', 'portrait', 'rumor') or verb-related ('anger', 'hate', 'attack'). To avoid a coreferential implicit PRO the nouns we used were either unaccusative or such that a potential implicit agent would be disjoint as, e.g. with 'rumor'. The PP arguments mostly involved prepositions selected by the noun (*an* 'at/to', *über* 'about', *für* 'for' …) rather than just *von* 'of'. Of course, it is contested whether nouns take arguments at all. Our classification of argument versus adjunct PP is based on the type of examples that have been argued to display the contrast.

[14] The models were fit following the recommendations for identifying parsimonious models by Bates et al. (2015a) and using the R packages lme4 and lmerTest (Bates et al. 2015b; R Core Team 2016; Kuznetsova, Brockhoff & Christensen 2017).
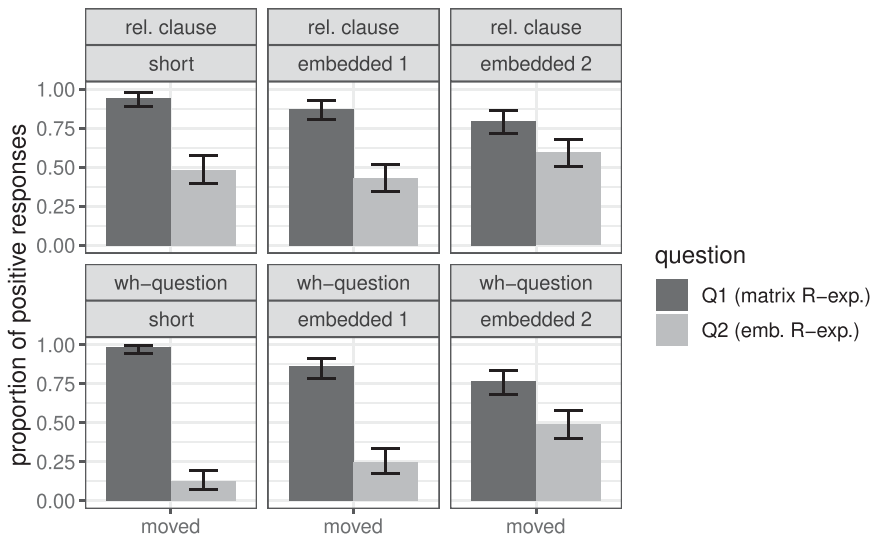
597

Figure 1
Proportion of positive answers to the coreference questions in Experiment 1. Error bars represent 95% Clopper–Pearson confidence intervals.

Dependency	Distance	Rating	Q1	Q2
wh-question	short	6	98.5%	12.5%
wh-question	embedded 1	3	85.9%	25.0%
wh-question	embedded 2	2	76.6%	49.2%
relative clause	short	6	94.5%	48.4%
relative clause	embedded 1	3	87.5%	43.0%
relative clause	embedded 2	3	79.7%	59.4%

Table 5
Proportion of positive answers to the coreference questions and median acceptability ratings (1–7 scale) in Experiment 1.

DEPENDENCY will represent a simple effect – the difference between relative clause and *wh*-question within the baseline level of distance (short) – giving us an impression of the difference between the categories in the basic case. The interaction terms (DEPENDENCY*DISTANCE) indicate if increased distance between R-expression and pronoun changes the basic difference between relative clauses and *wh*-questions.

According to the GLMM, there was a significant simple effect of DEPENDENCY (*wh*-movement vs. relativization) at the short baseline level of DISTANCE ($z = 6.672$, $p < 0.001$) with respect to Q2. There was a significant interaction between DEPENDENCY and DISTANCE at the other levels in comparison to the short baseline (embedded

598

	Estimate	Std. Error	*z*-value	*p*-value
distance2	−0.9848	0.2784	−3.537	0.000404
distance3	0.4070	0.2406	1.692	0.090681
distance4	1.5685	0.2390	6.564	5.25e-11
dependency	1.2495	0.1873	6.672	2.53e-11
distance2*dependency	−0.5746	0.2771	−2.074	0.038098
distance3*dependency	−0.7040	0.2412	−2.919	0.003507
distance4*dependency	−0.9725	0.2362	−4.117	3.83e-05

Table 6
Summary of fixed effects in the GLMM output for Experiment 1. Dummy variables: distance2 = coord vs. short, distance3 = embedded 1 vs. short, distance 4 = embedded 2 vs. short.

1: $z = 2.919$, $p = 0.004$; embedded 2: $z = 4.117$, $p < 0.001$) in the direction of a less pronounced difference between the two dependency types. The model results for the fixed effects are shown in Table 6.

4.5 *Discussion*

The results lend support to the hypothesis that the two types of DEPENDENCY differ regarding reconstruction: coreference between the lower R-expression and the pronoun is more available in relative clauses than in *wh*-questions. The higher availability in relative clauses is compatible with the matching analysis of relativization where there is no full representation of the external head in the RC-internal bottom position. Both versions of the matching analysis that we discussed (based on recoverability/vehicle change; recall (7)) predict the absence of a Condition C effect and thus higher availability of coreference in relative clauses. The raising analysis fails to predict the *wh*-relativization asymmetry.

   With respect to the factor DISTANCE, it is notable that *wh*-questions are more affected, while the percentages in relativization are quite similar in the three conditions (and do not increase monotonically with increasing distance). In addition, as in the AP/DP experiment, visual inspection suggests a decrease in positive answers to Q1 with increasing distance, which affects both dependency types.

   In line with similar findings reported by Wierzba et al. (2021), inspection of the acceptability ratings that were collected in Experiment 1 suggests that the effect is independent of how acceptable participants found these structures. Post hoc inspection of the data, in which we divided participants into three groups based on their acceptability rating for the embedded 1/2 conditions, revealed similar patterns in the coreference judgments across all groups. We interpret this as support for the view that coreference is generally more difficult to judge in the more complex structures, especially in embedded 2.

   Given that there is no Condition C effect under the matching analysis, our relativization examples are predicted to be fully grammatical. It may thus be surprising that the rate of positive answers to Q2 remains between 48% and 59%

599

rather than approaching 100%. There are two reasons suggesting that rates around 50% for Q2 may be close to the maximum one will obtain for grammatical sentences with this experimental setting. First, the rates for coreference in relativization are not much affected by distance, in contrast to what we observe for *wh*-movement; this points toward a ceiling effect. Second, we will see in Experiment 2 that even in in situ conditions without a Condition C violation, i.e. examples that are indisputably grammatical, the positive responses for Q2 remain between 56% and 66%.

There are two related remaining questions: First, is the difference between relative clauses and *wh*-questions really due to a difference in the syntactic structure, viz., the presence/absence of an R-expression in the bottom copy, or could it be caused by other syntactic or non-syntactic factors? As for other syntactic factors, an anonymous reviewers suggests that the fact that the R-expression is contained in an A-position in relative clauses (rather than in an A′-position in questions) could be responsible for the asymmetry. While this is indeed a syntactic difference (and RCs are similar to A-movement with regard to Condition C reconstruction), we are not aware of any syntactic accounts where this difference would translate into a Condition C asymmetry. Note also that this suggestion seems to imply that the reconstruction behavior in RCs would change if the head noun were A′-moved. However, we are not aware of any such effects. As for non-syntactic effects, as far as we can see, it is unlikely that the difference is related to plausibility (we do not see a straightforward reason why interpreting *Peter* and *he* as coreferential should be less plausible in 'John mentioned which statue of Peter he saw' than in 'John mentioned every statue of Peter that he saw'). An (inevitable) difference between the conditions is that the R-expression and the pronoun are directly adjacent in the *wh*-question, whereas one word (the relative pronoun) intervenes in the relative clause. We consider it unlikely that this is responsible for the difference and will come back to the issue of closeness in the discussion of Experiments 2 and 3 where we will see that high coreference values are certainly possible if *wh*-word and pronoun are adjacent. There is one factor that may indeed be at work here, though, namely the referential accessibility of (the phrase containing) the R-expression. We will discuss this factor in Experiment 3 and return to the implications for the *wh*-relativization contrast in the general discussion.

The second question concerns the interpretation of the results for *wh*-questions. We find similar values as in Wierzba et al. (2021) and the AP/DP experiment (confirming the reliability of the method): again, the availability of coreference with the lower R-expression is relatively low, but not at floor. Given the contrast with relativization, one a priori possible interpretation is that this indicates that there is reconstruction in *wh*-questions. However, this only holds as long as the *wh*-relativization asymmetry is related to a grammatical factor. Once the difference between RCs and questions can be related to a non-syntactic factor, this conclusion can no longer be drawn. We will come back to the interpretation of the *wh*-movement data in the general discussion.

5. Experiment 2: subjects versus objects (S/O)

Experiment 2 was designed to address crucial questions that the experiments in Wierzba et al. (2021), the AP/DP experiment and Experiment 1 left open: what does it mean that coreference between pronoun and embedded R-expression in German *wh*-movement of DPs is neither close to 0% nor to 100%? How can we disentangle the effects of grammatical principles and extra-syntactic factors? To tackle these issues, we compare *wh*-movement of objects with *wh*-movement of subjects. The crucial conditions differ in the base position of the fronted constituent, whereas the linear order and the distance between R-expression and pronoun are identical:

(25) (a) **Lisa** erzählt, [welche Geschichte über **Hanna**] sie _ ärgerlich
           Lisa tells (us) which   story   about Hanna   she   upsetting
           fand.
           found

      (b) **Lisa** erzählt, [welche Geschichte über **Hanna**] _ sie
           Lisa tells (us) which   story   about Hanna   her
           verärgert hat.
           upset   has

In (25a) (object movement), Condition C is violated only under the assumption that there is reconstruction. In (25b) (subject movement), Condition C is not violated, irrespective of whether reconstruction is assumed or not.[15] In all other respects, the sentences are as similar as possible, especially with respect to plausibility, topicality, and linear distance. Note that such near-minimal pairs can only be constructed in a head-final language like German, while in English (as shown by the translations), there would be a difference in the distance between R-expression and pronoun in the two conditions (but see also Note 18). The benefit of this design thus is that the reconstruction hypothesis predicts a difference here (not the absence of a difference, as in the previous designs): coreference between the pronoun and the embedded R-expression would violate Condition C in (25a), but not (25b). Crucially, we then do not have to rely on the problematic interpretation of absolute values (whether the responses are close to 0% or 100%) in this type of design: (25b) will provide us with a baseline that will show us what proportion of positive responses we should expect in the absence of any grammatical violation, purely based on pronoun resolution preferences unrelated to binding. If we find fewer positive responses in (25a) than in (25b), even if they are not at zero, this would lend

---

[15] Crucially, only reconstruction of the A′-movement step is relevant here. Non-pronominal subjects in German can also follow weak object pronouns; in the case at hand, this would lead to a Condition C violation. We follow Müller (1999: 792) in interpreting the variable positions of subjects as resulting from optional A-movement to Spec,TP. Thus, only reconstruction to the higher subject position (Spec,TP), which is above the position of weak pronouns, is relevant in the case at hand. The grammaticality of (25b) shows that the lower subject position, viz., Spec,vP, is not available for reconstruction here (as it would lead to ungrammaticality).

601

itself to an explanation in terms of reconstruction (but see the discussion at the end of this section for why this conclusion may be premature).

The design, hypotheses, and planned analysis of Experiment 2 were pre-registered prior to data collection at https://osf.io/mjgpz.

5.1 *Participants and procedure*

The basic procedure was the same as described above for the AP/DP experiment and Experiment 1. A web-based questionnaire was set up using the platform L-Rex (Starschenko & Wierzba 2020). No acceptability ratings were collected. A total of 32 participants, recruited via prolific.co, were tested and 78 stimuli were presented to each participant (32 critical items, 44 fillers, and 2 items intended for exploratory investigation of an additional research question). For the critical items, 128 data points were collected per condition/question (four from each participant). On average, completing the questionnaire took 24 minutes.

5.2 *Design and materials*

Experiment 2 had a $2 \times 2 \times 2$ design. The eight conditions are illustrated in (26). The first manipulated factor was MOVEMENT (in situ/moved). The second factor was PHRASE: the R-expression was either contained in the subject (in that case, the object was a pronoun) or in the object (in that case, the subject was a pronoun). The third manipulated factor was ARG/ADJ: the R-expression was either contained in an argument PP of the noun or in a PP adjoined to the noun.[16]

(26) *Matrix clause*: **Lisa** erzählt … 'Lisa tells (us) …'
  (a) *Object in situ (argument)*
    … dass ⟨sie⟩ die neue Geschichte über **Hanna** ärgerlich fand.
       that   she   the new story   about Hanna upsetting found
    '… that she found the new story about Hanna upsetting.'
  (b) *Object in situ (adjunct)*
    … dass ⟨sie⟩ die neue Geschichte im Buch über **Hanna** ärgerlich fand.
    '… that she found the new story in the book about Hanna upsetting.'
  (c) *Object moved (argument)*
    … [welche Geschichte über **Hanna**] ⟨sie⟩ _ ärgerlich fand.
    '… which story about Hanna she found upsetting.'

---

[16] We decided to test the argument/adjunct distinction by means of PPs rather than with relative clauses versus complement clauses to nouns to limit the complexity of the items. Moreover, Adger et al. (2017) and Bruening & Al Khalaf (2019) found coreference to be more available with clausal modifiers than with PP modifiers. Thus, if there is a Condition C effect, it is more likely to be diagnosed with PP modifiers. As discussed in Note 13, we classified those PPs as arguments whose preposition was selected by the noun. PP adjuncts involved locative expressions.

   (d)   *Object moved (adjunct)*
       … [welche Geschichte im Buch über **Hanna**] ⌈sie⌉ __ ärgerlich fand.
       '… which story in the book about Hanna she found upsetting.'
   (e)   *Subject in situ (argument)*
       … dass die neue Geschichte über   **Hanna** ⌈sie⌉ verärgert hat.
          that the new story   about Hanna her   upset   has
       '… that the new story about Hanna upset her.'
   (f)   *Subject in situ (adjunct)*
       … dass die neue Geschichte im Buch über **Hanna** ⌈sie⌉ verärgert hat.
       '… that the new story in the book about Hanna upset her.'
   (g)   *Subject moved (argument)*
       … [welche Geschichte über **Hanna**] __ ⌈sie⌉ verärgert hat.
       '… which story about Hanna upset her.'
   (h)   *Subject moved (adjunct)*
       … [welche Geschichte im Buch über **Hanna**] __ ⌈sie⌉ verärgert hat.
       '… which story in the book about Hanna upset her.'

The corresponding questions, Q1 and Q2, were *Kann man den Satz so verstehen, dass Lisa/Hanna eine Geschichte ärgerlich fand?* 'Can this sentence be interpreted such that Lisa/Hanna found a story upsetting?' (object conditions) and *Kann man den Satz so verstehen, dass eine Geschichte Lisa/Hanna verärgert hat?* 'Can this sentence be interpreted such that a story upset Lisa/Hanna?' (subject conditions).

5.3 *Hypotheses and predictions*

All hypotheses below refer to effects on the proportion of 'yes' answers to the question about coreference between the pronoun and the R-expression in the embedded clause (Q2). For the evaluation of hypotheses H1, H2(a), and H2(b), we take into account only the argument conditions. Whether there is a difference between arguments and adjuncts is tested by means of hypothesis H3.

(H1)   Condition C hypothesis: R-expressions cannot be coreferential with a
        c-commanding expression.

Hypothesis H1 is the premise on which the experimental design relies; only if this holds are the results informative in the intended way. It predicts a simple effect of PHRASE in the following direction: there should be more positive responses to Q2 in the 'subject in situ (argument)' condition (in which there is no Condition C violation) than in the 'object in situ (argument)' condition.
   The two crucial hypotheses with respect to reconstruction are:

(H2)   (a)   Reconstruction hypothesis: the base position of moved phrases matters
           for Condition C.
       (b)   Surface hypothesis: the surface position of moved phrases matters for
           Condition C.

Both hypotheses presuppose that H1 holds. If that is the case, then H2(a) predicts that there should also be a simple effect of PHRASE in the condition with movement: there should be more positive responses to Q2 in the 'subject moved (argument)' condition than in the 'object moved (argument)' condition. Crucially, this would be evidence in favor of reconstruction that is not based on the lack of an effect. Hypothesis H2(b) predicts an interaction between MOVEMENT and PHRASE: The difference between 'object moved (argument)' and 'subject moved (argument)' should be smaller than between 'object in situ (argument)' and 'subject in situ (argument)'.

Hypotheses H2(a) and H2(b) are not mutually exclusive. Our design potentially allows us to distinguish between data patterns compatible with exceptionless reconstruction for Condition C (evidence for H2(a), no evidence for H2(b)), fully surface-oriented evaluation of Condition C (no evidence for H2(a), evidence for H2(b)), and patterns in which both the base position and the surface position of the moved phrase play a role (in case we find evidence for both H2(a) and H2(b)).

In addition, we test the argument/adjunct asymmetry hypothesis:

(H3) Argument/adjunct asymmetry hypothesis: in contrast to arguments, there is no reconstruction for adjuncts.

We consider two predictions of H3. First, if the prediction of the reconstruction hypothesis H2(a) is borne out, then there should be a simple interaction between PHRASE and ARG/ADJ within the 'moved' conditions in the following direction: there should be a smaller difference between 'subject moved (argument)' and 'subject moved (adjunct)' than between 'object moved (argument)' and 'object moved (adjunct)'. The reasoning behind this prediction is that if there is no reconstruction for adjuncts, then they should always show a high proportion of positive answers to Q2 in the conditions with movement: i.e. they should be more similar to arguments in the subject movement conditions than in the object movement conditions (where H2(a) predicts less 'yes' answers for arguments). Second, H3 predicts a simple interaction between MOVEMENT and ARG/ADJ within the 'object' condition in the following direction: there should be a larger difference between 'object in situ (adjunct)' and 'object moved (adjunct)' than between 'object in situ (argument)' and 'object moved (argument)'. This is based on the reasoning that the reconstruction hypothesis predicts a similar pattern in the in situ and moved conditions for arguments. If there is no reconstruction for adjuncts, there should be a difference between the in situ and moved conditions.

5.4 *Results*

The results are summarized in Table 7 and illustrated in Figure 2.

Two generalized linear mixed models were fit. The contrast coding was chosen in such a way that it allowed us to test all predictions described above. Thus, in both models, all factors were treatment-coded, with object as the baseline level of PHRASE

Phrase	Arg/adj	Movement	Q1	Q2
object	argument	in situ	97.8%	7.0%
object	argument	moved	94.5%	35.9%
object	adjunct	in situ	96.9%	7.8%
object	adjunct	moved	96.9%	35.9%
subject	argument	in situ	93.0%	65.6%
subject	argument	moved	95.3%	50.8%
subject	adjunct	in situ	96.1%	55.5%
subject	adjunct	moved	96.1%	51.6%

Table 7
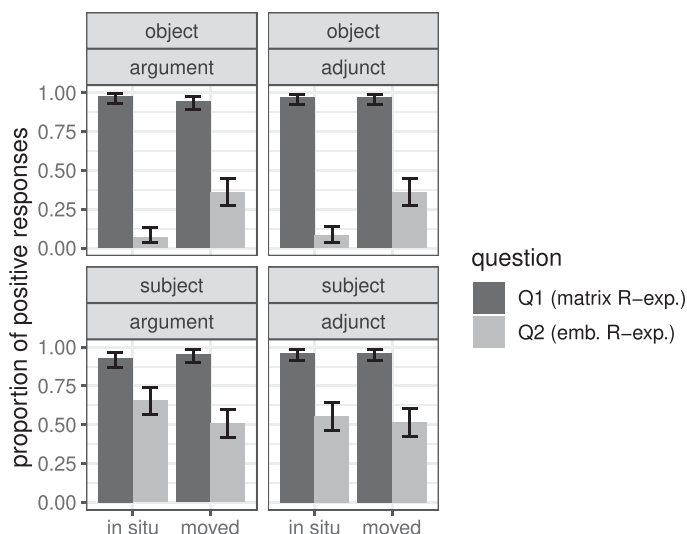Proportion of positive answers to the coreference questions in Experiment 2.

Figure 2
Proportion of positive answers to the coreference questions in Experiment 2. Error bars represent 95% Clopper–Pearson confidence intervals.

and argument as the baseline level of ARG/ADJ. For the factor MOVEMENT, two different kinds of contrast coding were required in order to test all of the predictions. In Model 1, 'in situ' was coded as the baseline level. This means that in the output of this model, PHRASE will represent a simple effect: the difference between 'object' and 'subject' within the levels 'argument' and 'in situ' of the other factors. This will allow for evaluation of the predictions of H1. In Model 2, 'moved' was coded as the baseline level for evaluation of H2(a) and the first prediction of H3, which predict simple effects/interactions within this level. For H2(b) and the second prediction of

605

H3, the contrast coding of the factor MOVEMENT is not relevant, thus, it can be evaluated based on the output of any of the models.

According to a generalized linear mixed model, the prediction of H1 (Condition C hypothesis) was confirmed: a simple effect of PHRASE was found within the levels 'in situ', 'argument' of the other factors ($z = 8.226$, $p < 0.001$ in Model 1). The prediction of H2(a) (reconstruction hypothesis) was confirmed: a simple effect of PHRASE was also found within the levels 'moved' and 'argument' of the other factors ($z = 2.391$, $p = 0.017$ in Model 2). The prediction of H2(b) (surface hypothesis) was confirmed: a simple interaction between MOVEMENT and PHRASE was found within the level 'argument' of the remaining factor ($|z| = 5.596$, $p < 0.001$ in Models 1/2). Neither of the predictions of H3 (argument/adjunct asymmetry hypothesis) was confirmed: there was no significant simple interaction between PHRASE and ARG/ADJ within the level 'moved' ($z = 0.087$, $p = 0.931$ in Model 2), nor a significant simple interaction between MOVEMENT and ARG/ADJ within the level 'object' ($|z| = 0.209$, $p = 0.834$ in Models 1/2). The full results of the models are shown in Tables 8 and 9.

	Estimate	Std. Error	z-value	p-value
phrase	3.2336	0.3931	8.226	< 2e-16
movement	2.0063	0.3922	5.115	3.13e-07
arg/adj	0.1139	0.4782	0.238	0.812
phrase*movement	−2.623	0.4688	−5.596	2.20e-08
phrase*arg/adj	−0.5417	0.5431	−0.998	0.319
movement*arg/adj	−0.1139	0.5446	−0.209	0.834
phrase*movement*arg/adj	0.5731	0.6522	0.879	0.380

Table 8
Summary of fixed effects in the output of Model 1 (with 'in situ' as the baseline level of MOVEMENT)
for Experiment 2.

	Estimate	Std. Error	z-value	p-value
phrase	0.6106	0.2554	2.391	0.01679
movement	−2.006	0.3922	−5.115	3.13e-07
arg/adj	1.045e-06	0.2606	0.000	1.00000
phrase*movement	2.623	0.4688	5.596	2.20e-08
phrase*arg/adj	0.03133	0.3612	0.087	0.93087
movement*arg/adj	0.1139	0.5446	0.209	0.83439
phrase*movement*arg/adj	−0.5731	0.6522	−0.879	0.37960

Table 9
Summary of fixed effects in the output of Model 2 (with 'moved' as the baseline level of MOVEMENT)
for Experiment 2.

5.5 *Discussion*

In Experiment 2, evidence for both the reconstruction hypothesis and the surface hypothesis was found: the finding that the 'moved subject' and 'moved object' conditions differ in spite of their similarity at the surface supports the view that the base position of the moved phrase plays a role. However, the difference between the 'moved object' and 'in situ object' conditions shows that the surface position matters as well. No evidence for an argument/adjunct asymmetry was found.

How can the finding that the base position plays a role (pointing toward reconstruction) be reconciled with the finding that the surface position also matters (speaking against reconstruction)? In other words, how can we interpret intermediate response patterns that neither correspond to the clear 0/100 divide that we would expect if coreference were fully determined by binding principles and reconstruction, nor to the complete absence of a difference between 'moved subject' and 'moved object' expected if reconstruction did not play a role at all? There are two basic possibilities: First, there is reconstruction but other factors lead to a higher availability of coreference than expected. Second, there is no reconstruction and, despite the fact that the subject/object conditions are near-minimal pairs, there are additional factors causing the asymmetry.

We will first discuss an interpretation in terms of reconstruction. In Section 2.3, we considered three possible scenarios in which the *wh*-phrase is reconstructed, but coreference in the moved condition is still available to some extent. The first two options had to do with a potential by-subject or by-item split: there might be a group of participants that reconstructs and another that does not (variation between dialects or idiolects); or a fixed group of items which reconstructs and another which does not (not necessarily related to our categorization as an argument/adjunct – otherwise we should have seen a difference between our conditions in this respect – but to some inherent property of the items). However, the idea of a by-subject or by-item split was not supported by post hoc analyses. We take the difference between the positive responses to Q2 in the 'moved subject' and 'moved object' conditions to be the main indicator of reconstruction. By-subject and by-item analyses of this measure revealed a gradient and unimodal distribution rather than a split between subgroups of speakers or items. Thus, whether coreference is available or not in the moved condition does not seem to vary systematically, based on idiolects or a specific property of the items, but it rather seems to vary individually from case to case.

Another possibility that we considered in Section 2.3 was that even if there is reconstruction (i.e. even if the PP modifier is present in the bottom copy), non-syntactic factors could still influence participants' judgments in this type of experimental task and lead them to respond with 'yes' in spite of a Condition C violation. In order to explain the pattern that we see, these would need to be factors that are likely to raise the availability of coreference in the 'object moved' more than in the 'object in situ' condition. One such factor could be closeness: coreference with the embedded R-expression could be judged to be possible because it is a very close

antecedent in the moved condition. Another factor could be linear order: coreference with the embedded R-expression would yield a (usually preferred) anaphoric instead of a cataphoric relation in the moved condition, potentially raising the number of positive responses. The degree to which participants' judgments are affected by such factors could plausibly vary on an individual basis, accounting for the gradience. This suggests treating Condition C as a soft factor, a point we return to in the general discussion. The subject–object asymmetry then follows under the assumption that the violation of a soft constraint still comes at a price, viz., reduces the availability of coreference.

Based on our data, one cannot assess conclusively how much these potential factors contribute to the patterns and how the various factors interact exactly. Thus, while there may be plausible explanations for why different response patterns to 'object in situ' and 'object moved' could emerge even if there is reconstruction, we cannot rule out that the asymmetry is due to non-reconstruction. This in turn leads us to the second possible interpretation of the pattern in this experiment, namely that the difference between 'object moved' and 'subject moved' is not explained in terms of reconstruction but by means of non-syntactic factors.

The anonymous reviewers suggested the following two possible alternative explanations of the contrast. First, it has been observed in pronoun resolution that parallelism between the function of the antecedent and the pronoun increases the likelihood of coreference (e.g. Stevenson, Nelson & Stenning 1995). Thus, a subject pronoun prefers a subject antecedent, while an object pronoun prefers an object antecedent. This could indeed have an effect in the design used in Experiment 2: in the object moved condition, the pronoun is a subject. Consequently, it could be more attracted to the matrix subject; the lower availability of coreference with the embedded R-expression could thus be related to this factor rather than reconstruction (while in the subject moved condition, the pronoun is an object, which would not be equally attracted to the matrix subject).[17] The second alternative capitalizes on the differential anaphoric availability of grammatical functions: in general, subjects tend to be more prominent antecedents than objects (recall from Section 2.3). This could affect coreference in Experiment 2 as follows: in the object moved condition, the matrix subject is far more salient than the R-expression within the *wh*-moved object, while in the subject moved condition, the asymmetry is not as substantial since both phrases bear the subject function.

These two explanations of the subject–object contrast based on non-syntactic factors are indeed obvious alternatives, which we address in the next experiment. Note that the role of non-syntactic factors is certainly visible in one part of the data. Although the examples with the subject in situ are uncontroversially grammatical, coreference with the lower R-expression is only available at 56–66% and thus substantially lower then the 100% one might a priori expect. This is clearly related

---

[17] However, the pronoun in the subject condition is an experiencer and thus may have different anaphoric properties than an object pronoun bearing the semantic role of theme/patient.

to the fact that the matrix subject is a more salient antecedent than an R-expression within a *wh*-phrase.[18]

6. EXPERIMENT 3: NON-SYNTACTIC FACTORS (NON-SYN)

Given the alternative explanations of the subject–object contrast detected in Experiment 2 that we discussed at the end of the last section, we designed another experiment to be able to tease apart the syntactic explanation (based on reconstruction) from the non-syntactic one (based on preferences in pronoun resolution). While we adopted the basic design of the previous experiment, to test both alternative explanations (parallel function and higher salience of subjects), certain modifications were necessary. We added a context sentence before the indirect question in which we introduced two R-expressions one of which would be repeated as the R-expression within the *wh*-phrase. In addition, we varied the grammatical function of the two R-expressions in the context sentence (subject vs. object). The context sentence is followed by the matrix clause that introduces the indirect question. Unlike in the previous experiment, no referent is introduced in the matrix clause; rather, an impersonal construction is used. The basic structure of an item would thus be as follows (in (27) with a *wh*-object):

(27) X/Y nahm   Y/X zu einer Party   mit. Es wurde unter   anderem
      X   took   Y   to a   party   with. It was   among   others
      darüber gesprochen, [welchen Bericht über X] er erstaunlich fand.
      about.it spoken   which   report on   X   he surprising found
      'X/Y took Y/X to a party. It was discussed which report on X he found
      surprising.'

---

[18] Stockwell et al. (2021) present an experiment that is based on a similar subject–object asymmetry (recall from note 11). They compare *wh*-objects as in *Which picture of Harry did he frame?* with *wh*-subjects in causative constructions as in *Which picture of Harry made him laugh?*. They find that coreference receives higher acceptability with *wh*-subjects than with *wh*-objects and that the acceptability of a different referent is much higher with *wh*-objects than with *wh*-subjects, pointing toward a Condition C effect. In both cases, coreference with the R-expression is increased by distance. Since in this experiment there is no additional R-expression in the context, coreference with the R-expression inside the *wh*-phrase cannot be reduced because of the higher salience of another R-expression (and since matrix questions are used, the distance between *wh*-phrase and pronoun is the same in both conditions). However, it cannot be ruled out that the different anaphoric preferences of subject and object pronouns can have an effect here (e.g. that the subject pronoun has a stronger preference to corefer with the topic than the object pronoun and therefore, since there is no topic in the sentence, a different referent is chosen with *wh*-objects but not necessarily with *wh*-subjects). Furthermore, it is remarkable that unquestionably grammatical conditions received surprisingly low ratings (4 out of 7 with local movement of *wh*-subjects for both the R-expression and a different referent and 2.7 for the other referent in long movement with *wh*-subjects), raising questions about potential task-related problems. It should also be mentioned that the subject/object conditions are not as similar as in our experiment in that the theta role of the *wh*-phrase (agent/causer vs. theme) and the position of the lexical verb differ. Thus, in our view, as in our experiment, one cannot rule out that other factors may be responsible for the asymmetry.

There are four conditions: The *wh*-phrase is either a subject or an object and the grammatical function of the R-expression in the context sentence that will be taken up within the *wh*-phrase is either subject or object. In what follows, R2 is the R-expression contained in the *wh*-phrase, R1 is the other R-expression that only occurs in the context sentence.

Under the parallel function hypothesis, the expectation is that coreference with the R-expression inside the *wh*-phrase (R2) will be higher if the referent of this R-expression is introduced (in the context sentence) with the same grammatical function as the pronoun. Thus, the availability of coreference with the lower R-expression (R2) does not depend on the grammatical function of the *wh*-phrase; rather, it is (indirectly) affected by the relationship between R2's grammatical function in the context sentence and the grammatical function of the pronoun. Under the subject prominence hypothesis, the expectation is that coreference with the R-expression within the *wh*-phrase (R2) is higher if R2 is introduced as a subject in the context sentence. Again, the availability of coreference with R2 is indirectly affected, namely by its grammatical function in the context sentence.

6.1 *Participants and procedure*

The procedure was the same as in Experiment 2; 32 participants were tested and 76 stimuli were presented to each participant (32 critical items, 44 fillers). For the critical items, 256 data points were collected per condition/question (8 from each participant). On average, completing the questionnaire took 26 minutes.

6.2 *Design and materials*

Experiment 3 had a 2×2 design. The four conditions are illustrated in (28). We partially followed the design of Experiment 2: we adopted the factor PHRASE (R-expression contained in subject/object). We dropped the factors MOVEMENT and ARGUMENT/ADJUNCT: R2 was always contained in a PP-argument of a *wh*-moved DP. FUNCTION IN CONTEXT was included as an additional factor: As mentioned above, we added a context sentence before the indirect question in which we varied which of the two proper names was introduced as a subject and which as an object.

(28) (a) *Wh-phrase = object, R2 introduced as object in the context*
      Kerstin hat Ilse zu einer Party mitgenommen. Es wurde unter
      Kerstin has Ilse to a   party   taken   it was   among
      anderem   darüber gesprochen, welches Geschenk für Ilse sie
      others   about.it discussed   which   present for Ilse she
      entzückend fand.
      lovely   found
      'Kerstin took Ilse to a party. It was discussed which present for Ilse she found lovely.'

(b) *Wh-phrase = object, R2 introduced as a subject in the context*
Ilse hat Kerstin zu einer Party mitgenommen. Es wurde unter anderem darüber gesprochen, welches Geschenk für Ilse sie entzückend fand.
'Ilse took Kerstin to a party. It was discussed which present for Ilse she found lovely.'

(c) *Wh-phrase = subject, R2 introduced as an object in the context*
Kerstin hat Ilse zu einer Party mitgenommen. Es wurde unter
Kerstin has Ilse to a   party taken   it was   among
anderem   daräber gesprochen, welches Geschenk für Ilse sie
others   about.it discussed   which   present for Ilse her
entzückt hat.
enraptured has
'Kerstin took Ilse to a party. It was discussed which present for Ilse enraptured her.'

(d) *Wh-phrase = subject, R2 introduced as a subject in the context*
Ilse hat Kerstin zu einer Party mitgenommen. Es wurde unter anderem darüber gesprochen, welches Geschenk für Ilse sie entzückt hat.
'Ilse took Kerstin to a party. It was discussed which present for Ilse enraptured her.'

The corresponding questions Q1 (coreference with R1) and Q2 (coreference with R2) were (object *wh*) *Kann man den Satz so verstehen, dass Kerstin ein Geschenk entzückend fand/Ilse ein Geschenk entzückend fand?* 'Can this sentence be understood such that Kerstin found a present enrapturing/Ilse found a present enrapturing' and (subject *wh*) *Kann man den Satz so verstehen, dass Ilse ein Geschenk entzückte/ Kerstin ein Geschenk entzückte?* 'Can this sentence be understood such that a present enraptured Ilse/a present enraptured Kerstin?'

The items were based on the indirect questions in Experiment 2, we only added a context sentence and an impersonal matrix clause. Context sentence and indirect question were identical in all items except that we varied the object of the preposition: 'party', 'meeting', 'celebration', 'festivity'. The fillers were identical to those in Experiment 2, except that we also added a context sentence containing two R-expressions to make the task similar to that of the items.

6.3 *Hypotheses and predictions*

(H1)   Reconstruction: the base position matters for Principle C.

H1 predicts a main effect of PHRASE in the direction of more 'yes' responses to Q2 in [subject *wh*-phrase] than in [object *wh*-phrase].

(H2)   Parallel function: a pronoun is more likely to be interpreted as coreferent with an R-expression if they have the same grammatical function (e.g. both are a subject or both are an object).

611

H2 predicts an interaction between PHRASE and FUNCTION IN CONTEXT. The interaction should be toward more 'yes' responses to Q2 in [object *wh*-phrase, R-exp. in *wh*-phrase introduced as subject] and [subject *wh*-phrase, R-exp. in *wh*-phrase introduced as object] than in [object *wh*-phrase, R-exp. in *wh*-phrase introduced as object] and [subject *wh*-phrase, R-exp. in *wh*-phrase introduced as subject].

(H3)   Subject prominence: subjects are more available as antecedents than elements with other grammatical functions.

H3 predicts a main effect of FUNCTION IN CONTEXT in the direction of more 'yes' responses to Q2 in [R-exp. in *wh*-phrase introduced as subject] than in [R-exp. in *wh*-phrase introduced as object].

6.4 *Results*

The results are summarized in Table 10 and illustrated in Figure 3. Generalized linear mixed models were fit. The contrast coding allowed us to test the predictions described above: both factors were sum-coded. The full model results are shown in

Phrase	Function in context	Q1	Q2
object	object	98.8%	44.5%
object	subject	96.5%	45.7%
subject	object	95.7%	50.8%
subject	subject	94.1%	51.2%

Table 10
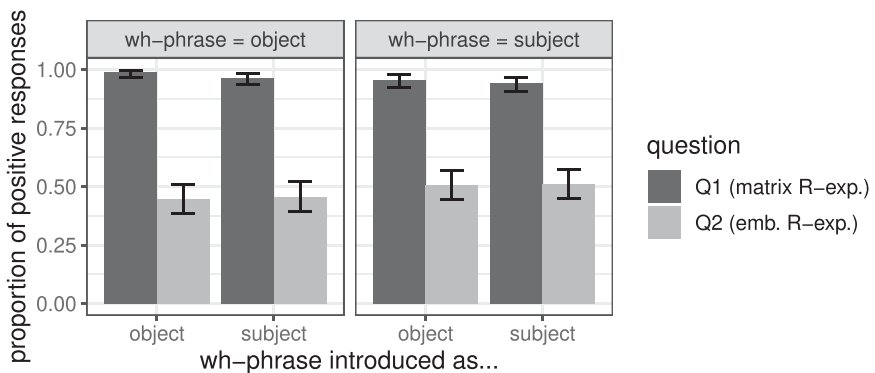Proportion of positive answers to the coreference questions in Experiment 3.

Figure 3
Proportion of positive answers to the coreference questions in Experiment 3. Error bars represent 95% Clopper–Pearson confidence intervals.

612

	Estimate	Std. Error	z-value	p-value
phrase	−0.27226	0.09546	−2.852	0.00434
function in context	−0.03679	0.09404	−0.391	0.69561
phrase*function in context	−0.01989	0.09432	−0.211	0.83298

Table 11
Summary of fixed effects in the output of the model for Experiment 3.

Table 11. The predictions of H1 (reconstruction hypothesis) were confirmed: there was a significant main effect of PHRASE in the predicted direction. The predictions of H2 (parallel function) and H3 (subject prominence) were not confirmed: there was neither a significant main effect of FUNCTION IN CONTEXT nor a significant interaction between the two factors.

6.5 *Discussion*

We did not find evidence for the hypotheses H2 and H3: the availability of coreference with R2 was not significantly affected by the grammatical function it bears in the context sentence. As in Experiment 2, there is a contrast between *wh*-subjects and *wh*-objects: Coreference with R2 is more available with *wh*-subjects than with *wh*-objects, as predicted by H1 (reconstruction hypothesis).

These findings speak against the view that the subject/object contrast that we observed in Experiment 2 can be fully reduced to pronoun resolution preferences: a residual contrast remains even if these factors are controlled for. However, numerically, the subject/object contrast is smaller than in the previous experiment. Compared to Experiment 2, where the difference between *wh*-subjects and *wh*-objects was between 36% and 51%, in Experiment 3, the difference was smaller (45 vs. 51%) which, given the results we have obtained so far in our experiments, is probably close to the maximum that one can get for Q2 in this experimental design. This difference between Experiments 2 and 3 (with the caveat that this is only a tentative post hoc observation across two separate experiments) would be compatible with the view that while none of the specific parallelism/salience-based hypotheses that we tested was confirmed, there might indeed be another important factor facilitating coreference here, namely the accessibility of the referent. If it has already been introduced in the previous discourse as in Experiment 3, coreference with R2 seems to become more available. This observation recalls the discussion in Bruening & Al Khalaf (2019: 268–269), who argue for an improvement in a similar configuration (without, however, investigating this experimentally). There remains an asymmetry, though, in that coreference with R2 with *wh*-subjects seems to be unaffected by this. The values for Q2 do not differ between Experiments 2 and 3 (51% for moved arguments in both experiments).

As for the consequences for syntactic theory, the results of Experiment 3 challenge the view that the main factors affecting coreference are reconstruction and a

613

hard Condition C constraint. This would imply a categorical split into grammatical (*wh*-subjects) and ungrammatical (*wh*-objects), which seems inadequate in view of the small size of the subject/object contrast that remains once the influence of potential confounds is reduced. Nevertheless, the contrast does not vanish completely, which also needs to be accounted for – we will discuss possible explanations in the next section. From a methodological point of view, Experiment 3 stresses the importance of the context sentences used in the materials and of the development of experimental designs that allow us to also detect subtle contrasts.

7. GENERAL DISCUSSION AND CONCLUSION

Given the results of Experiments 2 and 3, reconstruction and a hard Condition C constraint cannot be considered the main factors governing coreference in German A′-movement. We will now discuss the implications for DP movement in Wierzba et al. (2021), the AP/DP experiment and Experiment 1, differences between Wierzba et al. (2021), the AP/DP experiment and Experiment 1 versus Experiments 2 and 3, the AP/DP and *wh*-relativization contrast, possible differences between German and English, and the status of the residual subject/object asymmetry found in Experiment 3.

First, the data with *wh*-subjects in Experiments 2 and 3 support our arguments against the threshold-based logic of previous work (including our own) according to which values below a certain threshold close to zero indicate reconstruction, while values above it indicate non-reconstruction. In subject *wh*-movement where Condition C is not at stake, coreference with R2, which is fully grammatical, remains around 50%. This is clearly related to a non-syntactic factor, namely the preference for coreference with the matrix subject. It is plausible that this also contributes to the low coreference values for R2 that we obtained in Wierzba et al. (2021), the AP/DP experiment, and Experiment 1 with *wh*-movement of DP-*objects*. Together with the results of Experiment 3, which clearly argue against reconstruction and a hard Condition C constraint as the main factors governing coreference in A′-movement, a plausible interpretation of the DP movement data in Wierzba et al. (2021), the AP/DP experiment, and Experiment 1 is that the low values for Q2 do not result from reconstruction and thus a classical Condition C violation either (our conclusion is thus eventually similar to Bruening & Al Khalaf 2019, although their reasoning is based on a different non-syntactic factor).

Second, it is notable that the proportion of positive responses to Q2 in the 'moved object' conditions was overall higher in Experiments 2 and 3 (36–46%) than in experiments 2 and 4 of Wierzba et al. (2021), the AP/DP experiment, and Experiment 1 (7%, 11%, 21%, and 13%, respectively, in the short conditions). We think that this may be due to changes in the materials: in order to construct semantically similar object and subject variants, we mainly chose predicates expressing emotion or evaluation, for which pairs of the type 'X found Y upsetting' and 'Y upset X' could be constructed. It is possible that this change made the items more uniform and overall increased the plausibility of the coreferential reading. In experiments

2 and 4 of Wierzba et al. (2021), the AP/DP experiment, and Experiment 1, all kinds of transitive verbs were used, which perhaps introduced a larger amount of variability in plausibility; it might be generally quite easy to imagine having some emotion or attitude toward a book/report/rumor/etc. about oneself, whereas it might vary more whether performing a specific action to it (reading/hiding/falsifying/etc.) is perceived as likely. These substantial differences also provide another argument against threshold-based reasoning as they show how much the values can vary between experiments based on properties of the materials, even though the basic grammatical configuration is the same (object *wh*-movement).

Third, what remains open at this point is what the reasoning so far implies for the AP/DP-contrast in Wierzba et al. (2021), the AP/DP experiment, and the *wh*-/relativization contrast in Experiment 1. Starting with the AP/DP-contrast, while the facts are compatible with a theory where only APs/predicates reconstruct (see, e.g. Adger et al. 2017; Bruening & Al Khalaf 2019), we cannot rule out that the factors that increased coreference with R2 for object *wh*-movement in Experiments 2 and 3 could also lead to a significant improvement with AP-movement. Thus, perhaps, coreference with APs also becomes more available with experiencer predicates, e.g. in a sentence like *Peter erzählt, wie ungerecht gegen Hans ihm das Urteil erscheint* 'Peter tells us how unjust against John the judgment seems to him'. Furthermore, it may be possible to further increase the availability of coreference by introducing the R-expression inside the AP in the previous context as we did for DPs in Experiment 3. The (remaining) AP/DP-asymmetry could then perhaps be related to non-syntactic factors as well (e.g. a difference in plausibility). At this point, any conclusions about AP-reconstruction and the AP/DP-contrast strike us as premature, and we intend to investigate the different possibilities in future work. As for the *wh*-relativization contrast in Experiment 1, if *wh*-movement of DPs does not involve a typical Condition C violation, just like relativization, a non-syntactic explanation is required. In Section 4 we already hinted at such a possibility: a difference in referential accessibility. The difference between Experiments 2 and 3 can be interpreted such that coreference with R2 is more available if the referent is established in the prior discourse and thus more accessible. Importantly, relative clauses are sometimes analyzed as involving topicalization of some sort. For instance, in cartographic work, relative pronouns have been argued to occupy a topic position (see Bianchi 1999). Since the pronoun refers back to the head of the relative clause, this could make it and the R-expression contained in it referentially more accessible and thus facilitate coreference. It is not clear at this point, though, whether this is sufficient to account for the asymmetry. Note that the external head in our experiments is headed by a universal quantifier, which does not make an ideal topic. Conversely, the *wh*-phrase in our experiments is headed by 'which', which is normally associated with D-linking. Thus, it is not clear whether the referential asymmetry is large enough to account for the difference in coreference (it would arguably have to be the relative pronoun that makes the difference). Given these limitations, we leave an exploration of this hypothesis for the *wh*-relativization asymmetry for future work.

The results of the experiments in Wierzba et al. (2021), the AP/DP experiment, and Experiment 1 deviated from what had been reported for English in that we found coreference with DPs to be much less available. Initially, we investigated whether the differences with regard to coreference could be related to differences in the method or in the materials. However, as described in the Appendix (A.1), an additional experiment based on the materials of the AP/DP experiment but using the forced-choice method of Bruening & Al Khalaf (2019) did not change the results as the availability of coreference remained very low. We also replicated the second experiment of Bruening & Al Khalaf (2019) by translating their items into German (and using their forced-choice method) and still found a substantial difference in the availability of coreference, see the Appendix (A.2). This puzzling asymmetry has changed with Experiments 2 and 3 where the values for *wh*-objects are closer to the English facts and which invite the same conclusion, namely that reconstruction and, consequently, a hard Condition C constraint are not the main factors affecting coreference in A′-movement. Thus, the picture has become less clear-cut with the evidence for reconstruction in German fading and new evidence for reconstruction in English being introduced into the discussion (as in Stockwell et al. 2021). Given that crosslinguistic differences in this area are a priori unexpected, a thorough crosslinguistic comparison of Condition C reconstruction remains an important topic for future research. We should add that our results converge with previous work on English with regard to the lack of an argument–adjunct asymmetry (but see Stockwell, Meltzer-Asscher & Sportiche, to appear, for different results).

The final aspect we need to address is the residual subject–object asymmetry in Experiments 2 and, especially, 3. While the coreference values for subject and object *wh*-movement become very similar in Experiment 3 and a classification into grammatical (*wh*-subjects) and ungrammatical (*wh*-objects) seems inadequate, there remains a difference: (i) the small numerical difference is significant in Experiment 3 and (ii) only objects are affected by the design change in Experiment 3 where the R-expression within the *wh*-phrase is introduced in the previous discourse, while *wh*-subjects show the same coreference values in Experiments 2 and 3. We can think of two principled possibilities, (i) another non-syntactic factor or (ii) a possibly soft/violable syntactic factor. One possible non-syntactic factor could be the grammatical relation of the *wh*-phrase that contains the antecedent. Given that subjects are more salient antecedents than objects, if the grammatical function of an XP also affects the accessibility of R-expressions contained in that XP, we expect R-expressions inside subjects to be more accessible than R-expressions inside objects. While not implausible, we are not aware of any independent work supporting this assumption. In addition, the predictions are the same as any syntactic account that relies on the subject–object asymmetry, which is why they cannot easily be teased apart in the case at hand.[19]

---

[19] Note also that our items are based on experiencer predicates with the *wh*-subject bearing the theme/stimulus function. It is thus not fully obvious whether these subjects are equally salient as external/agent arguments of regular transitive verbs.

The alternative to a non-syntactic factor is a soft/violable grammatical factor. It has long been known that Principle C is sometimes violable. Recent work on backward anaphora by Gor & Syrett (2019) and Gor (2020) has shown that while a Condition C-configuration leads to an expectation of obviation (see Safir 2004), it can be overridden under certain conditions, including various pragmatic conditions and especially under high plausibility of coreference. Thus, a violable Condition C constraint could interact with other constraints (plausibility, referential accessibility) and cause the weak subject–object asymmetry. Given the variable results we have obtained for *wh*-objects, it is conceivable that the Condition C effect is visible to various extents, depending on the strength of the other factors. Note that this view requires reconstruction (viz., presence of PP modifiers in the bottom copy) after all and thus something else must be said to account for the in situ/moved contrast. One possibility is that in the moved condition, the initial parse (before reconstruction) would involve forward anaphora rather than backward anaphora. Since forward anaphora is preferred over backward anaphora, this could facilitate coreference.[20] Our data do not abjudicate between these two views. What they show quite clearly is that a theory which includes both reconstruction and a hard Condition C constraint fails. They are compatible with either a theory without reconstruction (viz., no PP modifiers in the bottom copy), but with a possibly strong Condition C constraint (supplemented by some additional constraint to account for the subject–object asymmetry) or a theory that includes reconstruction (viz., PP modifiers present in the bottom copy) and a soft Condition C constraint (and some additional factor to account for the object moved/object in situ asymmetry).

We will conclude by stressing an important methodological point: we have criticized previous work for relying on differences in coreference between moved/in situ. We do believe that as long as only object *wh*-movement is investigated, the problems we have identified remain serious: the two structures differ in other respects that may affect coreference judgments (forward/backward anaphora, linear distance). Our design in Experiments 2 and 3 is crucially different in that we

---

According to another alternative suggested by one of the anonymous reviewers, the subject/object difference could arise because there is a competitor in the object condition, namely one that uses a reflexive instead of the R-expression (e.g. *which report on himself John finds interesting*). If speakers somehow have access to this alternative, this may decrease accessibility of coreference with the R-expression in the object condition given that there are better ways of expressing the same meaning. We agree that the role of competing structures is an interesting point that is worth being explored further. However, as far as we can tell, the logic of this argument only goes through if there is obligatory reconstruction. If instead there is no obligatory reconstruction, no grammatical violation obtains in the object condition and thus no pressure to use a different construction should arise. If, however, there is a general preference to use a reflexive whenever possible, then this preference should be equally visible in the subject condition, where a reflexive is a possibility as well, given that the subject can reconstruct below the experiencer; it could then no longer be used to motivate the subject–object asymmetry.

[20] Potential indirect support for this view could come from an experiment on reconstruction of extraposition in Gor (2020: ch. 3). She observes that extraposition has no effect on coreference/Condition C (the sentences are just as good/bad as without extraposition), suggesting that it reconstructs. Since the linear order between pronoun and name is the same in both conditions, no other obvious factors interact with Principle C, unlike in *wh*-questions.

don't have to rely exclusively on the difference between moved and in situ in the object conditions. Rather, we can compare the values for object moved with two reference points: object in situ, which provides the baseline for ungrammatical structures, and the subject conditions, which provide the baseline for grammatical structures. We can then determine whether the values in 'object moved' are closer to the ungrammatical baseline or the grammatical baseline and draw conclusions based on that rather than having to rely on thresholds of absolute values. Importantly, the comparison moved/in situ remains important in our design; without it, we would only obtain a subject–object contrast and crucially could not conclude that Condition C is best viewed as a soft constraint.

REFERENCES

Adger, David, Alex Drummond, David Hall & Coppe van Urk. 2017. Is there Condition C reconstruction? In Andrew Lamont & Katerina Tetzloff (eds.), *Nels 47: Proceedings of the forty-seventh annual meeting of the North East Linguistic Society*, vol. 1, 21–30. Amherst: GLSA.

Bader, Markus & Gert Webelhuth. 2020. Experiments on coreference in German and C-command. Paper presented at Linguistic Evidence 2020, University of Tübingen.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015a. Parsimonious mixed models, arXiv:1506.04967 [stat.ME].

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015b. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1, 1–48.

Bianchi, Valentina. 1999. *Consequences of antisymmetry: Headed relative clauses*. Berlin: Mouton de Gruyter.

Brennan, Susan E. 1995. Centering attention in discourse. *Language and Cognitive Processes* 10, 137–167.

Bruening, Benjamin & Eman Al Khalaf. 2019. No argument–adjunct asymmetry in reconstruction for Binding Condition C. *Journal of Linguistics* 55.2, 247–276.

Citko, Barbara. 2001. Deletion under identity in relative clauses. *Proceedings of the North Eastern Linguistic Society* (NELS) 31, 131–145.

Cowles, H. Wind, Matthew Walenski & Robert Kluender. 2007. Linguistic and cognitive prominence in anaphor resolution: Topic, contrastive focus and pronouns. *Topoi* 26, 3–18.

Crain, Stephen & Cecile McKee. 1985. The acquisition of structural restrictions on anaphora. *Proceedings of the North Eastern Linguistic Society* (NELS) 16, 94–110.

Cunnings, Ian, Clare Patterson & Claudia Felser. 2014. Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language* 71.1, 39–56.

Donati, Caterina & Carlo Cecchetto. 2011. Relabeling heads: A unified account for relativization structures. *Linguistic Inquiry* 42.4, 519–560.

Fischer, Silke. 2002. Reanalyzing reconstruction effects: An optimality-theoretic account of the relation between pronouns and R-expressions. In Marjo van Koppen, Erica Thrift, Erik Jan van der Torre & Malte Zimmermann (eds.), *Proceedings of Console* IX, 68–81. Leiden: SOLE.

Fischer, Silke. 2004. *Towards an optimal theory of reflexivization*. Ph.D. dissertation, University of Tübingen.

Frey, Werner. 1993. *Syntaktische Bedingungen für die semantische Interpretation: Über Bindung, implizite Argumente und Skopus*. Berlin: Akademie.

Gor, Vera. 2020. *Experimental investigations of Principle C at the syntax-pragmatics interface*. Ph.D. dissertation, Rutgers University.

Gor, Vera & Kristen Syrett. 2019. Beyond Principle C: (Not)-at-issueness and plausibility influence acceptability of coconstrual. In Eszter Ronai, Laura Stigliano & Yenan Sun (eds.), *Proceedings of the 54th Annual Meeting of the Chicago Linguistic Society*, 109–124. Chicago: Chicago Linguistic Society.

Gordon, Peter C. & Randall Hendrick. 1997. Intuitive knowledge of linguistic co-reference. *Cognition* 62.10, 325–370.

Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12, 175–204.

Heycock, Caroline. 1995. Asymmetries in reconstruction. *Linguistic Inquiry* 26.4, 547–570.

Huang, C. T. James. 1993. Reconstruction and the structure of VP: Some theoretical consequences. *Linguistic Inquiry* 24.1, 103–138.

Kaiser, Elsi. 2011. Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and cognitive processes* 26.10, 1625–1666.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82.13, 1–26.

Lasnik, Howard. 1998. Some reconstruction riddles. *University of Pennsylvania Working Papers in Linguistics* 5.1, 83–98.

Lebeaux, David. 1991. Relative clauses, licensing, and the nature of the derivation. In Susan Rothstein (ed.), *Syntax and semantics*. Vol. 25: *Perspectives on phrase structure: Heads and licensing*, 209–239. New York: Academic Press.

Leiner, Dominik. 2018. SoSci Survey [software]. www.soscisurvey.com (accessed 13 May 2022).

Müller, Gereon. 1999. Optimality, markedness, and word order in German. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 37.5, 777–818.

Patterson, Clare & Claudia Felser. 2019. Delayed application of binding Condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research* 48.2, 453–475.

R Core Team. 2016. The R project for statistical computing. www.R-project.org (accessed 13 May 2022).

Safir, Ken. 1999. Vehicle change and reconstruction in A′-chains. *Linguistic Inquiry* 30.4, 587–620.

Safir, Ken. 2004. *The syntax of anaphora*. New York: Oxford University Press.

Salzmann, Martin. 2017. *Reconstruction and resumption in indirect A′-dependencies: On the syntax of prolepsis and relativization in (Swiss) German and beyond*. Berlin: De Gruyter Mouton.

Sauerland, Uli. 2003. Unpronounced heads in relative clauses. In Kerstin Schwabe & Susanne Winkler (eds.), *The interfaces: Deriving and interpreting omitted structures*, 205–226. Amsterdam: John Benjamins.

Schumacher, Petra B., Manuel Dangl & Elyesa Uzun. 2016. Thematic role as prominence cue during pronoun resolution in German. In Anke Holler & Katja Suckow (eds.), *Empirical perspectives on anaphora resolution*, 213–240. Berlin: De Gruyter.

Starschenko, Alexej & Marta Wierzba. 2020. L-Rex: Linguistic rating experiments [software]. GNU General Public License v3.0. https://github.com/2e2a/l-rex (accessed 13 May 2022).

Stevenson, Rosemary J., Alexander W. R. Nelson & Keith Stenning. 1995. The role of parallelism in strategies of pronoun comprehension. *Language and Speech* 38.4, 393–418.

Stockwell, Richard, Aya Meltzer-Asscher & Dominique Sportiche. 2021. There is reconstruction for Condition C in English questions. In Alessa Farinelly & Angelica Hill (eds.), *Proceedings of the fifty-first annual meeting of the North East Linguistic Society* (Nels 51), vol. 2, 205–214. Amherst, MA: GLSA.

Stockwell, Richard, Aya Meltzer-Asscher & Dominique Sportiche. To appear. Experimental evidence for the Condition C argument–adjunct asymmetry in English questions. In *Proceedings of the 52nd Annual Meeting of the North East Linguistic Society*, Amherst, MA: GLSA.

Wierzba, Marta, Martin Salzmann & Doreen Georgi. 2021. An experimental investigation of reconstruction for Condition C in German A′-movement. In *Proceedings of the 56th Annual Meeting of the Chicago Linguistic Society (2020)*, 479–491. Chicago: Chicago Linguistic Society.

Yoshida, Masaya, David Potter & Tim Hunter. 2019. Condition C reconstruction, clausal ellipsis and island repair. *Natural Language & Linguistic Theory* 37.4, 1515–1544.

Appendix

A. Further methodological aspects

Given the discrepancies between Wierzba et al. (2021) and the results reported for English in Adger et al. (2017) and Bruening & Al Khalaf (2019), we conducted an

experiment testing to what extent the differences between the languages could be related to the method or materials. It was run parallel to the AP/DP experiment, using the same materials, but adopting Bruening & Al Khalaf's (2019) forced-choice method. The experiment not only contained a replication of the AP/DP experiment, but also a replication of Bruening & Al Khalaf's (2019) study with the same materials (translated into German)[21] and an exploratory investigation of subject *wh*-movement (a pilot for Experiment 2). Participants were recruited via prolific.co and 36 native speakers of German took part. We used the platform L-Rex (Starschenko & Wierzba 2020) for the web-based questionnaires. The instructions were adopted from Bruening & Al Khalaf (2019) with 100 stimuli presented to each participant: 48 critical items, 36 fillers, 8 items for the exploratory investigation, and 8 items that were direct translations of Bruening-AlKhalaf's materials. For the critical items, 144 data points were collected per condition/question (four from each participant). On average, the questionnaires took about 25 minutes to complete.

A.1. *Replication of the AP/DP experiment with the forced-choice method*

The results of our parallel study to the AP/DP experiment are summarized in Table 12.

Fitting a statistical model to all data from this experiment was impeded by the presence of complete separation (i.e. 100% positive responses and thus 0 variance) in the short AP conditions. This made it impossible to fit a converging model to the

Category	Movement	Distance	Preference for matrix / emb. R-expr.
AP	in situ	short	100.0% / 0.0%
AP	moved	short	100.0% / 0.0%
AP	in situ	embedded 1	97.9% / 2.1%
AP	moved	embedded 1	94.4% / 5.6%
AP	in situ	embedded 2	98.6% / 1.4%
AP	moved	embedded 2	88.9% / 11.1%
DP	in situ	short	96.5% / 3.5%
DP	moved	short	99.3% / 0.7%
DP	in situ	embedded 1	99.3% / 0.7%
DP	moved	embedded 1	96.5% / 3.5%
DP	in situ	embedded 2	99.3% / 0.7%
DP	moved	embedded 2	86.8% / 13.2%

Table 12
Responses to the forced-choice question in our replication of the AP/DP experiment.

[21] We thank the authors for making the materials, questions, and instructions available to us.

	Estimate	Std. Error	z-value	p-value
distance	0.29186	0.24676	1.183	0.236898
movement	1.05470	0.30278	3.483	0.000495
category	–0.20144	0.27961	–0.720	0.471266
distance*movement	–0.39656	0.24192	–1.639	0.101170
distance*category	–0.13423	0.23658	–0.567	0.570454
movement*category	–0.20584	0.24777	–0.831	0.406103
distance*movement*category	0.04624	0.23528	0.197	0.844189

Table 13
Summary of fixed effects for the forced-choice replication of the AP/DP experiment (subset of the data, excluding short conditions).

complete data set. However, visual inspection of the percentages in the short conditions shows that there is no trend toward less robust reconstruction with DPs (in contrast to the AP/DP experiment with the two-question method): the embedded R-expression was not chosen more frequently as a referent in the moved than in the in situ condition. For statistical analysis of the remaining data (embedded 1 and embedded 2), we decided to use sum-coding for DISTANCE (in a post hoc decision) to be able to compare these two levels to each other. The factors MOVEMENT and CATEGORY were also sum-coded, as in the AP/DP experiment. Within this subpart of the data, a significant main effect of MOVEMENT was found ($z = 3.48$, $p < 0.001$), but no significant main effect of CATEGORY nor DISTANCE. None of the interactions was significant. The full model output is shown in Table 13.

In comparison to the AP/DP experiment, the results of the replication show a divergence: the AP/DP asymmetry found in the AP/DP experiment was not detected in the forced-choice replication and coreference with DP movement was so low that, given a threshold-based logic, it would arguably be interpreted as evidence for a Condition C effect, a conclusion that is less obvious under our two-question method, where coreference with the embedded R-expression was more available.

A.2. Replication of Bruening & Al Khalaf's (2019) second experiment

The results of our replication are shown in Table 14, in comparison to the original English experiment. Both factors (MOVEMENT, ARGUMENT/ADJUNCT) were sum-coded. No main effect of MOVEMENT ($z = 0.37$, $p = 0.71$) nor ARGUMENT/ADJUNCT ($z = –0.18$, $p = 0.86$) was found, nor a significant interaction ($z = –0.15$, $p = 0.89$). The results show that coreference is much less available in German under forced choice as well, even though the same method and direct translations of the materials were used here. In the *wh*-movement conditions, coreference with the embedded R-expression was preferred much more frequently in the English study (arguments: 22%, adjuncts: 31%) than in our German replication (8%/6%).

621

movement	arg/adj	preference (German)	original results (English)
in situ	argument	98.6% / 1.4%	97.3% / 2.7%
in situ	adjunct	97.2% / 2.8%	98.7% / 1.3%
moved	argument	91.7% / 8.3%	78.0% / 22.0%
moved	adjunct	94.4% / 5.6%	69.3% / 30.7%

Table 14
Responses to the forced-choice question in our replication of Bruening & Al Khalaf's (2019) second experiment, in comparison to the results of the original English experiment.

movement	pronoun	positive answers 2-quest	preference forced-ch
in situ	personal pr.	Q1: 88.9% / Q2: 69.4%	83.3% / 16.7%
moved	personal pr.	Q1: 91.7% / Q2: 58.3%	94.4% / 5.6%

Table 15
Exploratory item set investigating the role of grammatical relation in coreference. Proportion of positive answers to the coreference questions in the AP/DP experiment in comparison to responses to the forced-choice question in the parallel experiment.

A.3. *Subject questions under forced choice*

In four exploratory items, we investigated subject *wh*-movement, a pilot for our Experiment 2. These items were also part of the AP/DP experiment. The results in both experiments are shown in Table 15. With the two-question method, coreference with the embedded R-expression is available to a substantial extent, as expected given that Condition C is not violated. With the forced-choice method, however, coreference with the embedded R-expression is available only to a very small extent, even though such examples are unquestionably grammatical. Given the threshold-based logic of such approaches, the 5.6% may be considered close enough to zero and thus – wrongly – be taken to indicate ungrammaticality.

Authors' addresses: (Salzmann)
University of Pennsylvania, Department of Linguistics, 3401-C Walnut Street, Philadelphia, PA 19104-6228, USA
msalzm@ling.upenn.edu

(Wierzba)
Universität Potsdam, Department Linguistik, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany
marta.wierzba@mailbox.org

(Georgi)
Universität Potsdam, Department Linguistik, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany
doreen.georgi@uni-potsdam.de