GLOSSA psycholinguistics

Philipp, M., & Pregla, D. (2025). Similarities in the processing of scrambling and quantifier scope ambiguities – A shared source? *Glossa Psycholinguistics*, 4(1): 14, pp. 1–35. DOI: https://doi.org/10.5070/G6011.6627

eScholarship University of California

Similarities in the processing of scrambling and quantifier scope ambiguities – A shared source?

Mareike Philipp, mareike.philipp@email.de

Dorothea Pregla, Health and Medical University Potsdam, dorothea.pregla@hmu-potsdam.de

In psycholinguistics, phenomena which are commonly analysed as involving some type of movement are associated with increased costs in sentence processing (Anderson, 2004; Bornkessel et al., 2002; Friederici et al., 2006; Gibson, 2000). But these increased costs are not the same for all speakers. In particular, while some speakers can rely on interpretations that require movement with no additional effort, other speakers either seem to reject them or have more difficulty with these interpretations (Brasoveanu & Dotlačil, 2015; Caplan et al., 2011; Caplan & Waters, 1999; Gil, 1982; Kemper & Liu, 2007; King & Just, 1991; Philipp & Zimmermann, 2020; Pregla et al., 2021; Vogelzang et al., 2019; Wingfield et al., 2003). However, what factors drive these individual differences is currently unclear. Knowing their sources will help us to sharpen our knowledge about the linguistic system. We therefore aim to investigate inter-individual differences in the processing of movement at the interface of semantics and syntax. Specifically, we take two phenomena in German which exhibit the variable processing pattern mentioned above: non-canonical object-before-subject word orders (scrambling) and inverse scope readings in quantifier scope ambiguities. We present a self-paced-reading experiment which gives additional insight into the processing of these two phenomena. While previous findings regarding processing difficulties can be replicated for both scrambling and quantifier scope, no evidence for an underlying connection can be found on the basis of individual participant data.

1. Background

The starting point for this article is the observation that two seemingly distinct phenomena exhibit a surprisingly similar behaviour in speakers. Both in the case of non-canonical object-before-subject word order as well as in the case of scope inversion in doubly-quantified sentences, participants show increased processing difficulty in experiments when reading or listening to such sentences (Anderson, 2004; Bornkessel et al., 2002; Friederici et al., 2006; Gibson, 2000). Further, in both cases, these processing difficulties are not stable, but vary greatly between participants (Brasoveanu & Dotlačil, 2015; Caplan et al., 2011; Caplan & Waters, 1999; Gil, 1982; Kemper & Liu, 2007; King & Just, 1991; Philipp & Zimmermann, 2020; Pregla et al., 2021; Vogelzang et al., 2019; Wingfield et al., 2003). Interestingly, these two phenomena not only cause similar patterns in psycholinguistic experiments, they are also both commonly analysed as involving some type of movement in theoretical linguistics. The overarching goal of this article is, therefore, to test if the similarity in observable patterns is related to the same underlying source, namely, movement.

In the remainder of this section, we will first provide some background information on each of the two phenomena: scrambling (1.1) and quantifier scope ambiguities (1.2) in German. We will then provide the basis for the experiment by showing the commonalities of the two phenomena which justify discussing them under the same umbrella (1.3).

1.1 Scrambling

In German declarative sentences, different word orders are possible. The subject (S) can precede the object (O), which is the more common, *canonical* word order, or the subject can follow the object, which is a non-canonical word order. An example is given in (1).

(1) *German*

a. SO: Hier sieht **der** Mann **den** Jungen.
here sees the.NOM man the.ACC boy
'Here, the man sees the boy.'

b. OS: Hier sieht **den** Mann **der** Junge.
here sees the.ACC man the.NOM boy
'Here, the boy sees the man.'

In the present study, we consider sentences in which the subject and object occur in the Mittelfeld, that is, after the finite verb in the main clause (Abels, 2015). The free variation of word order in the German Mittelfeld is called *scrambling*. There are different approaches to analyzing scrambling, e.g., the different word orders could be base-generated or result from a post-syntactic

PF-linearization mechanism[1] (Abels, 2015). However, the most prominent approach is to assume a movement analysis in which, in the non-canonical sentence, the object is moved from its base position (Salzmann, to appear). Many authors provide theoretical arguments to support this account (e.g., Haider & Rosengren, 2003; Sabel, 2005; Stechow & Sternefeld, 1988). For example, the word orders produced by scrambling obey the same constraints as wh-movement, and, as of yet, only movement theories of scrambling can explain this parallel (Haider, 2017; Salzmann, to appear). Another argument that is currently only accounted for by movement theories is the distribution of floating quantifiers (Heck & Himmelreich, 2017). Additionally, the results of psycholinguistic experiments support the movement account over the base-generation account (e.g., Clahsen & Featherston, 1999).

It has been shown with various experimental methods that declarative sentences in which the object is scrambled over the subject are difficult to process. Since this study investigates scrambling, the following review only includes research on sentences in which the subject and object occur in the Mittelfeld (for research on sentence-initial objects, see Bader & Portele, 2021). Furthermore, the review is restricted to studies that used items with masculine singular nouns which have unambiguous case marking, because this study is concerned with the processing of movement rather than garden paths or ambiguity resolution (for a literature review on noun phrases that have an ambiguous case marking, see Schunack, 2016). This means that in the studies presented below, participants knew if they were confronted with a non-canonical structure as soon as they encountered the first noun phrase (NP). Nevertheless, participants had difficulty processing scrambled declaratives. For example, in a study using event related potentials (ERP), Bornkessel et al. (2002) showed that a negativity occurs from 300–450 ms post onset of the determiner of the first argument in sentences with OS order, in contrast to sentences with SO order. This effect has been called *scrambling negativity,* because it also occurred in other ERP studies that investigated scrambled sentences (Bornkessel & Schlesewsky, 2006; Rösler et al., 1998; Schlesewsky et al., 2003). This negativity might be due to a mismatch between the predicted canonical word order and the occurring non-canonical word order, which leads to an increase in processing (Bornkessel et al., 2002). Furthermore, it was demonstrated by means of functional magnetic resonance imaging that scrambling leads to an increased activation in the left frontal and temporal regions (Friederici et al., 2006;

[1] Following Kayne (1994) and its integration in Chomsky (1995), it is a common generative notion that syntax only determines hierarchical relations. Linear order, on the other hand, is not represented in syntax. Linearization is only necessary because human articulatory anatomy doesn't allow for simultaneous production of lexical items. Therefore, linear order is thought of as an extrasyntactic artefact that requires mapping rules from hierarchical relations to linear relations. Such mapping rules are commonly described as post-syntactic rules that merely affect phonological form (PF), devoid of semantic effects (cf. Chomsky, 2013).

Grewe et al., 2005; Röder et al., 2002; Vogelzang et al., 2020). The increased activity was attributed to the need to perform language-internal operations to derive the underlying word order (Friederici et al., 2006). In visual world eye-tracking, the proportion of looks to the target picture was lower in sentences with OS order than in sentences with SO order from the first NP until the end of the sentence (Pregla et al., 2022). Additionally, listening times for the first NP in self-paced listening were increased in OS sentences in comparison to SO sentences (Pregla, 2023). The findings in eye-tracking and self-paced listening were explained by a greater demand of cognitive resources in scrambled sentences. The resource demand was said to be increased in OS sentences due to the syntactic operation of associating the object with its base position, which is not necessary in the SO sentences. The difference between scrambled and canonical word order is also reflected in acceptability or grammaticality judgements, where sentences with OS order were judged worse than sentences with SO order (Bader & Häussler, 2010; Keller, 2000; Meng et al., 1999; Pechmann et al., 1996). An explanation for the worse judgement might be the lower frequency of OS sentences (Bader & Häussler, 2010). Finally, the increased processing effort for declarative sentences with scrambled arguments is visible in reduced response accuracy and increased reaction times in sentence comprehension tasks (e.g., Pregla et al., 2021; Vogelzang et al., 2019).

As summarized in the previous paragraph, declarative sentences with scrambled case-unambiguous arguments are difficult to process. In a movement analysis, this difficulty could stem from the increased processing effort required to derive the underlying SO order, or the need to revise the predicted SO order. Outside a movement analysis, scrambled sentences could be more difficult because they are less frequent or because they are grammatically dispreferred (Bornkessel et al., 2002). Additionally, OS sentences could be more difficult because they are semantically more complex than SO sentences, under the assumption that the central aim of the processing system is to identify the actor role (Bornkessel-Schlesewsky & Schlesewsky, 2014). From this assumption, it follows that the processing system prefers an order in which the actor comes first and that every argument is preferably assumed to be an actor (Bornkessel-Schlesewsky & Schlesewsky, 2014). In sum, processing difficulty for scrambled OS sentences should occur regardless of whether scrambling is assumed to be the result of movement.

Importantly, not all participants show increased response times for non-canonical sentences, and comprehension difficulties are variable (Pregla et al., 2021; Vogelzang et al., 2019). Wingfield et al. (2003) found individual differences in the comprehension of non-canonical sentences by younger and older participants, as difficulty increased in older participants. Furthermore, participants with different working memory capacities exhibited differences in syntactic processing (Kemper & Liu, 2007; King & Just, 1991) or in task performance (Caplan et al., 2011; Caplan & Waters, 1999) for non-canonical sentences. Thus, the non-canonical word order seems to affect participants differently.

1.2 Quantifier scope ambiguities

Quantifier scope ambiguities can arise whenever two or more quantificational expressions (*some*, *every*, …) occur in the same clause. An example is given in (2), where the subject *canal* is quantified by the existential *a* and the direct object *field* is quantified by the universal *every*. This sentence can have two different readings. Under the surface reading, there is only a single canal, and this canal irrigates all the fields. Under the inverse reading, different fields may be irrigated by different canals. In the surface reading, the existential, as the linearly first occurring quantifier, has scope over the universal, while under the inverse reading, the universal has scope over the existential.

(2) A canal irrigated every field.
   a. Surface reading: $\exists x(\text{canal}(x) \,\&\, \forall y(\text{field}(y) \rightarrow \text{irrigate}(x,y)))$ $\exists > \forall$
   b. Inverse reading: $\forall y(\text{field}(y) \rightarrow \exists x(\text{canal}(x) \,\&\, \text{irrigate}(x,y)))$ $\forall > \exists$

One very common way of representing these two different readings is via Quantifier Raising (QR), introduced in May (1977, 1985). This is a covert movement operation at the level of Logical Form (LF). The relative ordering of quantifiers at LF determines the reading: in order to arrive at the inverse reading, the universal in (2) has to move covertly to a c-commanding position above the existential; see (3).

(3) a. [A canal] irrigated [every field].
   b. $[\text{every field}]_i$ [a canal] irrigated $[t_i]$.

May's account has been one of the standard accounts of quantifier ambiguities since its first introduction, and has been reused in different modified versions (e.g., Fox, 2000; Reinhart, 2006). However, there are also approaches to quantifier scope ambiguities which do not rely on some kind of movement operation and the postulation of an abstract level of LF, such as the reconstruction account of Frey (1993), the multi-factorial account of Ioup (1975) or Pafel (2005), and the semantically based accounts of Barker (2005, 2012), Jacobson (1996), Steedman (2012), and others. One main reason for the popularity of May's account is that the availability of inverse readings strongly parallels the acceptability of overt movement. That is, in environments where overt movement, such as wh-movement, is banned or leads to a strong reduction in acceptability (known as *syntactic islands*), inverse readings are also either unavailable or much more difficult to obtain. This is supported by both introspective judgments in the literature (e.g., Huang, 1995; May, 1985) as well as experimental data (Philipp & Zimmermann, 2023; Scontras et al., 2017; Tanaka, 2015).[2]

[2] Over time, this claim has received some criticism, though. Some authors have pointed out examples where inverse readings across island boundaries seem to be available (e.g., Hulsey & Sauerland, 2006; Szabolcsi, 2012). Further, the experiments in Scontras et al. (2017) and Tanaka (2015) only show a reduction in availability, not a complete

The presence of more than one quantificational expression is a necessary but not a sufficient condition for quantifier scope ambiguities. A number of different factors have been proposed to have an impact on the availability of inverse readings. In some languages, such as English, inverse readings are generally and readily available (Anderson, 2004; Gillen, 1991; Kurtzman & MacDonald, 1993; Tunstall, 1998), while in other languages, such as Chinese, inverse readings are either non-existent or strongly dispreferred (Aoun & Li, 1993; Scontras et al., 2017). The reason for these cross-linguistic differences is not fully known yet. It is possible to assume that whether or not QR as a covert movement operation is available at all depends on the specific grammar of the language. Other authors, like Bobaljik and Wurmbrand (2012), suggest that QR is universally available and, instead, differences in, e.g., word order flexibility can serve as an explanation. However, also within a specific language, there is a lot of variation dependent on factors such as intonation (e.g., Büring, 1997; Krifka, 1998), lexical semantics of the quantifiers (e.g., Beghelli & Stowell, 1997; Ioup, 1975), information structure (e.g., Partee, 1991; Surányi & Turi, 2017), context/world knowledge (e.g., Saba & Corriveau, 2001; Villalta, 2003), grammatical role (e.g., Ioup, 1975; VanLehn, 1978), and others.

All of the variation described above is generally acknowledged to have an impact on the availability or non-availability of inverse readings only, while the surface reading should always be an option.[3] The inverse reading is not only shown to be generally dispreferred and less available than the surface reading (Anderson, 2004; Gillen, 1991; Kurtzman & MacDonald, 1993; Reinhart, 2006; Tunstall, 1998), but is also associated with higher processing costs (e.g., Anderson, 2004; Bott & Schlotterbeck, 2015; Gillen, 1991; Wurmbrand, 2018), which is commonly thought of as the source for its general dispreference. This dispreference, along with the higher processing costs, fits neatly into the covert movement approach described above: the inverse reading requires an additional movement step, as the lower quantifier has to be raised to a position where it can c-command the other quantifier. This additional movement step is responsible for the increase in processing costs, which, in turn, is responsible for the reduced availability (Anderson, 2004; Wurmbrand, 2018). While the connection between QR-steps and processing costs has been most prominent in the literature, it would, in principle, also be possible to explain increased processing costs with other analyses of quantifier scope. For example, under type shifting accounts (e.g., Hendriks, 1988, 1993; Jacobson, 1992; Shan & Barker, 2006), the inverse reading would be associated with a shift to a more complex semantic type and that, in turn, could cause increased processing costs. In addition to increased processing costs, it

absence, of inverse scope. On the other hand, Tanaka (2015), shows that overt wh-movement, too, only exhibits a reduction in acceptability, not a complete blocking. She also shows that this reduction varies with the type of island and that these island-dependent acceptability patterns of overt wh-movement parallel the availability pattern of inverse readings.

[3] But note that several experiments on inverse scope have observed non-ceiling effects in the case of the surface readings (e.g., Fanselow et al., 2022; Kurtzman & MacDonald, 1993; Philipp & Zimmermann, 2023).

has also been shown that the difficulty of obtaining the inverse reading varies greatly between speakers of the same language, with some speakers accepting inverse readings across the board, some speakers only accepting them sometimes, and some speakers rejecting them altogether (Brasoveanu & Dotlačil, 2015; Gil, 1982; Philipp & Zimmermann, 2020). It is not clear yet, which factors contribute to these inter-individual differences.

In this article, we will look specifically at quantifier scope ambiguities in German. While in the theoretical literature, inverse readings were thought of as being restricted to only very specific environments (Bobaljik & Wurmbrand, 2012; Frey, 1993; Pafel, 2005), experimental work has shown that inverse readings, despite being clearly dispreferred in German, are, in fact, available in more environments than predicted by the different theoretical accounts (Bott & Schlotterbeck, 2015; Fanselow et al., 2022; Philipp & Zimmermann, 2020, 2023; Radó & Bott, 2018). Similarly to other languages, these experiments provide evidence from judgment tasks that there is a general preference for surface over inverse readings (Bott & Radó, 2007; Bott & Schlotterbeck, 2012, 2015; Radó & Bott, 2012; 2018). They also show that additional factors play a role, such as syntactic construction (Bott & Schlotterbeck, 2015; Philipp & Zimmermann, 2020, 2023; Radó & Bott, 2012), the semantics of the quantifiers (Bott & Radó, 2007; 2009; Radó & Bott, 2012, 2018), and context/world knowledge (Philipp & Zimmermann, 2020). Additionally, Bott and Schlotterbeck (2012, 2015) provide data on the processing difficulty associated with quantifier scope ambiguities in German. Bott and Schlotterbeck (2012) employed an Incremental Truth Value Judgment Task, and found longer reading times for the ambiguous vs. the unambiguous sentences, which they attribute to participants trying to resolve the scope conflict in the case of ambiguity. Using self-paced reading and eye-tracking methods, Bott and Schlotterbeck (2015) found that processing difficulty associated with scope inversion does not immediately arise on the second quantifier, but only later, at the sentence boundary. They interpret these results as indicating that inverse scope in German is not computed immediately, but only once the predication is complete.

1.3 Scrambling and quantifier scope united

Section 1.1 introduced the topic of scrambling, and 1.2 introduced the topic of quantifier scope ambiguities. Although these are two very distinct phenomena, they show certain similarities. Both of them are commonly accounted for by assuming some type of movement, overt A-movement in the case of scrambling and covert A'-movement in the case of quantifier scope.[4] In both

[4] A-movement and A'-movement are both instances of phrasal movement. *A-movement* is short for argument movement, and it describes movement where the landing site is an argument position. This is a position in the syntactic tree where a theta role can be assigned. *A'-movement* is the opposite, namely, movement into a non-argument position. These definitions are not relevant for the remainder of this article, but it is worth mentioning that we are dealing with two different types of movement.

cases, there is one preferred option (canonical word order in scrambling and surface reading in quantifier scope) and one dispreferred but available option (non-canonical word order vs. inverse reading). In both cases, the dispreferred option is the one which requires additional movement steps and is associated with higher processing costs visible in longer reading or reaction times. And finally, both scrambling and quantifier scope show strong inter-speaker variability, with some speakers having no difficulty with the dispreferred option at all, while others showing some or even great difficulty. While, in the case of scrambling, there is evidence that differences in working memory can explain at least part of this individual variability in off-line responses (Caplan et al., 2011; Kemper & Liu, 2007; King & Just, 1991), no such studies have been conducted yet in the case of quantifier scope.

In light of these similarities between the two phenomena, we hypothesize that there could be a very basic, shared source for the processing difficulties associated with them. That is, the similar patterns of preferred vs. dispreferred option and the inter-individual variability regarding that preference could be associated with additional movement steps, in that some speakers have generally more difficulty with movement than other speakers. However, the difficulty might not be related to any kind of movement per se. As pointed out by a reviewer, this would otherwise imply that the same should be observed with movement that obligatorily applies to satisfy strict word order requirements, as is the case with wh-questions and verb-second sentences in German. This seems implausible, and, as far as we know, there is no evidence for processing difficulties with such sentences. Instead, we would like to suggest that difficulties arise with *optional* movement, as is the case both in scrambling and scope inversion. The underlying issue with optional movement might stem from the fact that speakers tend to follow a default strategy (e.g., expecting canonical word order and surface scope), where deviation from this default incurs processing costs. These costs would, then, not be caused by movement as such, but by movement as an additional step to revise an initial default expectation. Individual differences would, then, arise because some speakers have greater difficulty revising their expectations and deviating from the default syntax. This would predict that those participants who exhibit processing difficulties with non-canonical word orders should also have more difficulty with inverse readings, and vice versa. Alternatively, the difficulties could be unrelated and only arise due to problems which are specific to each phenomenon.

Besides that, there are also alternative explanations why processing difficulties are observed in both phenomena; see also the alternative accounts mentioned in 1.1 and 1.2. Difficulties might also arise because one instantiation of the phenomenon is simply more frequent in the input than the other (SO > OS, surface reading > inverse reading). In fact, for both phenomena, scrambling and quantifier scope, there is a debate in the literature whether difficulties arise due to an underlying syntactic principle or mere surface frequency; see, e.g., Bornkessel et al. (2002). Yet another possibility is that difficulties are related to greater semantic complexity. With quantifier scope, this could be type shifting in the case of inverse readings, and with scrambling, it could

be the unusual order of thematic role assignment in OS sentences. Importantly, even under those alternative explanations, there is a connection between the phenomena: Both the inverse scope reading and the OS structure are less frequent and semantically more complex than their counterparts. Thus, this would not undermine our main claim that processing difficulties could be due to a shared underlying source and, thus, be correlated, be it syntactic movement, frequency, or semantic complexity (or even a combination of these). If the hypothesis is borne out, we should see consistent intra-individual behaviour across linguistic phenomena in either case, supporting the assumption that there are more general speaker-specific processing difficulties associated with some level of increased complexity. On the other hand, if we do not see such a consistent intra-individual behaviour across the two phenomena, this indicates that the source of difficulty varies.

In Section 2, we present an experiment that investigates the hypothesis that scrambling and quantifier scope share an underlying mechanism which causes the observed similarity in processing patterns.

2. Design

This study has the following goals: (i) We want to see if the results of previous scrambling and scope experiments can be replicated, especially with regards to by-participant variability; (ii) We want to investigate if this variability is reflected in on-line processing data; (iii) We want to compare participants' behaviour in items with scrambled word order to their behaviour in items with two scopally interacting quantifiers, in order to test the hypothesis that the difficulties could arise due to a shared underlying source.

2.1 Materials

The full set of conditions used in this experiment is given in (4). The experiment had three conditions: a baseline condition (A), a scrambled condition (B), and a scope-ambiguous condition (C). Conditions A and B were unambiguous, since there was only one quantifier in the sentence, namely, in NP2, while NP1 contained a definite article. Condition C was ambiguous, since it contained a second quantifier, namely, the indefinite article as an existential quantifier on NP1. Conditions A and C had canonical SO order, while condition B had the scrambled OS order. Thus, condition A was the unambiguous and canonical baseline condition, which the other two conditions, B and C, were manipulated for the two phenomena of interest. There were 60 target items in total, which amounts to 20 items per condition and participant. Each participant saw each item only in one of the three conditions. Items consisted of a context sentence and a target sentence. The context introduced the content of the subject and object NP (in (4): *Äcker* 'fields', *Kanäle* 'canals') and the verb (in (4): *bewässern* 'irrigate') of the target sentence, such that all lexical material in the target sentence qualifies as given. This was done to control for information structure and intonation, thereby avoiding potentially confounding effects in the case of scope interpretation

(see, e.g., Krifka, 1998). The target sentence appeared in either conditions A, B, or C. Each item was followed by one of the two possible statements exemplified in (4), which had to be judged as true or false with respect to the preceding target sentence. Condition C was always followed by the plural statement to assess the availability of inverse readings. When participants judged the plural statement as true, we took this to mean that they had obtained the inverse reading. For the unambiguous conditions A and B, for which the statements were either clearly false or true, the presented statement alternated. If the inverse reading for the types of sentences we are using in our study does in fact exist in German, the acceptance of the plural statement in condition C should be higher than the acceptance of the false statement in the unambiguous conditions A and B.

(4) *German*

Context: Der Agrarexperte hatte empfohlen, dass die Äcker durch Kanäle bewässert werden sollten.

'The agriculture expert had advised that the fields be irrigated by canals.'

(A) Baseline condition: canonical word order, definite NP 1 (unambiguous)[5]

Und tatsächlich bewässerte **der breite Kanal** dann **jeden**
and indeed irrigated the.NOM wide.NOM canal then every.ACC
Acker in der Region der Bauern.
field in the region the.POSS farmers.POSS
'And then, indeed, the wide canal irrigated every field in the region of the farmers.'

Follow-up statements:

a. Singular:
Insgesamt gab es hier **genau einen** Kanal, der Äcker bewässert hat.
overall there.was here exactly one canal that fields irrigated has
'Overall, there was exactly one canal which irrigated fields.'

b. Plural:
Insgesamt gab es hier **mehrere** Kanäle, die Äcker bewässert haben.
overall there.were here several canals that fields irrigated have
'Overall, there were several canals which irrigated fields.'

(B) Scrambled condition: non-canonical word order, definite NP 1 (unambiguous)[6]

Und tatsächlich bewässerte **den breiten Acker** dann **jeder**
and indeed irrigated the.ACC wide.ACC field then every.NOM

[5] Note that there is a potential confound in this condition. The definite determiner carries a uniqueness presupposition, which is violated here. We will discuss this issue in 4.2.

[6] Note that, while the canonical word order is subject-before-object, it is known that information structure can improve the non-canonical object-before-subject word order. According to Lenerz (1977) and Skopeteas and Fanselow (2009), the non-canonical word order is best if the object is given and the subject is new. In our example, however, both subject and object are given. Therefore, it is conceivable that potential processing effort that participants might exhibit in condition B is not, or at least not only, due to the non-canonical word order, but to the combination of a non-canonical word order and an information structure that does not support this word order.

Kanal in der Region der Bauern.
canal in the region the.POSS farmers.POSS
'And then, indeed, every canal irrigated the wide field in the region of the farmers.'

Follow-up statements:

a. Singular:
Insgesamt gab es hier **genau einen** Kanal, der einen Acker
overall there.was here exactly one canal that a canal
bewässert hat.
irrigated has
'Overall, there was exactly one canal which irrigated a field.'

b. Plural:
Insgesamt gab es hier **mehrere** Kanäle, die einen Acker
overall there.were here several canals that a field
bewässert haben.
irrigated have
'Overall, there were several canals which irrigated a field.'

(C) Scope-ambiguous condition: canonical word order, indefinite NP 1 (ambiguous)
Und tatsächlich bewässerte **ein breiter Kanal** dann **jeden**
and indeed irrigated a.NOM wide.NOM canal then every.ACC
Acker in der Region der Bauern.
field in the region the.POSS farmers.POSS
'And then, indeed, a wide canal irrigated every field in the region of the farmers.'

Follow-up statement:

a. Plural:
Insgesamt gab es hier **mehrere** Kanäle, die Äcker bewässert
overall there.were here several canals that fields irrigated
haben.
have
'Overall, there were several canals which irrigated fields.'

2.2 Participants

We tested 66 participants,[7] who were acquired via the SONA participant pool of the University of Potsdam. 54 of the participants were female and 10 were male, and they were, on average, 26 years old (18–59). Two participants did not provide information about age and sex. The participants could choose between 1h of course credit or a monetary reimbursement of 7€. 19 participants were excluded from the analysis because they did not perform above threshold in the control items (less than 17 correct responses in the 20 unambiguous and canonical sentences

[7] The sample size was limited by availability of participants and financial means.

of condition A).[8] This left us with 47 participants for the final analysis, with a mean age of 25 (18–59) and 39 female and 8 male participants.

2.3 Method & procedure

We ran a cumulative self-paced-reading (SPR) experiment, together with an off-line truth-value-judgment task (TVJ). We ran this experiment as a web-based study on Ibex Farm. In each trial, participants were first presented with the context sentence (see example (4)). After pressing the space bar, the first word of the target sentence was displayed. Participants uncovered the target sentence incrementally, word by word, by pressing the space bar, thereby controlling the speed of presentation themselves. The context and the words already uncovered remained on the screen until the whole target sentence was uncovered. We opted for this rather unusual procedure in SPR, because pre-tests showed that participants perceived the sentences as relatively long and complex and had difficulties understanding the sentence and keeping it in memory in order to do the off-line task accurately. We will come back to potential problems with this procedure in the discussion in Section 4. After each item, participants pressed the button one more time, which caused the item to disappear from the screen and the task appeared instead. In the task, participants had to judge a statement about the target sentence with respect to its truth by pressing *v* if the statement was true or *n* if the statement was false (see (4)). At this point, the target sentence was not visible anymore. Both the reading times and the reaction times to the off-line task were recorded.

2.4 Predictions

We expected longer reading times on the first NP without the noun[9] in conditions B vs. A (***den breiten*** *Acker* 'the wide field' vs. ***der breite*** *Kanal* 'the wide canal') and potentially on the spillover region *dann* ('then') after the NP, since non-canonical word orders are known to cause processing difficulties. We also expected longer reading times on the second NP in conditions C vs. A (*jeden Acker* 'every field'), since, in the former case, participants potentially have to resolve a scope ambiguity at that point. As for the off-line data, the responses allow us to conclude (i) if they paid attention at all, and (ii) how they interpreted the ambiguous sentences. According to the

[8] This relatively high number of exclusions can be explained by a specific subgroup of participants who misunderstood the experiment task: they did not judge the truth of the statement against the second sentence only, as demanded in the task description, but against the whole item. Their response patterns were, thus, distinctively different from the other participants and easy to identify. In the unambiguous conditions A and B, which should only allow for one answer, they responded correctly in only 8–16 out of a total of 20 items in each condition. The response patterns of the remaining 47 participants show that they understood the task correctly.

[9] Because the noun itself varied between conditions A and B of the same item, we only took the reading times of the NP without the noun, i.e., determiner + adjective. Note that in the comparison of conditions A vs. C, dropping the head noun was not necessary, as the order of nouns was identical in these conditions.

hypothesis that the processing difficulties of the two phenomena – scrambling and quantifier scope ambiguities – share an underlying source, we expected those participants who more often allow the inverse reading in condition C to show smaller differences in reading times between conditions A and B, and vice versa. Smaller differences are taken to indicate less difficulty with non-canonical sentences.

2.5 Data analysis

Data analysis was performed on reading times in the SPR task and on accuracy as well as response time in the TVJ task. Target sentences were presented word by word. In the analysis, we summed up the reading times of certain words to form regions. Specifically, we analysed the following regions: 1) determiner + adjective of the first NP; 2) *dann* 'then'; 3) determiner + noun of the second NP; 4) preposition + determiner + noun of the PP; 5) determiner and noun of the outro. We fit Bayesian linear models using R (Version 3.6.3; R Core Team, 2020) and the R-package brms (Version 2.13.0; Bürkner 2017, 2018). Reading times were log-transformed, because they cannot be negative and have a longer right tail. TVJs are binary (0 and 1) and, therefore, we fit generalized linear models with a logistic link function. We will report estimates back-transformed into milliseconds and proportions for ease of interpretation.

With respect to the TVJs, we used Helmert contrasts comparing, first, conditions A and B (–1 baseline, +1 scrambling) and, second, the mean of conditions A and B to condition C (–1 baseline, –1 scrambling, +2 quantifier scope). With respect to the reading times, we used sum contrasts and fit the models separately to compare conditions A and B (–1 baseline, +1 scrambling) and to compare conditions A and C (–1 baseline, +1 quantifier scope). The comparison of conditions A and B additionally included the number of inverse readings per participant, centred around the mean, as a predictor and the interaction of condition and number of inverse readings. For the spill-over region *dann* and the outro, the comparisons of conditions A and B furthermore included the length and frequency on the Zipf scale (cf. Van Heuven et al., 2014) of the preceding noun as predictors.[10] The length and frequency information was taken from the DWDS corpus (Geyken et al., 2017). The comparison of conditions A and C additionally included accuracy as a predictor, as well as the interaction of condition and accuracy. Correct judgements (condition A) and surface readings (condition C) were coded as +1, and incorrect judgements (condition A) and inverse readings (condition C), as –1. Models included random intercepts and slopes for items and participants.

[10] These predictors were included, since a t.test revealed a significant difference between the nouns of condition A and B in word length (SO: word length = 8.7, OS: word length = 7.8, t = 2.6) and in frequency (SO: Zipf = 3.1, OS Zipf = 3.7, t = –2.9). We want to thank the editor for pointing out the potentially confounding effects of word length and frequency.

The prior distributions for the parameters in the models for the accuracy of the TVJs were specified as follows: The prior of the intercept was set to Normal(0, 1.5), the priors of the slopes were set to Normal(0, 1), and the prior standard deviations of the random effects, to Normal + (0, 1) truncated at zero, because standard deviations cannot be negative. The prior distributions for the parameters in the models for the reading times in the SPR task and for the response times in the TVJ task were specified as follows: The prior of the intercept was set to Normal(0, 10), the priors of the slopes were set to Normal(0, 1), and the prior standard deviations of the random effects and the residual error, to Normal + (0, 1) truncated at zero.[11] The model estimates the posterior distributions of the parameters. From this, we extracted the mean and the 95% credible interval (CrI). The CrI is the range of plausible values of the parameters, given the data and model.

3. Results

3.1 Accuracy judgments

Accuracies are depicted in **Figure 1**. Participants gave the correct response in 94.9% of the sentences in condition A and in 95.1% of the sentences in condition B, and the estimated difference between conditions A and B was negligible (1.5% CrI: [–0.3,3.4]). The inverse reading was accepted in 14.2% of the trials in condition C. The acceptance of inverse readings is about 10% higher than the noise level in conditions A and B (∼5%). Furthermore, the proportion of correct responses in conditions A and B was estimated to be 4% higher than the proportion of surface readings selected in condition C (4.3% CrI: [0.9, 8.7]). We take this as support for the hypothesis that we are dealing with true inverse interpretations and not just with spurious effects, but see 4.2.

3.2 Reading times

The reading times for the whole sentence and the critical regions of the sentence are shown in **Figures 2** and **3**. Differences between the sentences in SO and OS order were negligible on the first NP's determiner and adjective (7 ms CrI: [–2, 17]). But, as expected, we found longer reading times in sentences with OS order in the spillover region (*dann* 'then', 32 ms CrI: [18, 48], see **Figure 3**). This is in line with previous studies of non-canonical word order in German,

[11] A reviewer pointed out that there is, in fact, data from previous research, e.g., on scrambled sentences, which could have informed the priors. Nevertheless, the reason why we set them in this way is that previous research used non-cumulative paradigms. Since, in our case, the words stay on the screen, participants can reassess the critical region after uncovering the whole sentence. That is, they might delay at least some of the processing of the critical region until the whole sentence is displayed. Given this possible difference in processing at the critical word, the effect size might be different between the two versions of the task. Therefore, the previous data were not used to inform the prior. As can be seen in Section 3, our concerns about this were indeed justified.

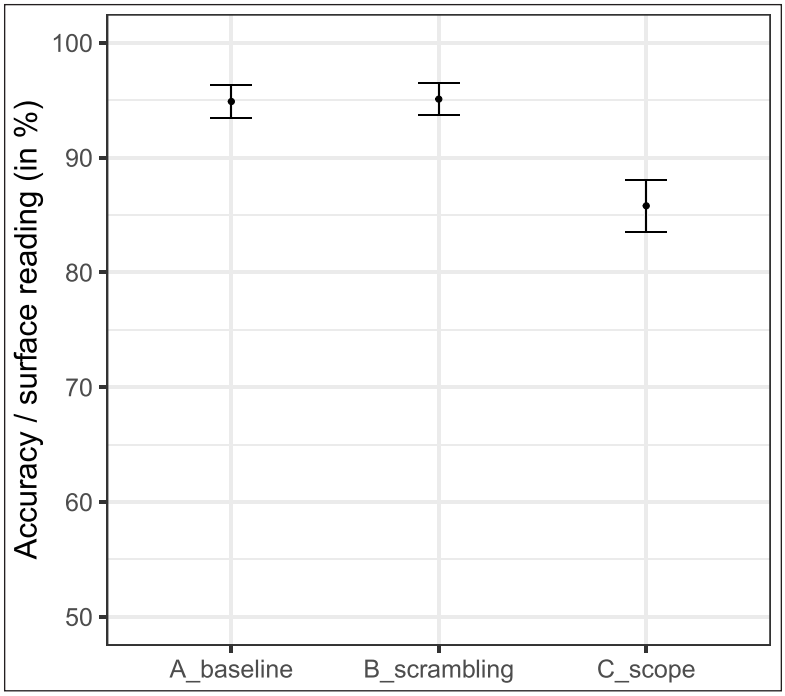

Figure 1: Mean accuracy per condition in the truth-value-judgment task. For the unambiguous conditions A and B, accuracy always refers to the proportion of correct responses. For the ambiguous condition C, accuracy always refers to the proportion of surface reading responses.

indicating that there is more processing effort involved in non-canonical OS order compared to canonical SO order. With respect to the baseline condition A and the quantifier scope condition C, we did not find any difference on the second NP or on the subsequent spillover region,[12] as the estimated difference was situated around zero in both regions (NP2: –17 ms CrI: [–45, 10]; spillover region: –7 ms CrI: [–46, 32]). However, there was an interaction between condition and accuracy (inverse reading in condition C was coded as 'correct') on the second NP and the subsequent spillover region (NP2: 41 ms CrI: [14, 69]; spillover region: 59 ms CrI: [20, 100]). To further explore this interaction, we investigated the effect of accuracy nested under condition; see **Figure 4**. In condition A, reading times were longer when participants chose the incorrect answer in the TVJ task (NP2: 55 ms CrI: [–4, 118]; spillover region: 54 ms CrI: [–13, 129]). In condition C, reading times were longer when participants chose the inverse interpretation in the TVJ task (NP2: 32 ms CrI: [–1, 66]; spillover region: 73 ms CrI: [23, 128]). Thus, the analysis of

[12] As mentioned before, the first NP's noun was not included in the main analysis, because it differed between the OS and SO conditions. Based on a reviewer's advice, we did an additional analysis of the whole NP and N1 separately. These analyses included word length and frequency of the noun as predictors. As expected, we see longer reading times in sentences with OS order, both in the whole first NP and in the first noun (NP1: 39 ms CrI: [5, 73], N1: 16 ms CrI: [2, 30]).

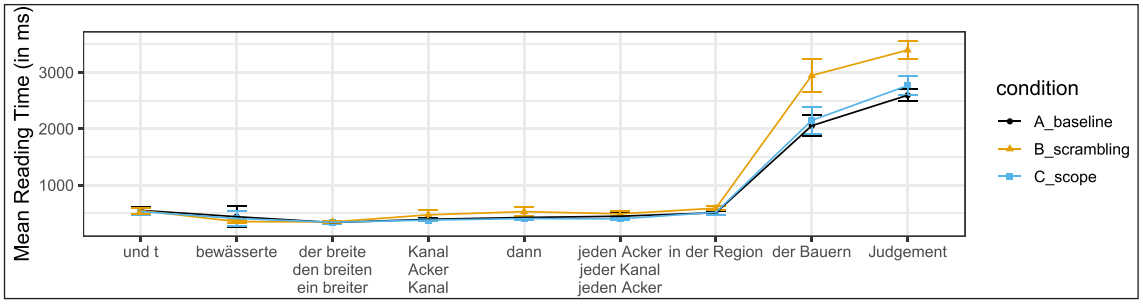

Figure 2: Reading times per condition in the target sentence and response times of the truth-value-judgment task. Error bars represent confidence intervals.

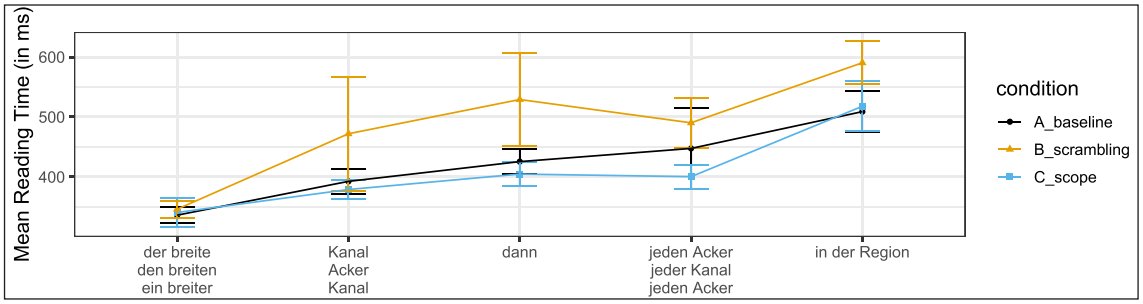

Figure 3: Reading times per condition across the critical regions of the target sentence. Error bars represent confidence intervals.

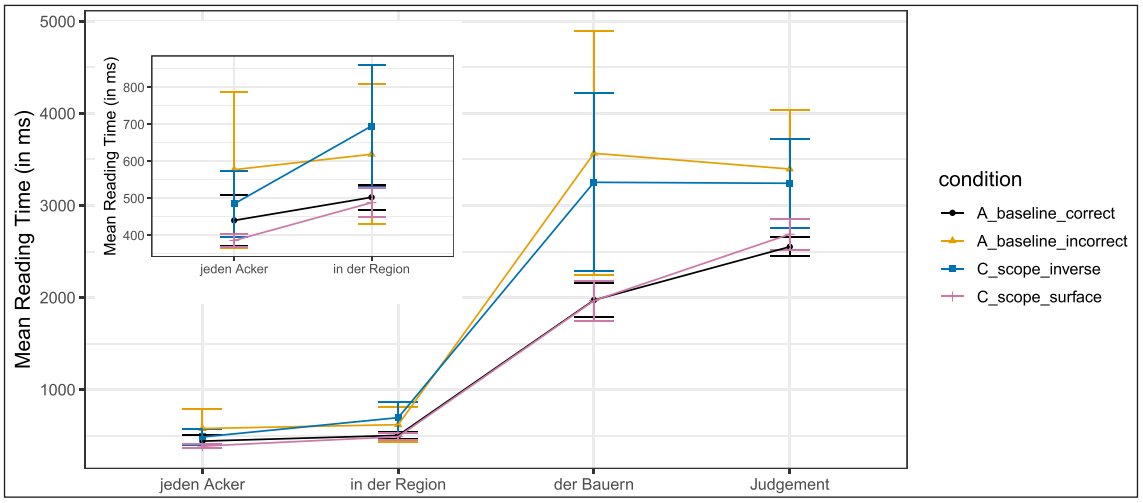

Figure 4: Reading times on the second NP, the spillover region and the final region, and response times in the truth-value-judgment task, for the baseline condition A and the quantifier scope condition C, depending on the judgement. For condition A, black colour refers to sentences in which the judgement was correct, and yellow colour refers to sentences in which the judgement was incorrect. For condition C, blue colour refers to an inverse reading response, and pink colour refers to a surface reading response. Error bars represent standard errors.

the nested effects showed that reading times were both longer in trials where participants opted for the wrong interpretation in condition A and in trials where participants opted for the inverse interpretation in condition C. Based on this result, we refit the previous model that included an interaction term for accuracy × condition. As before, we coded the inverse reading as incorrect and the surface reading as correct in condition C. In these analyses, the interaction turned out to be situated around zero (NP2: –12 ms CrI: [–38, 14]; spillover region: 9 ms CrI: [–28, 47]).

As can be seen in **Figure 2**, reading times are relatively short throughout the sentence, but they are longer on the final region. Based on this result, we suspect that participants might have uncovered the whole sentence first before fully processing it. Therefore, we decided to also explore the difference in reading times between conditions in the sentence final region and the response times in the TVJ task. With respect to the difference between conditions A and B, similar results as in the earlier regions were obtained, namely, sentences with OS order were read more slowly than sentences with SO order. However, effect sizes were more than ten times larger than at the critical or spillover region (sentence final region: 498 ms CrI: [267, 828], TVJ: 613 ms CrI: [439, 800]). Also, the result for the difference between conditions A and C was similar to the critical and the spillover region. The difference between A and C was situated around zero (sentence final region: –77 ms CrI: [–287, 112], TVJ: –84 ms CrI: [–369, 204]). But there was an interaction between condition and accuracy (inverse reading in condition C was coded as 'correct') in both regions (sentence final region: 350 ms CrI: [127, 604], TVJ: 475 ms CrI: [258, 710]). We further explored this interaction by looking at the effect of accuracy nested under condition. Although effect sizes were larger, effects were in the same direction as the effects at NP2 and the spillover region: In condition A, reading times were longer when participants chose the incorrect answer in the TVJ task (sentence final region: 392 ms CrI: [35, 801]; TVJ: 645 ms CrI: [291, 1037]). In condition C, reading times were longer when participants chose the inverse interpretation in the TVJ task (sentence final region: 302 ms CrI: [55, 575]; TVJ: 459 ms CrI: [55, 889]). In parallel to the analyses for NP2 and the spillover region, we refit the previous model that included an interaction term for accuracy × condition and coded the inverse reading as incorrect and the surface reading as correct. The interaction turned out to be situated around zero (sentence final region: –65 ms CrI: [–274, 133]; TVJ: –119 ms CrI: [–335, 101]).

3.3 Interaction: reading times in conditions A vs. B × number of inverse readings in condition C

Finally, as for our research question (iii), we tested for an interaction between the reading time differences between conditions A and B (i.e., the increase in reading times in non-canonical sentences) and the number of inverse readings in condition C. Our hypothesis was that participants who have more difficulty processing non-canonical sentences should have more difficulty establishing inverse interpretations. Therefore, we expected a negative interaction, i.e.,

the larger the difference between conditions A and B, the lower the number of inverse readings in condition C. However, we did not find support for such an interaction. In all sentence regions analyzed, the interaction estimates were close to zero or situated around zero, and were, thus, uninformative with respect to the question whether there was an interaction (determiner and adjective of NP1: 0 ms CrI: [–3, 3]; *dann* 'then': 4 ms CrI: [0, 9]; sentence final region: –8 ms CrI: [–45, 29]; TVJ: 30 ms CrI: [–17, 78]).

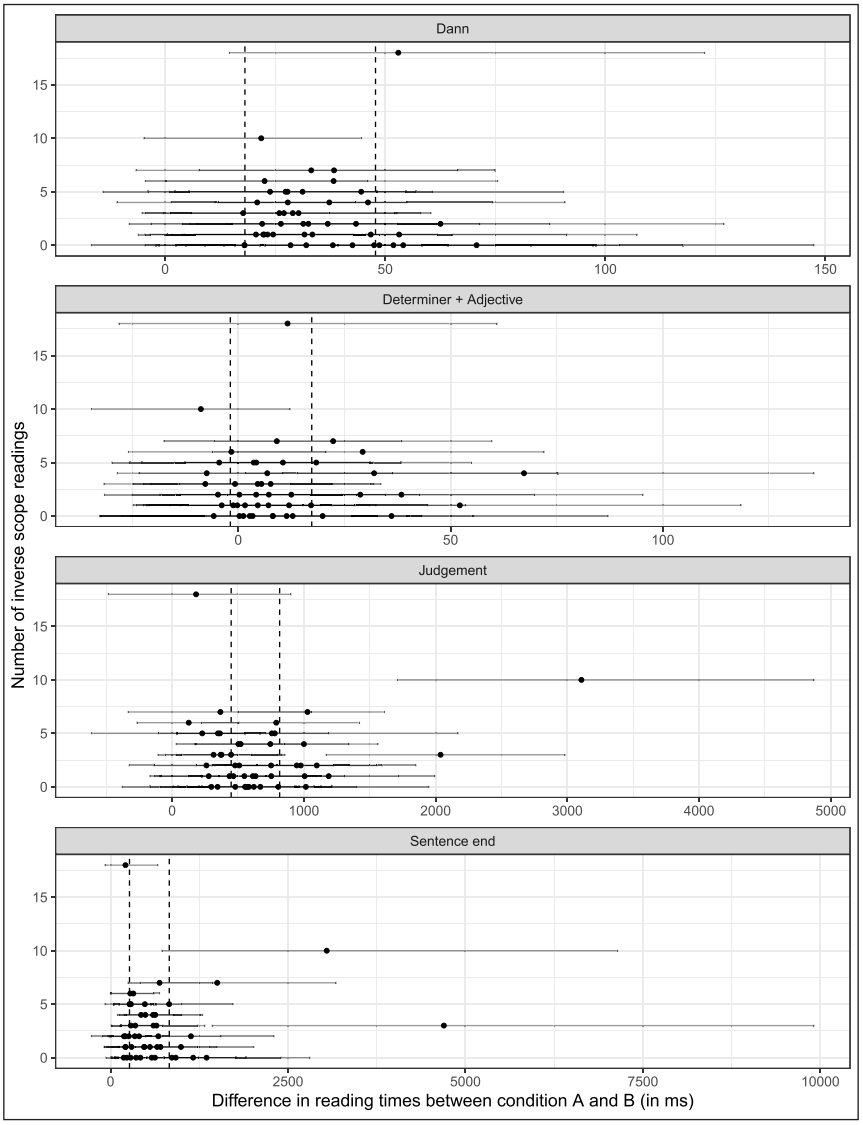

Figure 5: Connection between the number of inverse reading responses in condition C and the difference in reading times between A and B at the determiner and adjective of NP1, at the word *dann* 'then', at the end of the sentence and in the response times of the truth-value-judgement task. Dots represent the estimated mean difference between A and B per participant, and horizontal lines represent the 95% credible intervals around the mean. Dashed vertical lines represent the estimated 95% credible interval of the difference between A and B at the group level.

The results are depicted in **Figure 5**. The x-axis shows the estimated reading time differences between conditions A and B for each participant for the four sentence regions that were analyzed (determiner and adjective of NP1, *dann* 'then', end of the sentence, truth-value-judgement task). The 47 dots represent the estimated mean difference between conditions A and B for each of the 47 participants and the horizontal lines represent the 95% credible intervals around these means. The y-axis shows how often each of the 47 participants accepted the inverse reading, ranging from zero to 18 times. If there was a negative interaction, higher dots should be situated further to the left than lower dots. Yet, this pattern is not visible in **Figure 5**, illustrating that there is no support for the expected interaction.

4. Discussion

4.1 Effects of scrambling

Based on previous research on scrambling in German (e.g., Bornkessel et al., 2002; Pregla et al., 2021), we had predicted that declarative sentences with a clause-initial object should be more difficult to process than declarative sentences with a clause-initial subject. Our findings are in line with this prediction. The first indication of processing difficulty could be seen on the spillover region following the critical NP (NP1), where reading times for object-initial sentences were longer than for the subject-initial sentences. In a previous self-paced listening experiment, the effect of movement difficulties first emerged at NP1 (Pregla, 2023). Our result of an increase in reading times directly after NP1 is consistent with the earlier result, although in our cumulative paradigm, words uncovered by the participant remained visible, unlike in the typical moving window paradigm. That is, although the participants could have delayed processing until the sentence was fully uncovered, processing difficulties were already visible directly after the critical region. This suggests that the participants processed the sentences while uncovering the words.

In addition to the immediate slowdown in reading times at the spillover region, we also found prolonged reading times for object-initial sentences at the end of the sentence and in the TVJ task. Similarly, other studies had also found longer reaction times for object-initial sentences in comprehension tasks (e.g., Pregla et al., 2021; Vogelzang et al., 2019). This suggests that the difficulties in processing scrambled sentences cannot be overcome immediately and persist throughout the sentence. Different from Pregla et al. (2021) and Vogelzang et al. (2019), comprehension accuracy did not decrease in sentences with scrambling, though. This difference might be ascribed to the factor age, since our study included only younger adults, while the earlier studies included elderly participants. It has been found that auditory comprehension of sentences involving movement decreases with age (Wingfield et al., 2003). However, as pointed out by a reviewer, there is another difference between these two studies and ours which could explain this effect, as well. While both former studies were listening studies, our items were

presented written, such that case marking might have been more noticeable to participants. More than that, the cumulative presentation gave participants the option to double-read. This could have made the subsequent task easier and increased accuracy.

Finally, **Figure 5** illustrates the variability in differences between sentences with and without scrambling between participants. In general, all participants' differences are larger than zero at the spillover region (i.e., *dann* 'then'), at the sentence end, and in the TVJ task. This seems to suggest that all participants had some difficulty with the object-initial sentences. However, while many participants' estimates fall within the range of plausible values of the group, a number of participants show effect sizes that are outside the range. In particular, some participants show much larger slowdowns compared to the mean and, thus, they seem to have more difficulty with these sentences than the average. This result is in line with previous studies that observed variability between participants in the difficulty they had with scrambled sentences (Pregla et al., 2021; Vogelzang et al., 2019).

4.2 Effects of scope

Similar to the case of scrambling described above, we also found effects of scope, both in reading times and response times. Even though condition C was not generally associated with longer reading times compared to the baseline condition A, as can be seen in **Figures 2** and **3**, participants who later opted for the inverse reading did, indeed, show both longer reading and response times compared to those participants who opted for the surface reading, as was shown in **Figure 4**. However, while in the case of scrambling, the effect was clearly visible already on the spillover region immediately after the critical NP, this was less pronounced in the case of scope. Even though there was a slight increase already on the second NP and the spillover region, the effect in reading times became most apparent on the final region of the sentence. These longer reading and response times are in line with previous assumptions about inverse scope being associated with additional processing costs (Anderson, 2004; Blok, 2019; Wurmbrand, 2018; a.o.). The fact that these effects become most apparent by the very end of the sentence and in the judgment task could indicate that scope ambiguities in German are not resolved immediately, as was shown in previous experiments on inverse scope in English (see Clark & Kar, 2011; Dwivedi et al., 2010; Raffray & Pickering, 2010; Sanford & Sturt, 2002; Villalta, 2003; Zhou & Gao, 2009). This is, thus, additional evidence against older theories of scope that suggest immediate resolution of ambiguities, such as Crain and Steedman (1985) and Altmann and Steedman (1988). Nevertheless, considering the slight increase already starting on the second NP, it also seems too strong to claim that scope ambiguities are left underspecified and only resolved if necessary. Otherwise, we would expect an effect *only* on the response times and not also earlier in the sentence. Instead, we obtained a more nuanced picture, in that scope resolution is initiated at an early stage and continues until the end of the sentence. Our experiment is in line with an early SPR-experiment

on quantifier scope in German (Bott & Schlotterbeck, 2015), where scope resolution started immediately on the second NP, but only in case the main predicate of the sentence had already been encountered. Just as in our study, the longer reading times continued and amplified until the end of the sentence. Note, though, that in the study by Bott and Schlotterbeck, the items were different from ours, in that the word order was non-canonical OS, suggesting a reconstruction effect rather than QR.[13] The choice of quantifiers differed too, in that the first NP was a partitive construction with a universal and the second NP was a modified numeral (*genau ein* 'exactly one'); see (5). Further, there was only one more phrase after the critical item – the second NP – occurred and before the sentence ended.

(5) *German*
Jeden [seiner/dieser] Schüler lobte genau ein Lehrer voller
each [of-his/of-these] pupils.ACC praised exactly one teacher.NOM full-of
Wohlwollen.
goodwill.
'Exactly one teacher praised each of his/these pupils full of goodwill.'
Bott & Schlotterbeck (2015, p. 64)

Furthermore, some conditions contained a bound possessive pronoun. Finally, Bott and Schlotterbeck employed the moving window SPR paradigm, in contrast to our cumulative paradigm. It is, therefore, noteworthy that, despite those differences in the design of the items and the procedure, the processing results are so similar.

Looking at **Figure 4**, we see that participants who respond incorrectly in the judgment task in the baseline condition A already show an effect of longer reading times in the sentence. This indicates that participants did not merely give an incorrect response because they could not remember the sentence anymore, but because they already had problems processing the sentence while reading. In fact, the patterns of inverse scope judgments in condition C and incorrect judgments in condition A are strikingly similar. This may lead to the conclusion that we are not dealing with inverse interpretations at all, but simply with cases of misinterpretations, i.e., participants responding incorrectly. However, there are several reasons why this is not plausible. As described above, we obtained an inverse response ~15% of the time in condition C and an incorrect response ~5% of the time in condition A. If we assume that the 15% are merely incorrect responses, we have to explain why, in condition C, there are three times as many incorrect responses as in A. There is no difference in structural complexity between the two sentences. In fact, they are fully identical apart from the determiner on the first NP, which is definite in A

[13] Under reconstruction, an overtly moved NP is interpreted in its base position, which is in a position below another NP. This way, an inverse reading can arise without the need for QR. German is known to allow for inverse readings under reconstruction quite readily (Frey, 1993).

and indefinite in C. If we assume that the longer reading times in C have nothing to do with the interaction of the two quantifiers, then we can only attribute it to the lexical difference between the definite determiner in condition A and the indefinite determiner in condition B. But it seems implausible that the difference between those two determiners would cause a 3x-increase in incorrect responses. Even though multiple studies have reported that the definite determiner results in lower processing costs compared to the indefinite determiner, this is only the case if it refers back to an already introduced discourse referent (Irwin et al., 1982; Kirsten et al., 2014; Murphy, 1984). If the uniqueness presupposition of the definite determiner is violated, it causes processing costs in the same way as when an indefinite determiner is used despite the uniqueness of the referent (Bade & Schwarz, 2019). As for the items in our experiment, the uniqueness requirement of the definite determiner is not met, due to the use of a bare plural in the context. This potentially requires participants to accommodate in order to render its occurrence in this context felicitous, thereby making condition A with the definite determiner more difficult; see also the discussion in 4.4 below. Because the referents have already been introduced in the context, there is no additional cost that would be associated with the introduction of a new discourse referent, as might otherwise be the case with the indefinite determiner. Thus, condition C with the indefinite determiner should not be any more difficult than condition A with the definite determiner. If no scope interaction took place at all, we would expect at least the same accuracy in condition C, compared to condition A. Moreover, despite the more difficult sentence structure in the scrambled condition B, which also caused processing difficulties for participants, accuracy was not any lower than in the non-scrambled condition A. Thus, even if it was the case that condition C was more difficult to process than condition A, it is still unclear why that would be reflected by a dramatic 3-fold increase of incorrect answers, as we see no such increase in the scrambled condition B. If we assume that, in condition C, just as in A and B, 5% of the responses are truly incorrect, then we are still left with 10% of the cases which can only be interpreted as true inverse readings. We therefore attribute the similar patterns of inverse and incorrect responses to more general processing difficulties in both cases.

The overall percentage of inverse scope interpretations in the study at hand is lower than in comparable experiments on scope in German with the same type of sentence construction (Fanselow et al., 2022; Philipp & Zimmermann, 2020, 2023), where the acceptance across studies was 25–39%. We see two reasons for this effect. First, in all but one of the studies presented in the previous literature, the abbreviated form `*n* of the German indefinite *ein* was used. This was done because the full indefinite in German can also receive a numeral or a specific interpretation, both of which make the surface interpretation more likely. In fact, in the one study where the full indefinite was used (Philipp & Zimmermann, 2023), the acceptance of the inverse reading was clearly lower (25%) than in the other three studies (32%, 34%, 39%). If this 25% is also adjusted in the same way as we did with our own results above – by taking into account the percentage of incorrect responses in unambiguous conditions, as reported in the respective study (~5–10%)

– then we arrive at 15–20% of true inverse readings, which is much closer to the value of 10% we found in our own study. Finally, it is fair to assume that an SPR procedure is more difficult for participants than reading the sentences in a natural way. Additionally, in our study, participants had to perform the follow-up judgment task when the sentence had already disappeared, while in the studies mentioned above, the sentence could be read again when the question or task appeared. As inverse scope interpretations require a higher mental cost, it seems plausible to assume that additional factors that make the task more difficult, like SPR, or having to keep the sentence in memory, increase the burden on the participant and, consequently, reduce inverse scope interpretations.

Similar to previous experiments on scope in German (Fanselow et al., 2022; Philipp & Zimmermann, 2020, 2023) we found variability between speakers in the extent to which they accept inverse scope readings. While a number of participants consistently rejected the inverse scope interpretation, the majority of participants accepted them 5–35% of the time, as shown in **Figure 6**. Two outliers accepted the inverse scope reading unusually often, 50% and 90% of the time, respectively. However, because the overall acceptance of the inverse reading was much lower than in comparable studies, the variability we found is also less pronounced (floor effect). Nevertheless, it can be seen that the inverse responses were not mainly driven by a few outliers, but by a majority of participants, who gave them in a small portion of the items.

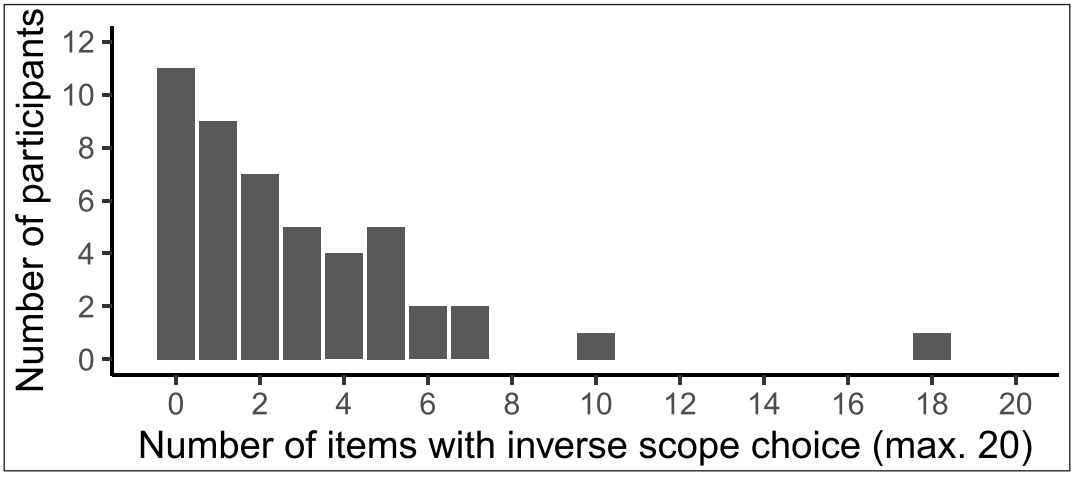

Figure 6: Participant distribution with respect to number of inverse scope choices.

4.3 Interaction between scrambling and scope

As shown in **Figure 5**, we did not find an interaction between the number of inverse responses and the difference in reading times between conditions A and B. Thus, our study does not provide support for the hypothesis that participants who have difficulty with processing non-canonical word order also have difficulty with inverse readings, and conversely, that participants who

have little difficulty with non-canonical word orders also find it easier to obtain the inverse reading. The results of this study do not give any indication that the difficulties associated with the two phenomena, and the variability found in both of them, share a common underlying source.

Even though we cannot draw conclusions from the lack of an effect in our study in favour of the null hypothesis, we still want to speculate a bit on the possibility that the two phenomena investigated here do *not,* in fact, share a common underlying source and that the similar patterns of processing difficulty in participants arise from two unrelated factors. This could either be because we are dealing with different types of movement (e.g., A- vs. A'-movement, overt vs. covert movement), or because one or both phenomena do not actually involve syntactic movement at all, contrary to what is often assumed; see also Section 1. One indication in favour of the assumption that scrambling and quantifier scope do not, in fact, share the same underlying source is provided by Varkanitsa et al. (2016), who show that participants with the acquired language disorder aphasia have *less* difficulty with the inverse reading than language unimpaired participants. Non-canonical word order, on the other hand, causes *more* difficulties in participants with aphasia than in unimpaired participants (Caplan et al., 2013, 2015; Hanne et al., 2011; Pregla et al., 2021). This pattern is unexpected if the underlying source is the same. If it was, we would expect participants with aphasia to exhibit the same pattern in both cases. Instead, we see that they diverge.

Another possibility would be that the two topics are, in fact, related but we were unable to find this relationship, possibly due to the measure we used for processing difficulty in the quantifier scope condition. We reasoned that the number of inverse readings would reflect the difficulty that a person has with covert movement. Our assumption was that a person who accepts a high number of inverse readings would have little difficulty with covert movement, whereas a person who accepts a low number of inverse readings would have more difficulty with covert movement. However, processing difficulty with covert movement might be better captured by the difference in reading times between surface reading and inverse reading. To study the connection between covert and overt movement, one would then have to compare the difference in reading times between the surface and the inverse readings and between scrambling and no scrambling. We did not carry out this analysis, because of the large group of participants who accepted only a very small number of inverse readings. For these participants, it would be difficult to get a representative estimate of the difference in reading times between surface and inverse readings. To get an estimate for this difference, one would, therefore, need more participants who accept a large number of inverse readings. The acceptance rate could, for example, be increased by using sentences that are biased towards the inverse reading. Nevertheless, this measure would only provide the relevant results under the assumption that the observed differences between participants are due to differences in processing difficulty. Instead, it could also be that participants all experience the same level of difficulty with inverse

readings, but with some participants, the parser is more "willing" to accept higher processing costs. In another study, it might, thus, be useful to analyse both the processing cost for inverse readings and the number of inverse readings.

4.4 Potential confounds

Our experiment contained two potential confounds that we are going to address in the following, showing that they do not undermine the main findings of our study.

4.4.1 Accommodation

As discussed in 4.2 above, the definite determiner carries two presuppositions, existence and uniqueness. In conditions A and B, where the definite determiner occurs, the uniqueness presupposition is partly violated. This can be seen in (6). In condition A, we refer to a unique wide canal in the target sentence after introducing a plurality of canals in the context. The same happens in (6b) for condition B, where we refer to a unique wide field after introducing a plurality of fields in the context. Even though the effect of this might be attenuated by the adjective *breit* ('wide'), referring to the only *wide* canal or field among all the other canals and fields, participants might still show an accommodation effect, because a unique antecedent has not been established in the context already. This may lead to additional processing costs, due to the accommodation procedure. At the same time, we would expect such an accommodation effect only on, or shortly after, the first definite NP. A potential effect of scope disambiguation in condition C, compared to the baseline condition A, should, therefore, still be visible on or after the second NP, which is exactly what we found. As for the effect of word order in the difference between conditions A and B, this should also be unaffected by accommodation, since this would apply equally in both conditions. The reading time data also does not show any sign of processing difficulties due to accommodation, as they do not differ between condition A with the definite determiner and condition C with the indefinite determiner on the first NP or the region after (see **Figure 3**).

(6) *German*

a. Condition A:
Der Agrarexperte hatte empfohlen, dass die Äcker durch **[Kanäle]** bewässert werden sollten. Und tatsächlich bewässerte **[der breite Kanal]** dann jeden Acker in der Region der Bauern.
'The agriculture expert had advised that the fields be irrigated by **[canals]**. And then, indeed, **[the wide canal]** irrigated every field in the region of the farmers.'

b. Condition B:
Der Agrarexperte hatte empfohlen, dass **[die Äcker]** durch Kanäle bewässert werden sollten. Und tatsächlich bewässerte **[den breiten Acker]** dann jeder Kanal in der Region der Bauern.

‘The agriculture expert had advised that **[the fields]** be irrigated by canals. And then, indeed, every canal irrigated **[the wide field]** in the region of the farmers.’

4.4.2 Cumulative presentation

The second potential confound is the cumulative presentation, which we employed in order to reduce the mental load for participants. A danger with this method is that participants uncover the whole sentence before actually reading and processing it, which would hide all potential on-line effects. Indeed, we can see that the reading times were very long on the final region. However, as we found on-line effects for both the scrambling and the scope phenomenon, these longer reading times at the final region seem to be an indicator of participants merely re-reading the sentence. They still seem to read and process the sentence while uncovering it. It is still possible, though, that the effects might be more pronounced in a non-cumulative presentation. It could, therefore, be interesting to see the outcome of the same study employing a different on-line method. One option would be eye-tracking, where the complete sentence can be shown, but different reading time measures still allow one to tease apart immediate processing from reanalysis.

5. Conclusion

In this article, we presented two seemingly unrelated phenomena, (i) scrambling and (ii) quantifier scope. We showed, based on previous literature, that both of them show similar processing patterns: both inverse scope readings and scrambling are associated with increased processing costs. Nevertheless, in both cases, there is great variability between participants, with some having little to no difficulty and others having much more difficulty. Both phenomena have been argued in the literature to involve some type of movement. We hypothesized that the processing difficulties may be caused by a shared underlying problem, e.g., by a general difficulty with or resistance to optional syntactic movement, which is stronger in some participants than in others. We conducted a cumulative self-paced reading experiment, where we manipulated the factors word order and presence of double-quantification. We measured accuracy, reading times, and response times to determine if participants with lower reading times for scrambled word order show a higher number of inverse interpretations. As such, this was the first study to investigate this type of question. We found the expected effect of scrambling after the first NP (clause-initial object) and we found the expected effect of inverse scope readings on the second NP and subsequent regions. Our results, thus, provide additional support for previous findings on these two topics. They also indicate that the cumulative paradigm that we employed is well suited to test these types of questions and can, thus, provide a viable methodological alternative in cases where another paradigm might be too demanding for participants. Although effects

were most pronounced at the end of the sentence and in the response times, we still observed on-line effects at the critical regions, too. Contrary to our initial hypothesis, we did not find a connection between the two phenomena when comparing the increase in reading times in the case of scrambling to the number of inverse interpretations for each participant. We discussed the possibility that the respective processing difficulties simply do not share any underlying source. This might be supported by results from independent studies, such as Varkanitsa et al. (2016). At the same time, we also discussed the second possibility that the specific design and measurements taken in our study might not have been appropriate for uncovering the hypothesized connection and suggested what improvements would allow future experiments to target this question in a better way.

Abbreviations

ACC = accusative, NOM = nominative, POSS = possessive

Data accessibility statement

Materials, data, and analysis of the experiment can be found at https://osf.io/3wk4z/.

Ethics and consent

The experiment was conducted within project B02 and C02 of the Collaborative Research Council 1287: The limits of variability in language. Within the overall proposal, there was no need to obtain individual clearing of sub-experiments from the ethics committee and the project was wholesale approved without an ethics statement by the DFG as proposed. In carrying out the experiment, we adhered to standard practices regarding data protection (in compliance with UP rules on data protection) and experimental ethics.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 317633489 – SFB 1287, Project B02 & C02. We would further like to express our gratitude to everyone who participated in our experiment. A special thanks goes to Malte Zimmermann, who gave valuable feedback on this article.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Mareike Philipp: Conzeptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. Dorothea Pregla: Conzeptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft.

References

Abels, K. (2015). Word order. In A. Alexiadou & T. Kiss. (Eds.), *Syntax – Theory and analysis: An international handbook* (pp. 1400–1445). Mouton de Gruyter. https://doi.org/10.1515/9783110363708-017

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity* [Doctoral dissertation]. Northwestern University.

Aoun, J., & Li, Y. A. (1993). *Syntax of scope*. MIT Press.

Bade, N., & Schwarz, F. (2019). (In-)definites, (anti-)uniqueness, and uniqueness expectations. In A. Goel, C. Seifert & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the Cognitive Science Society* (pp. 119–125). Cognitive Science Society.

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics, 46*, 273–330. https://doi.org/10.1017/S0022226709990260

Bader, M., & Portele, Y. (2021). Discourse and form constraints on licensing object-first sentences in German. *Languages, 6,* 82. https://doi.org/10.3390/languages6020082

Barker, C. (2005). Remark on Jacobson 1999: Crossover as a local constraint. *Linguistics and Philosophy*, *28*, 447–472. https://doi.org/10.1007/s10988-004-5327-1

Barker, C. (2012). Quantificational binding does not require c-command. *Linguistic Inquiry, 43*(4), 614–633. https://doi.org/10.1162/ling_a_00108

Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of *each* and *every*. In A. Szabolcsi (Ed.), *Ways of scope taking* (pp. 71–98). Springer. https://doi.org/10.1007/978-94-011-5814-5_3

Blok, D. (2019). *Scope oddity: On the semantic and pragmatic interactions of modified numerals, negative indefinites, focus operators, and modals* [Doctoral dissertation]. Utrecht University.

Bobaljik, J. D., & Wurmbrand, S. (2012). Word order and scope: Transparent interfaces and the 3/4 signature. *Linguistic Inquiry, 43,* 371–421. https://doi.org/10.1162/LING_a_00094

Bornkessel, I., & Schlesewsky, M. (2006). The role of contrast in the local licensing of scrambling in German: Evidence from online comprehension. *Journal of Germanic Linguistics, 18*(1), 1–43. https://doi.org/10.1017/S1470542706000018

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: Evidence from the online processing of flexible word order. *Cognition, 85*(2), B21–B30. https://doi.org/10.1016/S0010-0277(02)00076-8

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2014). Competition in argument interpretation: Evidence from the neurobiology of language. In B. MacWhinney, A. Malchukov & E. Moravcsik (Eds.), *Competing motivations in grammar and usage* (pp. 107–126). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709848.003.0007

Bott, O., & Radó, J. (2007). Quantifying quantifier scope: A cross-methodological comparison. In S. Featherston & W. Sternefeld (Eds.), *Roots. Linguistics in search of its evidential base* (pp. 53–74). Mouton de Gruyter. https://doi.org/10.1515/9783110198621.53

Bott, O., & Radó, J. (2009). How to provide exactly one interpretation for every sentence, or what eye movements reveal about quantifier scope. In S. Featherston & S. Winkler (Eds.), *The fruits of empirical linguistics* (Vol. 1, pp. 25–46). Mouton de Gruyter. https://doi.org/10.1515/9783110216141.25

Bott, O., & Schlotterbeck, F. (2012). Incremental truth value judgments. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory* (pp. 3–28). Mouton de Gruyter. https://doi.org/10.1515/9781614510888.3

Bott, O., & Schlotterbeck, F. (2015). The processing domain of scope interaction. *Journal of Semantics, 32*(1), 39–92. https://doi.org/10.1093/jos/fft015

Brasoveanu, A., & Dotlačil, J. (2015). Strategies for scope taking. *Natural Language Semantics, 23,* 1–19. https://doi.org/10.1007/s11050-014-9109-1

Büring, D. (1997). The great scope inversion conspiracy. *Linguistics and Philosophy, 20*(2), 175–194. https://doi.org/10.1023/A:1005397026866

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Caplan, D., DeDe, G., Waters, G. S., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging, 26*(2), 439–450. https://doi.org/10.1037/a0021837

Caplan, D., Michaud, J., & Hufford, R. (2013). Dissociations and associations of performance in syntactic comprehension in aphasia and their implications for the nature of aphasic deficits. *Brain and Language, 127,* 21–33. https://doi.org/10.1016/j.bandl.2013.07.007

Caplan, D., Michaud, J., & Hufford, R. (2015). Mechanisms underlying syntactic comprehension deficits in vascular aphasia: New evidence from self-paced listening. *Cognitive Neuropsychology, 32,* 283–313. https://doi.org/10.1080/02643294.2015.1058253

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*(1), 77–94. https://doi.org/10.1017/S0140525X99001788

Chomsky, N. (1995). *The minimalist program*. MIT Press.

Chomsky, N. (2013). Problems of projection. *Lingua, 130,* 33–49. https://doi.org/10.1016/j.lingua.2012.12.003

Clahsen, H., & Featherston, S. (1999). Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research, 28,* 415–437. https://doi.org/10.1023/A:1023293132656

Clark, D. G., & Kar, J. (2011). Bias of quantifier scope interpretation is attenuated in normal aging and sematic dementia. *Journal of Neurolinguistics, 24,* 401–419. https://doi.org/10.1016/j.jneuroling.2011.01.002

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 320–358). Cambridge University Press. https://doi.org/10.1017/CBO9780511597855.011

Dwivedi, V. D., Phillips, N. A., Einagel, S., & Baum, S. R. (2010). The neural underpinnings of semantic ambiguity and anaphora. *Brain Research, 1311*, 93–109. https://doi.org/10.1016/j.brainres.2009.09.102

Fanselow, G., Zimmermann, M., & Philipp, M. (2022). Assessing the availability of inverse scope in German in the covered box paradigm. *Glossa: A Journal of General Linguistics, 7*(1), 1–24. https://doi.org/10.16995/glossa.5766

Fox, D. (2000). *Economy and semantic interpretation*. MIT Press.

Frey, W. (1993). *Syntaktische Bedingungen für die semantische Interpretation: über Bindung, implizite Argumente und Skopus*. Akademie Verlag.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex, 16*(12), 1709–1717. https://doi.org/10.1093/cercor/bhj106

Geyken, A., Barbaresi, A., Didakowski, J., Jurish, B., Wiegand, F., & Lemnitzer, L. (2017). Die Korpusplattform des "Digitalen Wörterbuchs der deutschen Sprache" (DWDS). *Zeitschrift für germanistische Linguistik, 45,* 327–344. https://doi.org/10.1515/zgl-2017-0017

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). MIT Press. https://doi.org/10.7551/mitpress/3654.003.0008

Gil, D. (1982). Quantifier scope, linguistic variation, and natural language semantics. *Linguistics and Philosophy, 5,* 421–472. https://doi.org/10.1007/BF00355582

Gillen, K. (1991). *The comprehension of doubly quantified sentences* [Doctoral dissertation]. University of Durham.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., Von Cramon, D. Y., & Schlesewsky, M. (2005). The emergence of the unmarked: A new perspective on the language-specific function of Broca's area. *Human Brain Mapping, 26,* 178–190. https://doi.org/10.1002/hbm.20154

Haider, H. (2017). 'Mittelfeld' phenomena: Scrambling in Germanic. In H. van Riemsdijk & M Everaert (Eds.), *The Wiley Blackwell companion to syntax* (pp. 204–274). Blackwell. https://doi.org/10.1002/9781118358733.wbsyncom048

Haider, H., & Rosengren, I. (2003). Scrambling: Nontriggered chain formation in OV languages. *Journal of Germanic Linguistics, 15*(3), 203–267. https://doi.org/10.1017/S1470542703000291

Hanne, S., Sekerina, O. A., Vasishth, S., Burchert, F., & De Bleser, R. (2011). Chance in agrammatic sentence comprehension: What does it really mean? Evidence from eye movements of German agrammatic aphasic patients. *Aphasiology, 25,* 221–244. https://doi.org/10.1080/02687038.2010.489256

Heck, F., & Himmelreich, A. (2017). Opaque intervention. *Linguistic Inquiry, 48,* 47–97. https://doi.org/10.1162/LING_a_00235

Hendriks, H. (1988). Type change in semantics: The scope of quantification and coordination. In E. Klein & J. van Benthem (Eds.), *Categories, polymorphism and unification* (pp. 96–119). ITLI.

Hendriks, H. (1993). *Studies flexibility. Categories and types in syntax and semantics* [Doctoral dissertation]. University of Amsterdam.

Huang, C.-T. J. (1995). Logical Form. In G. Webelhuth (Ed.), *Government and Binding Theory and the Minimalist Program* (pp. 125–175). Blackwell.

Hulsey, S., & Sauerland, U. (2006). Sorting out relative clauses. *Natural Language Semantics, 14,* 111–137. https://doi.org/10.1007/s11050-005-3799-3

Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (Ed.), *Syntax and semantics* (pp. 37–58). New York: Academic Press. https://doi.org/10.1163/9789004368828_003

Irwin, D., Bock, J. K., & Stanovich, K. E. (1982). Effects of information structure cues on visual word processing. *Journal of Verbal Learning and Verbal Behavior, 21,* 307–325. https://doi.org/10.1016/S0022-5371(82)90637-5

Jacobson, P. (1992). Raising without movement. In R. K. Larson, S. Iatridou, U. Lahiri & J. Higginbotham (Eds.), *Control and grammar* (pp. 149–194). Springer. https://doi.org/10.1007/978-94-015-7959-9_5

Jacobson, P. (1996). The locality of interpretation: The case of binding and coordination. In T. Galloway & J. Spencer (Eds.), *Semantics and Linguistic Theory VI* (pp. 111–135). Cornell University. https://doi.org/10.3765/salt.v6i0.2771

Kayne, R. (1994). *The antisymmetry of syntax*. MIT Press.

Keller, F. (2000). Evaluating competition-based models of word order. In L. R. Gleitman & A. K. Joshi (Eds.), *22nd Annual Conference of the Cognitive Science Society* (pp. 747–752). Lawrence Erlbaum.

Kemper, S., & Liu, C. (2007). Eye movements of young and older adults during reading. *Psychology and Aging, 22*(1), 84–93. https://doi.org/10.1037/0882-7974.22.1.84

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

Kirsten, M., Tiemann, S., Seibold, V. C., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience, 29*(9), 1147–1162. https://doi.org/10.1080/23273798.2014.899378

Krifka, M. (1998). Scope inversion under the rise-fall pattern in German. *Linguistic Inquiry, 29,* 75–112. https://doi.org/10.1162/002438998553662

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition, 48,* 243–279. https://doi.org/10.1016/0010-0277(93)90042-T

Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Narr.

May, R. (1977). *The grammar of quantification* [Doctoral dissertation]. Massachusetts Institute of Technology.

May, R. (1985). *Logical Form: Its structure and derivation*. MIT Press.

Meng, M., Bader, M., & Bayer, J. (1999). Die Verarbeitung von Subjekt–Objekt-Ambiguitäten im Kontext. In I. Wachsmuth & B. Jung (Eds.), *Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft* (pp. 244–249). Infix Verlag.

Murphy, G. L. (1984). Establishing and accessing referents in discourse. *Memory & Cognition, 12,* 489–497. https://doi.org/10.3758/BF03198311

Pafel, J. (2005). *Quantifier scope in German*. John Benjamins. https://doi.org/10.1075/la.84

Partee, B. (1991). Topic, focus and quantification. In A. Wyner & S. Moore (Eds.), *Proceedings of the First Annual Conference on Semantics and Linguistic Theory* (pp. 159–187). Cornell University. https://doi.org/10.3765/salt.v1i0.2918

Pechmann, T., Uszkoreit, H., Engelkamp, J., & Zerbst, D. (1996). Wortstellung im deutschen Mittelfeld. Linguistische Theorie und psycholinguistische Evidenz. In C. Habel, S. Kanngießer & G. Rickheit (Eds.), *Perspektiven der Kognitiven Linguistik. Modelle und Methoden* (pp. 257–299). Westdeutscher Verlag. https://doi.org/10.1007/978-3-663-07678-0_11

Philipp, M., & Zimmermann, M. (2020). Empirical investigations on quantifier scope ambiguities in German. In M. Franke, N. Kompa, M. Liu, J. Mueller & J. Schwab (Eds.), *Proceedings of Sinn und Bedeutung 24* (Vol. 2, pp. 145–164).

Philipp, M., & Zimmermann, M. (2023). On trees blocking roads and cameras recording burglars: An experimental comparison of the availability of inverse scope in English and German. *Linguistic Inquiry*. https://doi.org/10.1162/ling_a_00493

Pregla, D. (2023). *Variability in sentence processing performance in German people with aphasia and unimpaired German native speakers* [Doctoral dissertation]. University of Potsdam.

Pregla, D., Lissón, P., Vasishth, S., Burchert, F., & Stadie, N. (2021). Variability in sentence comprehension in aphasia in German. *Brain and Language, 222,* 105008. https://doi.org/10.1016/j.bandl.2021.105008

Pregla, D., Vasishth, S., Lissón, P., Stadie, N., & Burchert, F. (2022). Can the resource reduction hypothesis explain sentence processing in aphasia? A visual world study in German. *Brain and Language*, *235*, 105204. https://doi.org/10.1016/j.bandl.2022.105204

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Radó, J., & Bott, O. (2012). Underspecified representations of scope ambiguity? In M. Aloni, V. Kimmelmann, F. Roelofsen, G. Sassoon, K. Schulz & M. Westera (Eds.), *Logic, language and meaning: 18th Amsterdam Colloquium* (pp. 180–189). Springer. https://doi.org/10.1007/978-3-642-31482-7_19

Radó, J., & Bott, O. (2018). What do speaker judgments tell us about theories of quantifier scope in German? *Glossa: a journal of general linguistics, 3*(1), 1–29. https://doi.org/10.5334/gjgl.309

Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension? *Psychological Science, 21,* 1090–1097. https://doi.org/10.1177/0956797610375446

Reinhart, T. (2006). *Interface strategies*. MIT Press. https://doi.org/10.7551/mitpress/3846.001.0001

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *NeuroImage, 15,* 1003–1014. https://doi.org/10.1006/nimg.2001.1026

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language, 38,* 150–176. https://doi.org/10.1006/jmla.1997.2551

Saba, W. S., & Corriveau, J.-P. (2001). Plausible reasoning and the resolution of quantifier scope ambiguities. *Studia Logica, 67,* 271–289. https://doi.org/10.1023/A:1010503321412

Sabel, J. (2005). String-vacuous scrambling and the Effect on Output Condition. In J. Sabel & M. Saito (Eds.), *The free word order phenomenon* (pp. 281–334). Mouton de Gruyter. https://doi.org/10.1515/9783110197266.281

Salzmann, M. (to appear). Word order in the German middle field – Scrambling. In K. Hartmann, J. Mursell & S. Wurmbrand (Eds.), *Handbook of Germanic syntax*. De Gruyter.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *TRENDS in Cognitive Science, 6*(9), 382–386. https://doi.org/10.1016/S1364-6613(02)01958-7

Schlesewsky, M., Bornkessel, I., & Frisch, S. (2003). The neurophysiological basis of word order variations in German. *Brain and Language, 86,* 116–128. https://doi.org/10.1016/S0093-934X(02)00540-0

Schunack, S. (2016). *Processing of non-canonical word orders in an L2: When small changes make no big difference* [Doctoral dissertation]. University of Potsdam.

Scontras, G., Polinsky, M., Tsai, C.-Y. E., & Mai, K. (2017). Cross-linguistic scope ambiguity: When two systems meet. *Glossa: a journal of general linguistics, 2*(1), 1–28. https://doi.org/10.5334/gjgl.198

Shan, C.-C., & Barker, C. (2006). Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy, 29*(1), 91–134. https://doi.org/10.1007/s10988-005-6580-7

Skopeteas, S., & Fanselow, G. (2009). Effects of givenness and constraint on free word order. In M. Zimmermann & C. Féry (Eds.), *Information structure: Theoretical, typological, and experimental perspectives* (pp. 307–331). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570959.003.0013

Stechow, A., & Sternefeld, W. (1988). *Bausteine syntaktischen Wissens: Ein Lehrbuch der generativen Grammatik*. Westdeutscher Verlag.

Steedman, M. (2012). *Taking scope. The natural semantics of quantifiers*. MIT Press. https://doi.org/10.7551/mitpress/9780262017077.001.0001

Surányi, B., & Turi, G. (2017). Focus and quantifier scope: An experimental study of Hungarian. In H. van der Hulst & A. Lipták (Eds.), *Approaches to Hungarian: Papers from the 2015 Leiden conference* (Vol. 15, pp. 209–238). John Benjamins. https://doi.org/10.1075/atoh.15.08sur

Szabolcsi, A. (2012). Strategies for scope taking. In A. Szabolcsi (Ed.), *Ways of scope taking*. Springer.

Tanaka, M. (2015). *Scoping out of adjuncts: Evidence for the parallelism between QR and wh-movement* [Doctoral dissertation]. University College London.

Tunstall, S. (1998). *The interpretation of quantifiers* [Doctoral dissertation]. University of Massachusetts.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

VanLehn, K. (1978). Determining the scope of English quantifiers. *Technical Report AI-TR 483*. Artificical Intelligence Laboratory, MIT.

Varkanitsa, M., Kasselimis, D., Fugard, A. J. B., Evdokimidis, I., Druks, J., Potagas, C., & Van de Koot, H. (2016). Syntactic predictions and asyntactic comprehension in aphasia: Evidence from scope relations. *Journal of Neurolinguistics, 40,* 15–36. https://doi.org/10.1016/j.jneuroling.2016.04.001

Villalta, E. (2003). The role of context in the resolution of quantifier scope ambiguities. *Journal of Semantics, 20,* 115–162. https://doi.org/10.1093/jos/20.2.115

Vogelzang, M., Thiel, C. M., Rosemann, S., Rieger, J., & Ruigendijk, E. (2019). Cognitive abilities to explain individual variation in the interpretation of complex sentences by older adults. In A. Goel, C. Seifert & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the Cognitive Science Society* (pp. 3036–3042). Cognitive Science Society.

Vogelzang, M., Thiel, C. M., Rosemann, S., Rieger, J., & Ruigendijk, E. (2020). Neural mechanisms underlying the processing of complex sentences: An fMRI study. *Neurobiology of Language, 1,* 226–248. https://doi.org/10.1162/nol_a_00011

Wingfield, A., Peelle, J. E., & Grossmann, M. (2003). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *A Journal on Normal and Dysfunctional Development, 10*(4), 310–322. https://doi.org/10.1076/anec.10.4.310.28974

Wurmbrand, S. (2018). The cost of raising quantifiers. *Glossa: a journal of general linguistics, 3*(1), 1–40. https://doi.org/10.5334/gjgl.329

Zhou, P., & Gao, L. (2009). Scope processing in Chinese. *Journal of Psycholinguistic Research, 38,* 11–24. https://doi.org/10.1007/s10936-008-9079-x