Anti-Locality Effect Without Head-Final Dependencies

Juliane Schwab[a], Ming Xiang[b], and Mingya Liu[c]

[a] Institute of Cognitive Science, Osnabrück University

[b] Department of Linguistics, University of Chicago

[c] Department of English and American Studies, Humboldt University of Berlin

Author note

Juliane Schwab, https://orcid.org/0000-0001-7899-5974

Ming Xiang, https://orcid.org/0000-0002-6027-8704

Mingya Liu, https://orcid.org/0000-0001-9426-3669

We have no conflicts of interest to disclose.

Data Availability: All materials, code and data associated with the study are available in the following repository: https://osf.io/h3qa8/

Correspondence concerning this article should be addressed to Juliane Schwab, Institute of Cognitive Science, Osnabrück University, Wachsbleiche 27, 49090 Osnabrück, Germany

Abstract

Anti-locality effects provide strong evidence for expectation-based sentence parsing models. Previous discussion of the anti-locality effect, however, largely focused on the argument-verb dependencies in verb-final constructions, for which a memory retrieval-based account has been argued to be equally adequate. To test whether the principles of expectation/memory-based accounts hold for a wider range of dependencies, we report on two self-paced reading experiments that compared two different determiners in German: the morphologically complex determiner *derjenige* 'the-jenig,' which obligatorily requires a relative clause, and the bare determiner *der* 'the,' which does not trigger such expectations. The first experiment did not show the expected anti-locality effect, but the reliability of our results was restricted by the experiment's low statistical power. In a large-scale second experiment we addressed confounds in the design of experiment 1 and found evidence for an anti-locality effect with the complex determiner. As the anti-locality effect found in our study does not involve argument-verb dependencies, the memory-based account cannot be extended to the current case. Thus, our findings provide novel empirical support for the expectation-based anti-locality effect. At the same time, the experiment attests to a processing cost in later sentence regions, hinting that memory- and expectation-based effects can co-occur within the same structure.

Keywords: anti-locality effect, syntactic parsing, determiner, relative clause, German

Anti-Locality Effect Without Head-Final Dependencies

A growing body of work has shown that humans form expectations about upcoming information when processing linguistic input. Expectation-based models of sentence comprehension (e.g., Hale, 2001; Levy, 2008) suggest that the processing cost of a word in a sentence is a function of its expectancy given the prior linguistic context. An equally large body of research has suggested that sentence processing is sensitive to memory constraints. Specifically, it has been shown that integrating syntactic constituents can be negatively affected by increased distance between dependents (for an overview on locality effects, see Gibson, 1998; see also Balling & Kizach, 2017; Chen et al., 2012), and is sensitive to retrieval interference from elements in memory that share some features with the target for retrieval (Jäger et al., 2020; Patil et al., 2016; Van Dyke, 2007; Van Dyke & McElree, 2006).

Anti-locality effects, as found for head-final dependencies in various languages (e.g., Konieczny, 2000; Nakatani & Gibson, 2010; Vasishth & Lewis, 2006), were considered some of the strongest pieces of evidence for expectation-based models (see an introduction below). However, Vasishth and Lewis (2006; also Lewis et al., 2006) argued that a memory-based account is equally adequate to account for anti-locality effects in head-final constructions. Previous empirical and theoretical discussion on anti-locality effects heavily focused on the head-final dependencies, especially the dependency relation between a verb and its dependents. In this study, we examine the (anti-)locality effects in a non-head-final syntactic dependency. Our empirical findings shed new light on the nature of anti-locality effects and the interaction between expectation- and memory-based sentence processing mechanisms.

Background on Anti-Locality Effects

In verb-final structures such as (1), Konieczny (2000) found that the reading time (RT) of the clause-final verb was shorter with a larger number of arguments preceding it, being, for example, shorter with the prepositional phrase than without it in (1).

(1) *Er hat den Abgeordneten (an das große Rednerpult) begleitet, und…*

He has the delegate (to the big lectern) escorted, and…

'He has escorted the delegate (to the big lectern), and…'

Similar anti-locality effects have been found for verb-final structures in other head-final languages, such as Hindi (Vasishth & Lewis, 2006) or Japanese (Nakatani & Gibson, 2010), and in non-head-final languages such as English (Jaeger et al., 2005).

Expectation-Based Account of Anti-Locality. The expectation-based model (Hale, 2001; Levy, 2008) provides a natural explanation for the anti-locality effect in a probabilistic framework pursuing several possible parses in parallel during sentence processing. The possible structures are ranked in order of preference by placing a probability distribution over them that ranks the likelihood of all possible structural representations given the previous sentential and extra-sentential context. Words that are unexpected continuations of a given sentence are assumed to lead to processing difficulties. Conversely, highly expected words are processed more quickly. Using surprisal as the complexity metric, the processing difficulty of the current word can be quantified as its negative log-probability given the previous context (Levy, 2008). In (1) above, the expectation-based model accounts for the anti-locality effect by way of expectations generated by the verb's dependents and the intervening prepositional phrase: For one, the accusative case marker on the definite determiner in *den Abgeordneten* ('the delegate') indicates that a transitive verb has to follow. Second, intervening phrases (e.g., 'to the big lectern') make the final verb more predictable, both by constraining the number of ways the sentence can continue and by raising the expectation of a specific verb that is compatible with the established context.

Memory-Based Account of Anti-Locality. Memory-based accounts of syntactic parsing (Gibson, 1998, 2000; Lewis & Vasishth, 2005) generally predict locality instead of anti-locality effects (for an overview on locality effects, see Gibson, 1998; see also Balling &

Kizach, 2017; Chen et al., 2012); that is, with increasing distance between sentential dependents, activation decay and interference from intervening items will increase the processing cost at the final constituent of the dependency relation. For example, the cue-based retrieval model of sentence processing (Lewis & Vasishth, 2005), built within the ACT-R (Adaptive Control of Thought-Rational) cognitive architecture (Anderson, 2005), assumes that upon encountering an element that forms a dependency with a previously parsed constituent, a memory retrieval will be initiated. Guided by the relevant retrieval cues, this process needs to access and reactivate the other element of the dependency. Longer distance can lead to increased processing difficulties due to two features of the system: activation decay and retrieval interference. Items in memory decay over time. After a long time of nonuse, the retrieval target could have a low activation level and poor quality of representation, leading to a longer processing time for memory retrieval and integration. A longer dependency length also increases the likelihood of introducing other elements into the memory that share semantic or syntactic features with the retrieval target, reducing the discriminability of the retrieval cues and contributing to more retrieval errors due to similarity-based interference (Patil et al., 2016; Van Dyke, 2007; Van Dyke & McElree, 2006).

Although memory-based models have been traditionally associated with locality effects, Vasishth and Lewis (2006) argued that the cue-based memory retrieval model could also explain anti-locality effects in verb-final structures. Using Hindi as an example, Vasishth and Lewis reasoned that in a verb-final argument–verb dependency, if an intervening element is also a dependent of the verb, such as an adverb or a prepositional modifier, it will activate the verb representation even before the verb is encountered. More pre-verbal material triggers repeated activation and therefore strengthens the verb representation in memory and facilitates the integration of the verb when it is encountered in the input. Such an account

offers an elegant explanation of the anti-locality effect under the general architecture of the cue-based memory retrieval model, and the proposal is significant because it can potentially unite the expectation- and memory-based processing models. The account, however, comes with a crucial constraint: It is only applicable when the interpolated material has properties sufficient to pre-activate the end element of the dependency. Cue-based retrieval works well for verb-final structures where the pre-verbal material is structurally dependent on the verb, such as in (1), where the prepositional phrase *to the big lectern* appears ahead of the verb it modifies. However, if an anti-locality effect is observed for dependency phenomena without a structural dependence between the intervening material and the forthcoming element, the explanation would have to make reference to the expectation based mechanism.

Interplay of Expectation-Based and Memory-Based Mechanisms. Both locality and anti-locality effects can be observed in the same experiment. Konieczny (2000), for instance, investigated both locality and anti-locality effects in sentence processing in German. Critically, he manipulated two factors within a self-paced reading experiment: (a) the position of the relative clause (RC; adjacent vs. extraposed) and (b) the lengths of the matrix clause and the RC (see an example in [2]). In doing so, he investigated what effect increased distance from their dependents has on two critical words, namely the verb of the matrix clause and the relative pronoun (each marked in bold).

(2) *Er hat die Rose [RC1] {**hingelegt**/ auf den (kleinen runden) Tisch **gelegt**} [RC2]*

He has the rose [RC1] {**laid-down**/ on the (small round) table **laid**} [RC2]

'He has laid the rose [RC1] down (on the [small round] table) [RC2]

With the RC of varying length at Position 1 or 2:

die *{wunderschön/ auffällig schön und farbenprächtig/ auffällig schön*

that {beautiful/ remarkably beautiful and colorful/ remarkably beautiful

gewachsen und besonders farbenprächtig} war

grown and particularly colorful was

'that was beautiful, …' / 'that was remarkably beautiful and colorful, …' / 'that had grown remarkably beautiful and was especially colorful, …'

The results showed that verbs following an RC were reliably read faster, confirming an anti-locality effect. The relative pronoun, on the other hand, was read faster in adjacent RCs than in extraposed RCs, confirming a locality effect. Konieczny (2000) argued that the differential patterns for verbs and relative pronouns were due to a strong syntactic expectation for the verb and a much weaker expectation for the RC. It should be noted, however, that the RC is not only less expected; extraposed RCs are also less frequent than adjacent RCs (see the corpus data in the following section, and Strunk, 2014; Blümel & Liu, 2020). The expectation-based model explains the anti-locality effect on the verb in the same way it explained the anti-locality effect in (1). The locality effect on the relative pronoun, on the other hand, may be a reflection of the lower frequency of extraposed RCs or may indicate that memory-based mechanisms play a larger role when the syntactic expectation is weak.

Similarly, Husain et al. (2014) found that locality and anti-locality effects can even emerge in the same type of structure. The authors reported on a study in Hindi wherein they manipulated the strength of the expectation for the final constituent in a long-distance argument–verb dependency such that either a specific verb or only the lexical category "verb" could be predicted. Although in both conditions, a verb was needed to complete the syntactic dependency, Husain et al. (2014) found an anti-locality effect only in the strong expectation condition. The authors took this as an indicator that memory effects affected the condition in which readers only had a generic verbal prediction, whereas strong predictions for the specific upcoming verb could override memory effects.

To summarize, memory- and expectation-based mechanisms can be seen as complementary, rather than mutually exclusive, constraints in sentence processing (similar

views have been advocated, for instance, by Demberg & Keller, 2008; Levy & Keller, 2013; and Vasishth & Drenhaus, 2011). The anti-locality effect has significant empirical status because it has been seen as a major piece of evidence for expectation-based syntactic parsing. However, previous investigations have primarily focused on head-final argument–verb dependencies, for which memory-based accounts like the cue-based retrieval model can make the right predictions as well. We aim to extend the empirical landscape on anti-locality effects through an investigation of a novel, non-head-final dependency, furthering our understanding of the respective roles expectation and memory mechanisms play in sentence comprehension.

Empirical Phenomenon: German Determiners That Need Relative Clauses

This study investigates anti-locality effects outside of head-final structures. In particular, we examine the dependency relation between a determiner and an RC within a determiner phrase (DP). As is documented in the linguistic literature (Alexiadouet al., 2000; Blümel, 2011; Sternefeld, 2008), German has two distinct sets of definite determiners that either can or must take an RC. The bare definite determiner *der* (as well as *das/die*) is a simplex determiner with gender and number markings that agrees with the following noun (e.g., *der Ball* [‘the ball’] vs. *die Bälle* [‘the balls’]). On the other hand, *derjenige* (‘the one’)[1] (henceforth referred to as *d-jenig*) is a morphologically complex determiner composed of the bare definite determiner *der* (which agrees with the noun), the stem of the demonstrative *jen-*, and a weak adjectival ending that agrees with the noun (*-ig*). Crucially, a restrictive RC is optional for the bare definite determiner, see (3a), but obligatory for the complex determiner *d-jenig* (3b). Although different syntactic or semantic analyses were proposed to capture the contrast (cf. Simonenko, 2015; see also Blümel & Liu, 2020), we stay analysis-neutral.

[1] Although we loosely translate *derjenige* to ‘the one’ throughout this paper, this translation does not perfectly capture the complex meaning of *derjenige* (see Blümel & Liu, 2020).

Critically relevant for the current purpose, it suffices to know that the complex determiner triggers a strong syntactic expectation for an upcoming RC, whereas the bare determiner does not generate such an expectation.

(3) a. *Der Ball(, der auf dem Tisch liegt,) ist blau.*
The ball(, that on the table lies,) is blue
'The ball that lies on the table is blue.'

b. *Derjenige Ball*(, der auf dem Tisch liegt,) ist blau.*
The-one ball*(, that on the table lies) is blue
'The one ball *(that lies on the table) is blue.'

Strunk (2014) and Blümel and Liu (2020) examined the differences between the bare and complex determiners using naturalness rating tasks. An example from the latter study is presented in (4). Both studies found that bare determiners in German received lower ratings with distant RCs than with locally attached RCs, whereas distant and local RCs were judged as equally good for the complex determiner *d-jenig*. Taken together, with regard to the co-occurring RCs, these studies found a locality preference for the bare determiner but not for the complex determiner. This provides preliminary evidence for processing differences between bare and complex determiners. Still, results from offline naturalness rating tasks do not address the question of how expectation- and memory-based mechanisms shape incremental online processing. We address that question in this study.

(4) a. *Maria hat {**dem/ demjenigen**} Mitarbeiter den Vorschlag einer Lohnerhöhung*
Maria has {**the/ the-one**} employee the suggestion of-a salary-raise
gemacht, der das Projekt bald beantragen sollte.
made who the project soon apply should

b. *Maria hat {**dem/ demjenigen**} Mitarbeiter, der das Projekt bald beantragen*
Maria has {**the/ the-one**} employee who the project soon apply

sollte,	*den*	*Vorschlag*	*einer*	*Lohnerhöhung*	*gemacht.*
should	the	suggestion	of-a	salary-raise	made

'Maria has made the suggestion of a salary raise to the colleague who should apply for the project soon.'

Before we conducted these experiments, we performed a corpus analysis to quantify the relevant differences between the two determiners. In particular, our corpus analysis was guided by two questions: (a) How often does an RC occur following a noun in a DP with the bare or complex determiner—that is, should the presence of either determiner give rise to expectations of an RC—and (b) when the noun is followed by an RC, how often is it a local or distant RC—that is, are readers likely to hold graded expectations regarding the location of the RC? To answer these questions, we needed a corpus annotated for dependency relations, allowing us to (a) match the RC to its hosting DP and (b) determine the distance between the head determiner and the RC. Because most dependency-annotated corpora of German that we are aware of are rather small, it is difficult to gather sufficient data for the infrequent determiner *d-jenig* to derive a robust estimate for the conditional probabilities. For instance, from TüBa-D/Z (Telljohannet al., 2004) we were able to extract only 17 instances of *d-jenig* as a determiner, and from TIGER (Brants et al., 2004) only seven. Eventually, we settled on the 2,100,000-sentence dependency-annotated corpus of German newspaper articles hosted at Treebank.info (Uhrig & Proisl, 2011), which allowed us to extract 133 instances of the complex determiner. The data are reported in Table 1.

Table 1

Conditional Probabilities for RCs in DPs Headed by Der *or* D-jenig

	Bare determiner (*der*) + noun			*Complex determiner (d-jenig)* + noun		
x	No RC	Local RC	Distant RC	No RC	Local RC	Distant RC
p(x|Det)	0.975	0.015	0.009	0.008	0.519	0.474
Frequency	1,997,306	31,417	19,245	1[2]	69	63

Note. Conditional probabilities for RCs in DPs headed by *der* or *d-jenig* (in any case or number) as estimated from corpus counts at Treebank.info. The conditional probabilities were calculated from the count for individual cells in the table, that is, the joint frequency of a bare/complex determiner and an RC type divided by the total frequency of the determiner, e.g., P(local RC|bare determiner) = 31,417/(31,417+1,997,306+19,245). RC = relative clause; DP = determiner phrase.

The Scope and Aim of This Study

This study takes advantage of two different kinds of German definite determiners. The comparison between the two provides an opportunity to tease apart expectation- and memory-based accounts for the anti-locality effect. Over the course of two experiments, we examine how a pre-nominal determiner influences the processing of an RC later within the sentence. In particular, we measure the RT of an RC-initial relative pronominal phrase in order to estimate the syntactic integration cost of attaching the (un-)expected RC. For the complex determiner, because an RC is obligatory, upon encountering *d-jenig*, a comprehender is predicted to have a strong expectation for an upcoming RC. As there is no

[2] The sentence (no. 922,624 in the corpus) is: *Und erinnert dabei vielleicht an dasjenige Land von vor etwa 15 oder 30 Jahren, oder das, was wir denken, das es war.* (‘And reminds therein maybe of the-jenig country of before roughly 15 or 30 years, or that, which we think that it was.’). We refer readers to Blümel and Liu (2020) for a discussion on *d-jenig* with prepositional phrase modifiers.

clear preference for close or distant RCs in the corpus data, the expectation should be formed immediately at *d-jenig*, but both local RC attachment and an extraposed RC position should be considered approximately equally likely. Due to syntactic restrictions in German, the RC can appear only at two positions, immediately after its host DP or after the main clause boundary. Thus, as comprehenders incrementally process the sentence, their expectation for the RC is predicted to sharpen to the clause boundary position once local attachment does not occur. This should facilitate attaching the RC in distant RC configurations once it appears in the linguistic input. In other words, expectation-based models would predict an anti-locality effect for the distant RC with the complex determiner. In contrast, for the bare determiner, there is no strong syntactic expectation for an upcoming RC in the first place: Contrary to the verb-final structures investigated in previous studies, the RC is not an obligatory element in the sentence (for the bare determiner, that is). Furthermore, as indicated by the corpus data, if an RC appears at all, local RC attachment is more frequent than distant RC attachment. The expectation-based account would therefore predict increased processing times for the relative pronominal phrase of distant RCs compared to close RCs.

The cue-based memory retrieval model account of anti-locality effects, as discussed in Vasishth and Lewis (2006), would critically not predict anti-locality for this phenomenon. Different from the verb-final structures discussed in Vasishth and Lewis (2006) and elsewhere, the RC is only structurally related to the determiner. In verb-final constructions, intervening prepositional phrase or adverbial modifiers of the verb can pre-activate and strengthen the representation of the upcoming verb. In the determiner–RC relation, on the other hand, no additional syntactic dependents of the RC are introduced between the two constituents of the dependency relation. Instead, under the memory-based models, increasing the distance between *d-jenig* and the RC would likely introduce more memory interference and possibly also more decay of the determiner representation, all of which could lead to a

standard profile of a locality effect, that is, increased integration cost at the distant RC. For the bare determiner, memory-based models would also predict a standard locality effect favoring the local RC over the distant RC.

We first present a self-paced reading experiment (Experiment 1) conducted to test these predictions. The experiment did not find the anti-locality effect predicted under the expectation-based account, which may be attributed to low statistical power and confounds in the experimental design. We then report a prospective power analysis based on the data of Experiment 1, which formed the basis for a large-sample conceptual replication experiment (Experiment 2). In this second—better powered—experiment, we confirmed the anti-locality effect with the complex determiner, but also found evidence for additional distance-based processing costs.

Experiment 1

As discussed earlier, because the complex determiner *d-jenig* induces a strong syntactic expectation for an RC, expectation-based models predict anti-locality effects will rise with increasing distance between the pre-nominal complex determiner and the post-nominal RC. Memory-based accounts, on the other hand, expect processing difficulties to increase as the distance between the complex determiner and the RC increases. In verb-final structures, pre-verb modifications might strengthen the representation of the verb head and thus counteract the locality effect. The lack of a verb–argument dependency in our study allowed us to test the predictions of expectation- and memory-based accounts without such confounding factors.

Materials and method

Participants

Fifty-four students (43 female) at Osnabrück University participated in the study for course credits. They were aged between 19 and 29 years, with a median age of 22. All

participants were native German speakers with normal or corrected-to-normal vision. They gave written informed consent to the experiment prior to participating. The experiment took roughly 40 min. Both reported experiments were approved by the ethics committee of Osnabrück University.

Design and Materials

The experiment was conducted in a moving-window self-paced reading paradigm on the online platform Ibex Farm (Drummond, 2013). The materials consisted of 24 experimental items and 48 fillers of comparable length. All items were 'chunked' such that the participants would be reading them phrase by phrase. The items are listed in Appendix A.1. An example item can be seen in (5), with the slashes indicating the division of words into regions. The critical region (CR) for the experimental items is the relative pronominal phrase (e.g., *dessen Mutter,* 'whose mother'). The experimental items were constructed using a 2 × 2 design, with the type of determiner (*den* vs. *d-jenig*) and the location of the RC (immediately after the related DP vs. extraposed to the end of the sentence) as the two factors. We used a Latin square design to ensure that participants saw each item only once and that all conditions were equally distributed. Participants were instructed to read the sentences region by region on a computer screen by pressing the space bar to continue. All sentences appeared on a single line in the middle of the screen. Participants initially saw a line of gray dashes where the words would appear. When they pressed the space bar, the first region appeared in place of the first gray dash. Once the participant had read a region and pressed the space bar, the subsequent region would appear, whereas the former was again obscured by a gray dash. Additionally, the subjects were instructed to give a naturalness rating on a 7-point Likert scale for each sentence after reading it.

(5) a. *Maria Richter / hat / den Mitarbeiter, / [dessen Mutter]*$_{CR}$ / *ein großes Haus /*

Maria Richter / has / the employee, / [whose mother]$_{CR}$ / a big house /

auf Mallorca / hat, / in einem Café / gesehen. **(bare-close)**

on Mallorca / has, / in a café seen

b. *Maria Richter / hat / den Mitarbeiter / in einem Café / gesehen, / [dessen*

Maria Richter / has / the employee / in a café / seen, / [whose

Mutter]$_{CR}$ */ ein großes Haus / auf Mallorca / hat.* **(bare-distant)**

mother]$_{CR}$ / a big house / on Mallorca / has

c. *Maria Richter / hat / denjenigen Mitarbeiter, / [dessen Mutter]*$_{CR}$ */ ein großes*

Maria Richter / has / the-one employee, / [whose mother]$_{CR}$ / a big

Haus / auf Mallorca / hat, / in einem Café / gesehen. **(complex-close)**

house / on Mallorca / has, / in a café seen

d. *Maria Richter / hat / denjenigen Mitarbeiter / in einem Café / gesehen, /*

Maria Richter / has / the-one employee / in a café / seen, /

[dessen Mutter]$_{CR}$ */ ein großes Haus / auf Mallorca / hat.* **(complex-distant)**

[whose mother]$_{CR}$ / a big house / on Mallorca / has

'Maria Richter has seen the employee [RC_1 whose mother owns a big house on Mallorca] in a cafe [RC_2].'

Predictions

Concerning the RT and following expectation-based models of language processing, our first hypothesis was that we would see an anti-locality effect in the *d-jenig* conditions in the form of a reduced RT at the CR (i.e., the relative pronominal phrase "*whose mother*") in the 'complex-distant' condition compared to in the 'complex-close' condition (5d vs. 5c). The cue-based memory retrieval model would not predict such an anti-locality effect, because in contrast to the Hindi case discussed in Vasishth and Lewis (2006), the additional interpolated material in the 'complex-distant' case does not include further dependents of the RC that can continuously pre-activate the RC representation. Our second hypothesis,

concerning the *der* conditions, was that we would not find an anti-locality effect, as the RC cannot be anticipated by the presence of the bare determiner. Instead, we hypothesized that we might find a longer RT at the relative pronominal phrase in the 'bare-distant' condition due to the locality preference as found in previous rating studies (Blümel & Liu, 2020). We therefore predicted an interaction between the type of determiner (bare vs. complex) and the location of the RC (close vs. distant)

Results

For all statistical analyses, we carried out Bayesian regression analyses using the brms package, version 2.12 (Bürkner, 2017) in R, version 4.0 (R Development Core Team, 2019). The rating data were analyzed using Bayesian ordinal regression models with a cumulative link function, and the RT data were analyzed using Bayesian linear mixed effects regression models. RTs were log-transformed before the analysis. The predictor variables *determiner* and *locality* were entered as sum-coded contrasts (0.5, −0.5) with interaction. When pairwise comparisons between conditions were needed to resolve an interaction, models were rerun with treatment-coded contrasts (1, 0). We used the maximal random effects structure for all models including random by-subject and by-item intercepts and slopes for all fixed effects and their interaction. We included weakly informative priors for the fixed effect estimates in both the rating and the RT data: We set the priors for the estimated parameters in the rating data to a normal distribution with a mean of 0 and a standard deviation of 6. For the log-transformed RT, we set the priors to a normal distribution with a mean of 0 and a standard deviation of 2. Both of these priors remain non-committal with regard to the size or direction of the effects, but constrain our prior assumptions to physically plausible effect sizes. For the rating model, four chains were run with 4,000 sampling iterations each using a warm-up period of 2,000 iterations. For the RT models, we used 8,000 sampling iterations with a warm-up period of 4,000 iterations. We report the posterior parameter estimates together with

the 95% credible intervals (CrI) and the posterior probability that the parameter value is bigger/smaller than 0.

The naturalness ratings (Table 2, Figure 1) were high overall (*M* = 5.48, *Md* = 6, *SD* = 1.51). The posterior was consistent with an interaction between determiner and locality, $\hat{\beta}$ = 0.46, CrI = [0.13, 0.80], P(β > 0) = 0.99.[3] For the bare determiner *der*, the analysis supported that distant RCs were rated less natural than close RCs, $\hat{\beta}$ = -0.38, CrI = [-0.65, -0.12], P(β < 0) = 1, which indicates a preference for locally attached RCs. For *d-jenig*, on the other hand, the posterior did not support a difference in the naturalness ratings between close and distant RCs, $\hat{\beta}$ = 0.09, CrI = [-0.18, 0.35], P(β < 0) = 0.24. In other words, we were able to replicate the rating results in Strunk (2014) and Blümel and Liu (2020).

Table 2

Naturalness Ratings and Raw RTs for Experiment 1

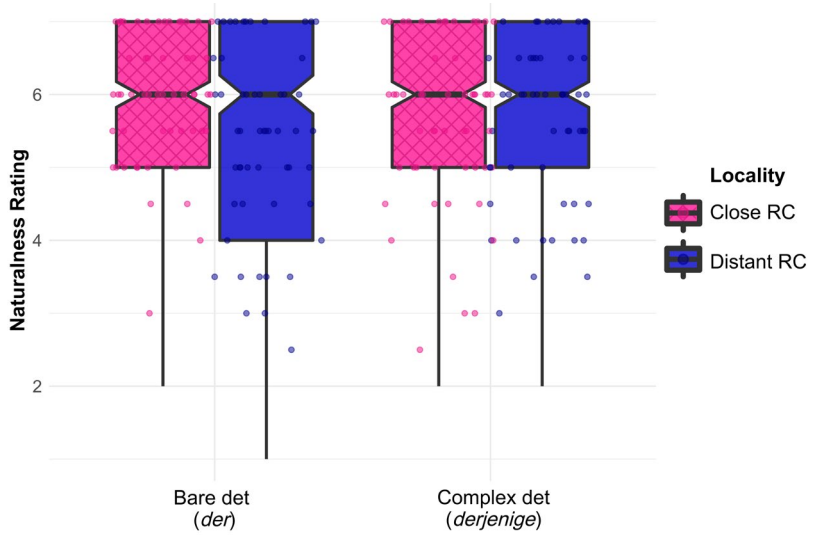
Condition	Naturalness rating	RT CR-1	**RT CR ("whose mother")**	RT CR+1
Bare-close	5.69 (1.38)	854.69 (446.35)	**802.97 (354.04)**	953.41 (331.69)
Bare-distant	5.20 (1.70)	823.06 (366.44)	**757.75 (271.07)**	952.24 (371.94)
Complex-close	5.49 (1.45)	1147.25 (500.12)	**868.64 (406.96)**	955.73 (323.22)
Complex-distant	5.52 (1.45)	783.30 (339.90)	**750.96 (267.83)**	934.29 (335.95)

Note. This table summarizes the mean naturalness ratings (on a 1-7 Likert scale) and raw RTs (in ms) for the critical region (CR) as well as the preceding (CR-1) and following (CR+1) regions for Experiment 1 (with standard deviation in parentheses). RT = reading time.

[3] This notation indicates the expected mean value of the estimated parameter, $\hat{\beta}$, the lower and upper bounds of the 95 percent credible interval, CrI, and the posterior probability, given the model and the data, for the hypothesis that the parameter value would be bigger than 0, P(β > 0).

Figure 1

Naturalness Ratings for Experiment 1

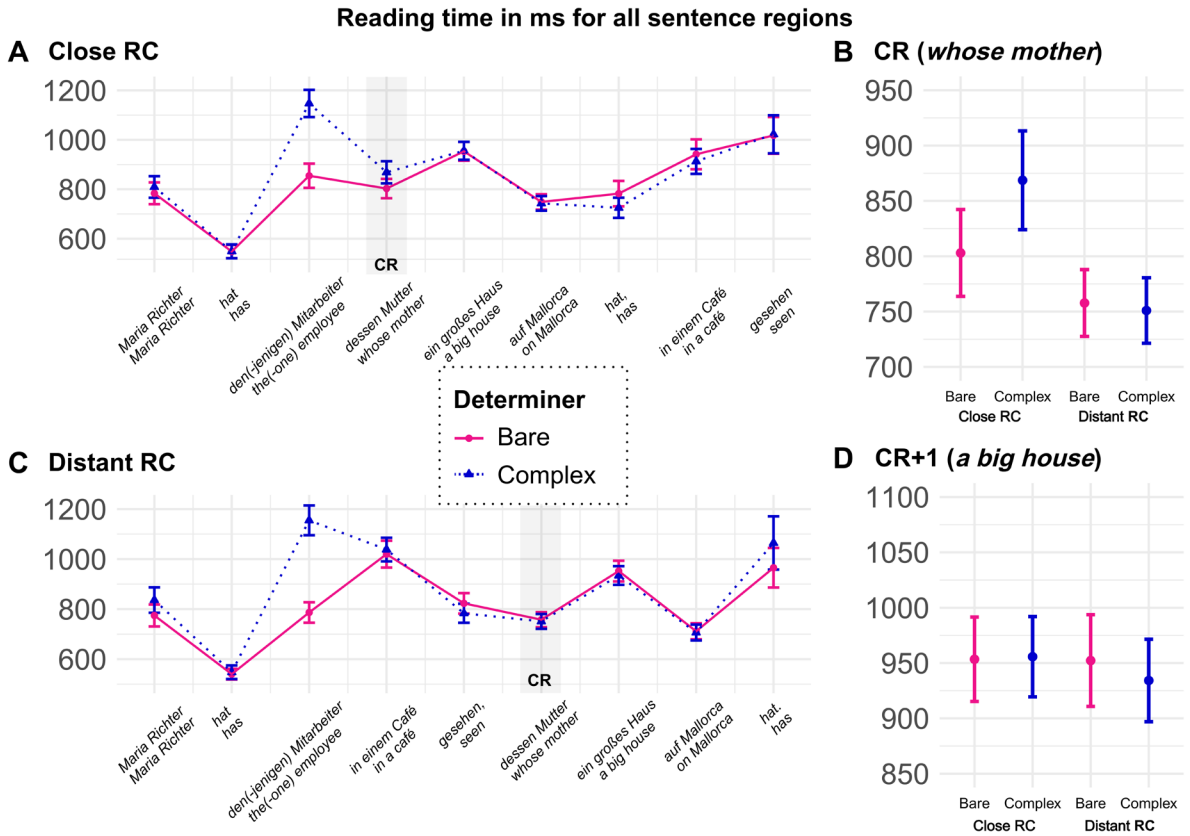


Note. Naturalness ratings (on a 1-7 Likert scale) for the four conditions of Experiment 1. The thick horizontal line indicates the median rating and the box the interquartile range (IQR). Whiskers extend to the smallest/largest value that is no further than 1.5-times the IQR away from the hinges of the box. Dots display individual participants' median naturalness rating.

For the RT (Table 2, Figure 2), separate models were fit for the CR (the relative pronominal phrase, e.g., *dessen Mutter*, 'whose mother' in [5]) as well as for the region immediately following it (CR+1; e.g., *ein großes Haus*, 'a big house' in [5]). An additional analysis for the pre-critical region (CR-1; e.g., *den[-jenigen] Mitarbeiter*, 'the[-one] employee' / *gesehen*, 'seen' in [5]) is available in Appendix B.

Figure 2

Mean RT for all Sentence Regions in Experiment 1.



Note. Panels A and C of this figure show the mean raw reading times in ms for all sentence regions of the four conditions of Experiment 1. The critical region is the relative pronominal phrase, indicated as CR. Because of the difference in the position of the CR, we grouped the plots by 'close' and 'distant' conditions. Panels B and D focus in on the mean raw RT[4] at the CR and CR+1. Error bars indicate the 95% confidence interval. RC = relative clause; RT = reading time; CR+1 = the region immediately following the CR.

We first focus on the results for the CR: Because there were large differences in the RTs at the CR-1, particularly due to a prolonged RT in the 'complex-close' condition, where

[4] Additional plots of the posterior predicted RT are available in the appendices and the OSF repository for this paper.

the region contained the complex DP *denjenigen Mitarbeiter* ('the-one employee')[5], we included the log-transformed RT from the CR-1 as additional predictor in our model for the CR. In doing so, variance explained through spillover from the previous region could be factored out in the statistical model, by which we hoped to account for spillover effects that might otherwise confound our findings (Vasishth, 2006). The posterior for the CR yielded no clear indication of the Determiner × Locality interaction we predicted, $\hat{\beta} = -0.03$, CrI = [-0.10, 0.05], $P(\beta < 0) = 0.76$, and did not support an effect for the determiner, $\hat{\beta} = 0.01$, CrI = [-0.03, 0.05], $P(\beta < 0) = 0.31$. However, the posterior was consistent with an effect of the RC location, $\hat{\beta} = -0.05$, CrI = [-0.09, -0.01], $P(\beta < 0) = 0.99$, such that distant CRs were processed more quickly overall.

At the CR+1 region, the posterior did not indicate any effect (all CrIs included 0, all posterior probabilities < 0.85).

Discussion

The offline ratings in Experiment 1 indicated a locality preference for only the bare determiner. The online RTs at the CR, on the other hand, showed facilitated processing in the 'distant' RC conditions for both determiners. Our predictions regarding an expectation-driven anti-locality effect were thus not confirmed. Several factors may have contributed to this result: First, it may reflect a confounding influence from the CR-1. Despite including the CR-1 log RT as a predictor, we may not have completely factored out the spillover effects from this region. Second, in the 'distant' RC conditions, the matrix verb *gesehen*, ('seen') immediately preceded the CR. Following the memory-based accounts of sentence processing, the verb's arguments are reactivated to integrate them with the verb. Thus, in both 'close' and

[5] The contrast in the CR-1 can be accounted for by processing difficulties incurred through the morphological complexity and the lower frequency of the complex determiner as compared to the bare determiner.

'distant' RC conditions, the bare/complex DP was activated immediately before the RC, which might have undermined the effect of the RC distance manipulation in Experiment 1.

A further serious issue regarding the reliability of the results from Experiment 1 was low statistical power: We report a prospective power analysis in the following section which shows that, given our sample size and number of items (54 subjects, 24 items), we would achieve sufficient power (> 80%) for the Determiner × Locality interaction only if the true effect was at least as large as 50 ms. The observed difference, however, was around 30 ms, for which, if it reflects the true effect, we only achieved 51% power. We therefore conducted a second experiment that increased the number of items and the sample size to the number required to achieve more than 80% power, according to the power analysis. To address the additional above-mentioned confounds at the CR-1, the second experiment used a slightly modified design. In the following section, we first report the prospective power analysis before turning to Experiment 2.

Prospective Power Analysis[6]

Our first experiment did not show a Determiner × Locality interaction in the RT measure. As discussed above, the experiment may have suffered from low statistical power, particularly if the true effect size is small. To address this possibility, and to estimate the required sample size for a follow-up experiment, we conducted a prospective power analysis determining the estimated power for the Determiner × Locality interaction given a range of plausible effect sizes. The prospective power analysis was implemented in R, version 4.0 (R Development Core Team, 2019) following a procedure outlined by Vasishth et al. (2018):

---

[6] Ideally, in a fully Bayesian framework, one would determine a study's *precision*, i.e., its ability to estimate the relevant parameter's value within a desired range of certainty, rather than its *power*, a term that stems from the frequentist framework (see Kruschke and Lidell (2018) and Schönbrodt and Wagenmakers (2018) for discussions of Bayesian sample size determination). One downside of these procedures is that they are computationally expensive (Wang & Gelfand, 2002). For this study, we therefore use the approach of simulating data based on the parameters drawn from a Bayesian model fit to the data of Experiment 1, but subsequently calculate the frequentist prospective power estimate. Similar "hybrid" approaches have been used, for instance, by Du and Wang (2016).

First, we extracted all parameter estimates (including random effects estimates and standard deviations) from the Bayesian linear mixed regression model that was fit to the CR. Then, we used these estimates, as well as a range of plausible estimates for the effect of interest (the Determiner × Locality interaction), to simulate 100 fake data sets for each assumed effect size. We set the magnitude of the effect to a range of 20 to 60 ms, which matches locality-based effects observed in the past (Chen et al., 2012; Husain et al., 2014; Konieczny, 2000). Finally, we fit each of the simulated data sets with a maximal linear mixed effects model to determine how often the interaction would turn significant at the $\alpha < 0.05$ level, giving us a power estimate as the percentage of simulated data sets in which significant effects were observed.

For Experiment 1, given our sample size and item number (54 subjects, 24 items), we were at 70% power if the true effect was as large as 40 ms (our observed raw RT difference in the Determiner × Locality interaction, as can be calculated from Table 2, was only around 30 ms). For a smaller effect magnitude of around 30 ms, we achieved 51% power; for a 20 ms effect, only 26% power.

We also used the prospective power analysis to determine the required sample size for the second experiment. To trade off the required number of participants against the number of items, we increased the number of items from 24 to 32. Subsequently, we estimated the number of participants required to detect an effect with more than 80% power. The analysis indicated that, given a sample size of 160 participants, we would have 85% power to detect a 30 ms effect, and 61% power for a 20 ms effect. It is preferable to have more than 80% power, even for the 20 ms effect. However, as this would have required more than 500 participants, we settled on a planned sample size of 160 participants for the second experiment.

Experiment 2

Experiment 2 was a large-scale extension of Experiment 1, with improvements to the stimulus material to address confounds from the previous experiment. The design of this experiment was similar to that of Experiment 1, but we modified the materials in the following way (see an example in the next section): First, by using stimuli in the present tense, we moved the lexical verb to a sentence position before its DP arguments, which reduced the likelihood of a reactivation of the DP ahead of the RC in the 'distant' conditions. Second, we added a prenominal adjective between determiner and noun, e.g. *der(-jenige) tapfere Junge* ('the(-one) brave boy'), and then further split the DP into word-by-word 'chunks' for the self-paced reading presentation. With this simple change in design we hoped to reduce the influence of the spillover effect from the CR-1, which we hypothesized was due largely to the long RTs on the complex determiner. Whereas in Experiment 1, the complex determiner appeared in the CR-1 (together with the noun) immediately before the RC, Experiment 2 had two regions (the adjective and the noun) between the complex determiner and the RC.

Materials and Method

Participants

We recruited 169 native German speakers through the crowd-sourcing platform Prolific (https://www.prolific.co/), nine of whom were later excluded from the data set because they did not pass attention checks (< 80% accuracy on comprehension questions). The remaining 160 participants (49 female) were aged between 18 and 60 years, with a median age of 28 years. The experiment took about 20 min.

Design and Materials

Again, the experiment was conducted in a moving-window self-paced reading paradigm on the online platform Ibex Farm (Drummond, 2013). The materials consisted of

32 experimental items and 36 grammatical filler sentences of comparable length but without RCs. As with the previous experiment, the experimental items were constructed using a 2x2 design with the factors 'Determiner' (*der* vs. *d-jenig*) and 'RC Location' (immediately after the related DP vs. extraposed to the end of the sentence). The CR for the experimental items was the relative pronominal phrase. All items were 'chunked' such that participants would read them region by region. An example item with the slashes indicating the division of words into regions can be seen in (6). The full list of items is available in Appendix A.2. Crucially, we added a prenominal adjective to the subject DP and split the DP regions into word-by-word chunks. This addressed the concern of Experiment 1 in which the complex determiner in the CR-1 of the 'complex-close' condition led to longer RTs that might have affected the RTs on the CR immediately following. Secondly, we used present tense stimuli to move the lexical verb to a position preceding the subject DP. This addressed the second concern of Experiment 1 that the verbal head in the CR-1 of 'distant' conditions might have reactivated the DP. The subject and object DPs used different genders such that the relative pronominal phrase, which in German must match gender with its DP host, unambiguously indicated that the RC attaches to the subject DP.

(6) a. *Heute / besucht / der*$_{\text{MASC}}$ / *tapfere / Junge, / [dessen*$_{\text{MASC}}$ *Bein]*$_{\text{CR}}$ / *komplizierte*
Today / visits / the$_{\text{MASC}}$ / brave / boy, / [whose$_{\text{MASC}}$ leg]$_{\text{CR}}$ / complicated
Brüche / hat, / zur Nachuntersuchung / die$_{\text{FEM}}$ / *Ärztin.* **(bare-close)**
fractures / has, / to-the follow-up / the$_{\text{FEM}}$ / doctor

b. *Heute / besucht / der*$_{\text{MASC}}$ / *tapfere / Junge / zur Nachuntersuchung / die*$_{\text{FEM}}$ /
Today / visits / the$_{\text{MASC}}$ / brave / boy / to-the follow-up / the$_{\text{FEM}}$ /
Ärztin, / [dessen$_{\text{MASC}}$ *Bein]*$_{\text{CR}}$ / *komplizierte Brüche / hat* **(bare-distant)**
doctor, / [whose$_{\text{MASC}}$ leg]$_{\text{CR}}$ / complicated fractures / has

c. *Heute / besucht / derjenige*$_{\text{MASC}}$ / *tapfere / Junge, / [dessen*$_{\text{MASC}}$ *Bein]*$_{\text{CR}}$ / *komplizierte*

Today / visits / the-one$_{MASC}$ / brave / boy, / [whose$_{MASC}$ leg]$_{CR}$ / complicated

Brüche / hat, / zur Nachuntersuchung / die$_{FEM}$ / *Ärztin.* **(complex-close)**

fractures / has, / to-the follow-up / the$_{FEM}$ / doctor

d. *Heute / besucht / derjenige*$_{MASC}$ / *tapfere / Junge / zur Nachuntersuchung /*

Today / visits / the-one$_{MASC}$ / brave / boy / to-the follow-up /

die$_{FEM}$ *Ärztin, / [dessen*$_{MASC}$ *Bein]*$_{CR}$ / *komplizierte Brüche / hat* **(complex-distant)**

the$_{FEM}$ doctor, / [whose$_{MASC}$ leg]$_{CR}$ / complicated fractures / has

'The brave boy whose leg has complicated fractures visits the doctor today for his follow-up examination.'

We used a Latin square design to ensure each item was only seen once by participants and all conditions were equally distributed. Half the sentences were followed by a comprehension question regarding the content of the sentence the participant had just read. This question could be answered with "yes" or "no". The procedure and predictions were the same as in Experiment 1.

Results

Just as for Experiment 1, we used the brms package, version 2.12 (Bürkner, 2017) in R, version 4.0 (R Development Core Team, 2019) for all analyses, and applied Bayesian ordinal regression models to the rating data, and Bayesian linear mixed effects regression models to the RT data. RTs were log-transformed before the analysis. The predictor variables *determiner* and *locality* were entered as sum-coded contrasts (0.5, –0.5) with interaction. When pairwise comparisons between conditions were needed to resolve an interaction, models were rerun with treatment-coded contrasts (1, 0). We used the maximal random effects structure for all models including random by-subject and by-item intercepts and slopes for all fixed effects and their interaction. The priors were the same as for Experiment 1. For the rating model, four chains were run with 4,000 sampling iterations each using a warm-up

period of 2,000 iterations. For the RT models, we used 8,000 sampling iterations with a warm-up period of 4,000 iterations. We report the posterior parameter estimates together with the 95% credible intervals (CrI) and the posterior probability that the parameter value is bigger/smaller than 0.

For the naturalness ratings (Table 3, Figure 3), the posterior was consistent with main effects of determiner, $\hat{\beta} = -0.67$, CrI = [-0.80, -0.55], $P(\beta < 0) = 1$), and RC location, $\hat{\beta} = -1.09$, CrI = [-1.30, -0.89], $P(\beta < 0) = 1$, as well as an interaction, $\hat{\beta} = 0.73$, CrI = [0.55, 0.91], $P(\beta > 0) = 1$. The analysis supports that sentences with either determiner were rated as more natural with a close than with a distant RC. Additionally, it indicates that the difference between 'close' and 'distant' conditions was stronger for the bare determiner, $\hat{\beta} = -1.45$, CrI = [-1.71, -1.180], $P(\beta < 0) = 1$, than for the complex determiner conditions, $\hat{\beta} = -0.73$, CrI = [-0.91, -0.55], $P(\beta < 0) = 1$, resulting in the aforementioned interaction. The contrast between the 'bare-close' and 'bare-distant' conditions replicated the finding from Experiment 1. The difference between the 'complex-close' and 'complex-distant' conditions, however, was not entirely consistent with the earlier finding that close and distant RCs were rated as equally natural, as in Strunk (2014), Blümel and Liu (2020), or our Experiment 1. The main effect of determiner shows that sentences with the complex determiner were rated less natural overall than those with the bare determiner. We attribute this to the low frequency of prenominal modifiers with the complex determiner (see General Discussion).

For the RT (Table 3, Figure 4), separate models were fit for the CR (the relative pronominal phrase; e.g., *dessen Bein*, 'whose leg' in [6]) as well as for the region immediately following it (CR+1; e.g., *komplizierte Brüche,* 'complicated fractures' in [6]). An additional analysis for the pre-critical region (CR-1; e.g., *Jungen*, 'boy' / *Ärztin*, 'doctor' in [6]) is available in Appendix B.

Table 3

Naturalness Ratings and Raw RTs for Experiment 2

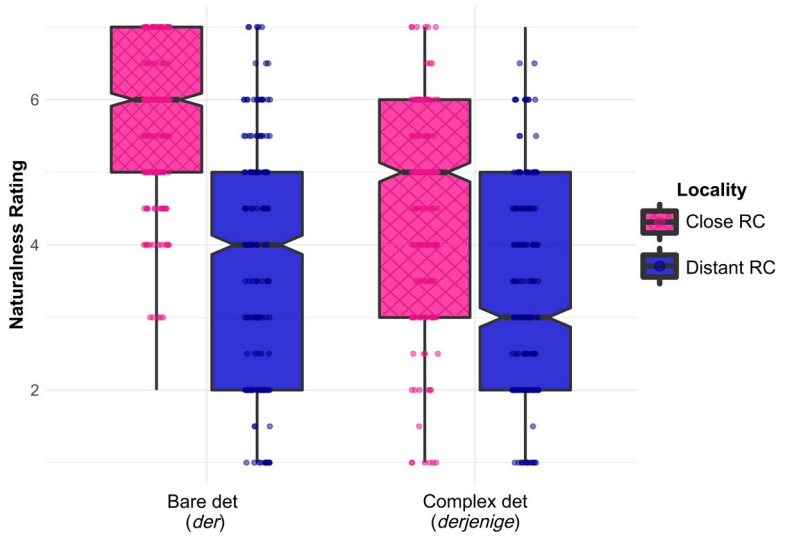
Condition	Rating	CR-1	**CR (“whose leg”)**	CR+1
Bare-close	5.48 (1.39)	509.81 (244.95)	**648.45 (305.46)**	675.49 (295.28)
Bare-distant	3.85 (1.87)	540.91 (294.75)	**665.40 (335.25)**	745.67 (360.38)
Complex-close	4.36 (1.69)	549.75 (262.25)	**658.20 (293.30)**	666.96 (296.62)
Complex-distant	3.49 (1.74)	542.25 (290.58)	**640.76 (322.28)**	727.39 (409.38)

Note. This table provides the mean naturalness ratings (on a 1-7 Likert scale) and raw RTs (in ms) for the critical region (CR), as well as the preceding (CR-1) and following (CR+1) regions of Experiment 2 (with standard deviation in parentheses). RT = reading time.

We first focus on the results for the CR: Again, we included the log-transformed RT from the CR-1 as an additional predictor in our model for the CR to account for any spillover from the previous region, $\hat{\beta} = 0.20$, CrI = [0.16, 0.24], $P(\beta > 0) = 1$. The model’s posterior distribution was consistent with the predicted Determiner × Locality interaction, $\hat{\beta} = -0.04$, CrI = [-0.08, -0.01], $P(\beta < 0) = 0.99$, such that the analysis supports that the ‘distant’ condition was processed faster with the complex determiner than with the bare determiner, $\hat{\beta} = -0.04$, CrI = [-0.07, -0.02], $P(\beta < 0) = 0.99$, whereas it did not support a difference between the two determiners in the ‘close’ conditions, $\hat{\beta} = 0.00$, CrI = [-0.02, 0.02], $P(\beta < 0) = 0.45$.

Figure 3

Naturalness Ratings for Experiment 2

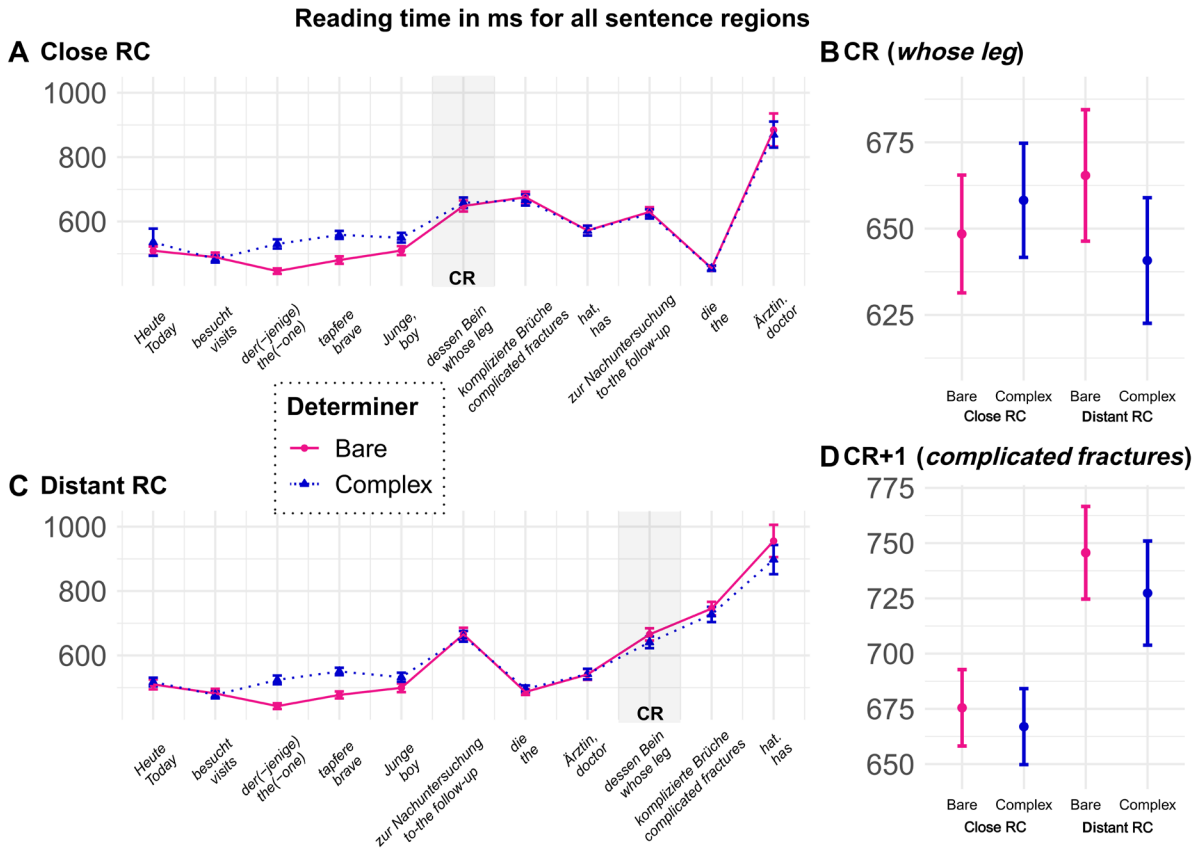


Note. Naturalness ratings (on a 1-7 Likert scale) for the four conditions of Experiment 2. The thick horizontal line indicates the median rating, and the box indicates the interquartile range (IQR). Whiskers extend to the smallest/largest value that is no further than 1.5-times the IQR away from the hinges of the box. Dots display individual participants' median naturalness rating. RC = relative clause.

At the CR+1, we find main effects of the determiner, $\hat{\beta} = -0.02$, CrI = [-0.04, -0.00], $P(\beta < 0) = 0.99$, and the RC location, $\hat{\beta} = 0.08$, CrI = [0.05, 0.11], $P(\beta > 0) = 1$, but no clear indication of an interaction, $\hat{\beta} = -0.02$, CrI = [-0.07, 0.02], $P(\beta < 0) = 0.87$. The analysis is consistent with faster processing of the region if the sentence included the complex determiner compared to the bare determiner, but it also supports that 'distant' conditions were processed more slowly overall than 'close' conditions, indicating distance-based processing costs or inhibitory interference from the object DP (see General Discussion).

Figure 4

Mean RT for All Sentence Regions of Experiment 2.



Note. Panels A and C of this figure show the mean raw reading times in ms for all sentence regions of the four conditions of Experiment 2. The critical region is the relative pronominal phrase, indicated as CR. Because of the difference in the position of the CR, we grouped the plots by 'close' and 'distant' conditions. Panels B and D focus in on the mean raw RT[7] at the CR and CR+1. Error bars indicate the 95% confidence interval. RC = relative clause; CR+1 = the region immediately following the CR.

To further quantify the evidence for the hypothesized anti-locality effect, we conducted a Bayes factor analysis on the Det x Locality interaction at the CR in both experiments (Table 4, see also Figure AC.1 in the appendices). We used the bridge sampling

[7] Additional plots of the posterior predicted RT are available in the appendices and the OSF repository for this paper.

method implemented in the *bayes_factor* function in the R package brms, version 2.12 (Bürkner, 2017) to compare a model with the Determiner x Locality interaction (H_1) to a reduced model in which the effect is assumed to be 0 (i.e., the interaction term is removed; H_0). The Bayes factor is known to be highly sensitive to the prior distribution (Du et al., 2019; Gelman et al., 2014) and to favor the null hypothesis under uninformative or weakly informative priors, particularly for small effects (Wagenmakers et al., 2010). We therefore calculated the Bayes factor for a range of increasingly informative priors on the interaction effect, using prior distributions with more and more narrow standard deviations around the mean (Jäger et al., 2020; Lee & Wagenmakers, 2013). By shrinking the standard deviation, the prior assigns less probability mass to large effect sizes, allowing us to test the evidence for small or medium-sized effects instead (Schad et al., 2021). The analysis indicates no evidence for the Determiner x Locality interaction in Experiment 1, regardless of prior informativity. For Experiment 2, the Bayes factor ranges from 0.48–4.98 with increasingly informative priors, which indicates evidence in favor of a small interaction effect. The fluctuation around 1 for less uninformative priors, on the other hand, shows that there is uncertainty about larger effect sizes. Therefore, to confirm the size and robustness of the effect, future replication attempts are certainly warranted.

Table 4

Bayes factor (BF_{10}) in favor of the Determiner × Locality interaction at the CR in Experiments 1 and 2

	Prior on Determiner × Locality interaction	BF_{10} Experiment 1	BF_{10} Experiment 2
1	Normal distribution ($\mu = 0, \sigma = 2$)	0.02	0.48
2	Normal distribution ($\mu = 0, \sigma = 1$)	0.05	0.90
3	Normal distribution ($\mu = 0, \sigma = 0.8$)	0.06	1.13
4	Normal distribution ($\mu = 0, \sigma = 0.6$)	0.13	1.77
5	Normal distribution ($\mu = 0, \sigma = 0.4$)	0.16	2.42
6	Normal distribution ($\mu = 0, \sigma = 0.2$)	0.25	4.98

Note. Results of a Bayes factor analysis quantifying the evidence in favor of the Determiner × Locality interaction at the critical region. We report the Bayes factor under a range of increasingly informative priors. Values below 1 indicate evidence in favor of the null hypothesis (no interaction), values above 1 indicate evidence in favor of the interaction.

General Discussion

Anti-locality effects in argument–verb dependencies have provided considerable evidence for expectation-based parsing. However, many prior investigations have focused on verb–final structures, wherein the effect is also compatible with a memory-based account. In our study, we focused on a different kind of dependency relation between the complex determiner *d-jenig* and an RC in German. The effect discovered for the complex determiner supports the expectation-based account of the anti-locality effect. Furthermore, we found both an anti-locality effect at the critical region (CR), and increased RTs at the post-critical region (CR+1) of 'distant' conditions, providing new empirical evidence to shed light on the trade-off between expectation and memory retrieval processes.

Expectation-Driven Anti-Locality and the Trade-Off Between Expectation and Memory Retrieval

The findings from the current study provide evidence for an anti-locality effect in the presence of strong expectations, which is in line with the existing evidence for anti-locality effects in head–final structures (Konieczny, 2000). More importantly, the current results support an expectation-based account of anti-locality effects and challenge the assumption that anti-locality effects can be wholly subsumed under a memory-based account (Vasishth & Lewis, 2006). According to the latter account, anti-locality effects in argument–verb dependencies can be explained in the memory-based model in terms of a strengthened representation of the upcoming verb, due to the dependency relationship between the verb and the pre-verbal arguments and adjuncts. However, this explanation cannot be applied to the determiner–RC relationship in our experiments: The intervening materials between the determiner and the RC are not dependents of the RC, which means that the activation of the intervening materials does not strengthen the representation of the upcoming RC. A potential way to account for the observed anti-locality effect under the cue-based retrieval model would be that the intervening material, specifically the object DP *the doctor* in the distant condition in Experiment 2, could have reactivated the subject DP *the boy* which the RC attaches to. This may be possible due to the fact that both DPs are co-arguments of the same verb, and the syntactic integration of the object DP would potentially trigger a retrieval of the other VP dependents, i.e. the subject DP. However, even if the subject DP were reactivated, we believe this is an unlikely explanation for the observed effect because the subject DP in the ‘bare-distant’ and ‘complex-distant’ conditions would be rendered equally accessible through reactivation, which is incompatible with the observed anti-locality effect for only the complex determiner.

In Experiment 2, we also found an interesting effect at the post-critical region (CR+1): An increased processing cost for both conditions with an extraposed RC. The effect at the CR+1 is compatible with the locality effect predicted under the standard memory retrieval account, which assumes that the final element of a syntactic dependency relation should take longer to process with increased distance from its dependents. As indicated by Husain et al. (2014), locality and anti-locality effects can emerge in the same structure such that strong expectations may override memory effects. Complementing this line of work, our study shows that anti-locality and locality effects potentially emerge within the same (high-expectation) structure, with anti-locality effects appearing on the critical word and locality effects appearing soon after. Critically, the determiner–RC relation investigated in our study differs from previous work on argument–verb dependencies in that neither the determiner nor the linguistic material intervening between determiner and RC are strongly constraining regarding the content of the upcoming RC. It seems plausible, therefore, that the expectation generated from the complex determiner *d-jenig* was primarily syntactic, enabling readers to pre-activate (parts of) the syntactic structure of the upcoming RC. The anti-locality effect at the critical region might thus reflect a facilitation in syntactically integrating a distant RC with the complex determiner. By contrast, memory-related costs might have re-emerged at the CR+1 because later integration processes could not benefit from the syntactic expectation generated by *d-jenig* to the same extent.

A strong alternative explanation for the effect at the CR+1 is that the intervening object DP *the doctor* in 'distant' conditions contributed to similarity-based interference during the retrieval of the subject DP. The cue-based retrieval model assumes that retrieving a target element from memory can be affected by inhibitory interference from items that partially or fully match the target's retrieval cues. In our particular case, to attach the RC, a DP that matches its number and gender cues (e.g., [+*singular*] and [+*masculine*] for (6))

must be retrieved from memory. Although the subject DP fully matches these retrieval cues, the object DP also matches at least one of them (number). The standard cue-based retrieval model thus predicts inhibitory interference due to cue overload. Notably, this effect was not present in Experiment 1, although the intervening material there, too, contained an additional DP inside the prepositional phrase (e.g., *in a café*). The difference between the two experiments may again be attributable to low statistical power in Experiment 1, but may also relate to the prominence of the respective intervening DPs: A recent revised version of the original cue-based retrieval model by Engelmann et al. (2019) adds, among other features, the assumption that the base-level activation of items in memory can differ depending on their prominence. Engelmann et al. suggest that nouns can be more prominent if occurring in salient grammatical or discourse positions such as, for instance, as subject or object of the sentence. In turn, similarity-based interference is assumed to be dependent on the relative activation levels of the target and distractor items, such that this model could, in principle, be compatible with increased interference from the object DP in Experiment 2 compared to the DP embedded inside the prepositional phrase in Experiment 1.

In any case, the presence of both effects, anti-locality for the complex determiner at the CR and increased processing costs at the CR+1 (due to locality effects or interference), suggest that expectations and memory effects are both at play in incremental sentence processing, although the exact mechanisms by which they interact are not well understood. Although there is evidence for locality and anti-locality effects in various languages and syntactic structures, convergent empirical findings from a wider range of constructions are needed to better understand the interaction between memory retrieval and syntactic expectation.

Offline Naturalness Ratings

Whereas the online RTs lent support to expectation-based parsing models, the offline naturalness rating data revealed additional factors distinguishing between Experiments 1 and 2. For the bare determiner conditions, in both experiments, the 'distant' condition was rated as significantly less natural than the 'close' condition. This was expected based on previous results from Strunk (2014) and Blümel and Liu (2020), as well as our own corpus data, wherein we found a preference for locally attached RCs with the bare determiner. The naturalness ratings for the complex determiner, on the other hand, were less consistent across the two experiments. Although in Experiment 1, we found a difference between complex and bare determiners, with no preference regarding local or distant RC attachment for the complex determiner conditions, Experiment 2 revealed a similar pattern for both types of determiners: Distant RC conditions were rated as less natural than close RC conditions.

One possible explanation for the offline naturalness ratings in Experiment 2 is that they are simply a reflection of the aforementioned inhibitory interference in 'distant' conditions. Alternatively, the data may indicate differential RC attachment preferences for *d-jenig* in a subject versus object position. Citing Shannon (1992), Strunk (2014) suggested that extraposition is more likely to be formed from object than from subject noun phrases (NPs). In Experiment 1, *d-jenig* was indeed the determiner of the object NP of the matrix clause, which may have rendered extraposition more acceptable. In Experiment 2, on the other hand, *d-jenig* was the determiner of the subject NP; extraposition might therefore be less common. We conducted a corpus search to test this idea: As before, all searches were conducted using Treebank.info (Uhrig & Proisl, 2011). We searched for all instances of *d-jenig* with or without an associate noun in the subject (SB) or accusative object (OA) position that headed an RC. The results of the corpus search are summarized in Table 5.

The corpus data are partially in line with the findings from our study's naturalness ratings: *D-jenig* in a subject position (with or without head noun) appears to prefer a locally attached RC. For *d-jenig* in the OA position, however, the corpus data also show a tendency for local RCs, albeit weaker than that found for subject position *d-jenig*. This is not directly compatible with the naturalness ratings of Experiment 1, Strunk (2014), and Blümel and Liu (2020), where local and distant RCs were rated equally natural. It will require further inquiry to determine whether the preference for a local RC by *d-jenig* in the OA position was not detected because Experiment 1 was underpowered, or whether a more large-scale corpus study would possibly yield a different picture.

Table 5

Conditional Probabilities for RC Positions with D-Jenig *as Subject or Accusative Object*

Condition	SB *d-jenig* +noun		OA *d-jenig* +noun		SB *d-jenig* -noun		OA *d-jenig* -noun	
RC	Local	Distant	Local	Distant	Local	Distant	Local	Distant
P(x|Condition)	0.77	0.23	0.42	0.58	0.87	0.13	0.66	0.34
Frequency	10	3	5	7	247	38	75	39

Note: This table shows the conditional probabilities for RC positions with *d-jenig* as subject or (accusative) object of the matrix clause as estimated from corpus counts at Treebank.info. The conditional probabilities were calculated by taking the counts from one cell in the table (e.g., subject position *d-jenig*+noun with local RC = 10) and dividing it by the total count for each condition (e.g., total count for subject position *d-jenig*+noun = 10+3). SB = subject; OA = accusative object; RC = relative clause; NP = noun phrase.

Conclusion

This paper investigated the interaction between memory- and expectation-based accounts of sentence processing in German relative clauses (RCs). We manipulated the distance between the RC and its corresponding noun phrase and the degree to which a determiner predicts an upcoming RC. Our results function as evidence supporting an expectation-based anti-locality effect, which is not subsumable under current versions of the memory-based account of anti-locality. Furthermore, our results show that memory-based effects and anti-locality can emerge within the same structure, supporting the view that memory- and expectation-based processes interact during online sentence processing.

Acknowledgements

This work has benefited from valuable comments of the audience at AmLaP 2018 and CUNY 2019, as well as from insightful discussions with Lyn Frazier, Andreas Blümel, and Lars Konieczny. We thank Danielle Benesch, Brianna Insley, and Birte Spekker for their help in the data acquisition. We also thank Peter Uhrig for providing access and assistance with the Treebank.info corpus.

Funding

J.S. is supported by the DFG-funded Research Training Group “Computational Cognition” (DFG-GRK 2340). The work was also supported by a NSF grant to M.X. (BCS1451635) and a DFG grant to M.L. (Project number 367088975).

Author Contributions

M.L. and M.X. conceived and planned Experiment 1. J.S., M.L. and M.X. conceived and planned Experiment 2. J.S. carried out both experiments and conducted data analyses. All three authors contributed to the interpretation of the data and writing of the manuscript, with J.S. preparing the first draft of the manuscript.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

All code and data associated with the study are available in the following repository: https://osf.io/h3qa8/

References

Alexiadou, A., Law, P., Meinunger, A., & Wilder, C., (2000). Introduction. In Alexiadou, A., Law, P., Meinunger, A., & Wilder, C. (eds.). *The syntax of relative clauses.* Amsterdam/Philadelphia: John Benjamins.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, *29*(3), 313–341. https://doi.org/10.1207/s15516709cog0000_22

Balling, L. W., & Kizach, J. (2017). Effects of surprisal and locality on Danish sentence processing: an eye-tracking investigation. *Journal of Psycholinguistic Research, 46*(5), 1119–1136. https://doi.org/10.1007/s10936-017-9482-2

Blümel, A. (2011). Derjenige determiner that wants a relative clause. *University of Pennsylvania Working Papers in Linguistics*, *17*(1), 21–29.

Blümel, A., & Liu, M. (2020). Revisiting obligatory relatives in German – empirical and theoretical perspectives. *Zeitschrift für Sprachwissenschaft 39*(1), 1–39. https://doi.org/10.1515/zfs-2019-2007

Brants, S., Dipper, S., Eisenberg, P., Hansen, S., König, E., Lezius, W., Rohrer, C., Smith,

G., & Uszkoreit, H. (2004). TIGER: Linguistic interpretation of a German corpus. *Journal of Language and Computation, 2004* (2), 597–620. https://doi.org/10.1007/s11168-004-7431-3

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

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, *52*(1), 144–169. https://doi.org/10.1016/j.jml.2004.10.001

Chen, Z., Jäger, L., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. In B. Stolterfoht & S. Featherston (Eds.), *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure* (pp. 43–62). Berlin, Boston: De Gruyter Mouton.

Chow, W. Y., Smith, C., Lau, E., & Phillips, C. (2016). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. https://doi.org/10.1080/23273798.2015.1066832

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Drummond, A. (2013). Ibex Farm. http://spellout.net/ibexfarm/ Accessed March 12, 2019

Du, H., Edwards, M. C., & Zhang, Z. (2019). Bayes factor in one-sample tests of means with a sensitivity analysis: A discussion of separate prior distributions. *Behavior Research Methods*, *51*(5), 1998–2021. https://doi.org/10.3758/s13428-019-01262-w

Du, H., & Wang, L. (2016). A Bayesian power analysis procedure considering uncertainty in effect size estimates from a meta-analysis. *Multivariate Behavioral Research*, *51*(5), 589–605. https://doi.org/10.1080/00273171.2016.1191324

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: A computational account. *Cognitive Science*, *43*(12). https://doi.org/10.1111/cogs.12800

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0077661

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton: CRC Press, Taylor & Francis Group. https://doi.org/https://doi.org/10.1201/b16018

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

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 Articulation Project Symposium* (pp. 95–126). Cambridge, MA: MIT Press

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001 - NAACL '01* (pp. 1–8). https://doi.org/10.3115/1073336.1073357

Hale, J. (2016). Information-theoretical complexity metrics. *Linguistics and Language Compass*. https://doi.org/10.1111/lnc3.12196

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: evidence from Hindi. *PloS One*. https://doi.org/10.1371/journal.pone.0100986

Jaeger, F., Fedorenko, E., & Gibson, E. (2005). Dissociation between production and comprehension complexity. Poster Presentation at the 18th CUNY Sentence

Processing Conference, University of Arizona.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language, 111*, https://doi.org/10.1016/j.jml.2019.104063

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645. https://doi.org/10.1023/A:1026528912821

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin and Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Logačev, P., & Vasishth, S. (2016). A multiple-channel model of task-dependent ambiguity

resolution in sentence comprehension. *Cognitive Science, 40*(2). https://doi.org/10.1111/cogs.12228

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

Nakatani, K., & Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science, 34*(1), 94–112. https://doi.org/10.1111/j.1551-6709.2009.01067.x

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00329

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

R Development Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods, 26*(1), 103–126. https://doi.org/10.1037/met0000275

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin and Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Shannon, T. (1992). Toward an adequate characterization of relative clause extraposition in

modern German. In I. Rauch, G. F. Carr, and R. L. Kyes (Eds.), *On Germanic Linguistics. Issues and Methods* (pp. 253–281). Berlin/New York: Mouton de Gruyter.

Simonenko, A. (2015). Structural triggers of the loss of scopelessness. In U. Steindl, T. Borer, H. Fang, A. G. Pardo, P. Guekguezian, B. Hsu, … I. C. Ouyang (Eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics* (pp. 191–200). Somerville, MA: Cascadilla Proceedings Project.

Smith, N. J., and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition 128*, 303–319. https://doi.org/10.1016/j.cognition.2013.02.013

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(2), 425–436. https://doi.org/10.1037/0278-7393.32.2.425

Sternefeld, W. (2008). *Syntax: Eine morphologisch motivierte generative Beschreibung des Deutschen*. Tübingen, DE: Stauffenburg.

Strunk, J. (2014). A statistical model of competing motivations affecting relative clause extraposition in German. In B. MacWhinney, A. Malchukov & E. Moravcsik (Eds.), *Competing Motivations in Grammar and Usage* (pp. 88–106). Oxford, UK: Oxford University Press.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition*. https://doi.org/10.3758/MC.36.1.201

Telljohann, H., Hinrichs, E., & Kübler, S. (2004). The TüBa-D/Z treebank: Annotating German with a context-free backbone. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)* (pp. 2229–2235), Lisbon, Portugal.

Traxler, M.J., Pickering, M.J., & Clifton Jr., C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558–592. https://doi.org/10.1006/jmla.1998.2600

Uhrig, P., & Proisl, T. (2011). The treebank.info dependency-annotated corpus interface. http://treebank.info/ Accessed January 29, 2019

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. http://dx.doi.org/10.1037/0278-7393.31.3.443

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(2), 407–430. https://doi.org/10.1037/0278-7393.33.2.407

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166. https://doi.org/10.1016/j.jml.2006.03.007

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2004.11.003

Vasishth, S. (2006). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories. In *Proceedings of the International Conference on Linguistic Evidence*. Tübingen, Germany.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language*, *82*(4), 767–794. https://doi.org/10.1353/lan.2006.0236

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2018.07.004

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Appendix A: Experimental Material

A.1 Experiment 1

We provide the experimental material in the condition determiner = bare, locality = distant. The three other conditions can be constructed by moving the RC after the object DP and/or changing the determiner of the object DP to *d-jenig.*

(1) Maria Richter hat den Mitarbeiter in einem Café gesehen, dessen Mutter ein großes Haus in Spanien besitzt.

(2) Tim Brandt hat den Freund auf einer Party kennengelernt, dessen Partner einen historischen Roman über China schreibt.

(3) Michaela Messner hat den Headhunter mit der Mitarbeitersuche beauftragt, dessen Manager ein kompetentes Team in Berlin leitet.

(4) Johanna Gabel hat den Anwalt für ihre Klage kontaktiert, dessen Kanzlei eine gute Bewertung im Internet hat.

(5) Das Vorstandsmitglied hat den Mitarbeiter auf die Bilanzfälschung angesprochen, dessen Team eine wichtige Instanz im Unternehmen ist.

(6) Das Amt hat den Architekten für ein Bauvorhaben eingestellt, dessen Büro ein günstiges Angebot zum Wohnungsbau macht.

(7) Nicole Schmidt hat den Manager auf einen Verbesserungsvorschlag angesprochen, dessen Geschäft einen neuen Standort in Frankfurt öffnet.

(8) Das Model hat den Fotografen auf einer Ausstellung getroffen, dessen Name eine bekannte Größe aus Modezeitschriften ist.

(9) Das Krankenhaus hat den Patienten von der Intensivstation verlegt, dessen Tochter einen kleinen Laden in Osnabrück hat.

(10) Das Team hat den Trainer für ein Jahr engagiert, dessen Taktik einen riesigen Erfolg im Turnier bewirkt.

(11) Das Kollektiv hat den Arbeiter von einem Streikaufruf überzeugt, dessen Unterstützung ein großer Vorteil in Verhandlungen ist.

(12) Jutta Wagner hat den Hausmeister für eine Reparatur gerufen, dessen Helfer eine zuverlässige Arbeit im Gebäude erledigt.

(13) Das Ensemble hat die Dirigentin für ein Konzert rekrutiert, deren Arbeit einen ausgezeichneten Ruf in Wien genießt.

(14) Moritz Stein hat die Studentin für eine Hilfsstelle angesprochen, deren Arbeit eine auffällige Qualität gegenüber anderen zeigt.

(15) Der Verein hat die Erzieherin in die Kitaleitung aufgenommen, deren Arbeitsweise eine starke Begeisterung unter Kindern auslöst.

(16) Das Ärzteteam hat die Patientin mit einem Knochenbruch diagnostiziert, deren Schmerz eine dringende Behandlung durch Medikamente erfordert.

(17) Das Generalskomitee hat die Roboterarmee auf einen Angriff programmiert, deren Codename ein offenes Geheimnis unter Angestellten ist.

(18) Merlin Auer hat die Frau durch einen Glücksfall angesprochen, deren Oma eine kleine Wohnung im Nachbarhaus besitzt.

(19) Die Richterin hat die Mutter vor einem Sorgerechtsentzug gewarnt, deren Kind ein bedenkliches Verhalten vor Gericht zeigt.

(20) Das Gremium hat die Sprecherin vor einem Publikum vorgestellt, deren Vortrag eine interessante Perspektive zur Diskussion hinzufügt.

(21) Thomas Lenz hat die Gärtnerin bei der Baumpflege beobachtet, deren Haus ein schönes Blumenbeet im Vorgarten hat.

(22) Hannes Strauß hat die Autofahrerin in einer Werkstatt angetroffen, deren Auto einen großen Schaden im Lack aufweist.

(23) Georg Reise hat die Auszubildende zu einem Gespräch gerufen, deren Launigkeit ein häufiges Thema im Betrieb ist.

(24) Marius Grund hat die Zahnärztin mit einem Kariesbefall aufgesucht, deren Kollegin eine neue Praxis im Hafenviertel eröffnet.

A.2 Experiment 2

We provide the experimental material in the condition determiner = bare, locality = distant. The three other conditions can be constructed by moving the RC after the subject DP and/or changing the determiner of the subject DP to *d-jenig.*

(1) Heute konsultiert der clevere Student für Rückfragen die Professorin, dessen Hausarbeit komplexe Fragen bearbeitet.

(2) Heute besucht der tapfere Junge zur Nachuntersuchung die Ärztin, dessen Bein komplizierte Brüche hat.

(3) Erstmals befragt der erfahrene Polizist zum Tathergang die Verdächtige, dessen Einheit wichtige Beweise sammelt.

(4) Mehrmals belügt der nervöse Zeuge im Kreuzverhör die Richterin, dessen Freund überzeugende Alibis benötigt.

(5) Letztendlich überzeugt der nette Architekt im Gespräch die Stadträtin, dessen Büro günstige Angebote macht.

(6) Normalerweise begleitet der erfahrene Assistent auf Reisen die Abgeordnete, dessen Sachkenntnis internationale Verhandlungen erleichtert.

(7) Zuverlässig kontaktiert der erfolgreiche Anwalt bei Neuigkeiten die Klientin, dessen Kanzlei viel Anerkennung genießt.

(8) Sicherlich kontaktiert der junge Patient in Notfällen die Psychologin, dessen Depression ernste Probleme bereitet.

(9) Mittlerweile übertrifft der talentierte Lehrling bei Feinarbeiten die Meisterin, dessen Ausbildung drei Jahre dauert.

(10) Nachher interviewt der renommierte Journalist für Nachforschungen die Ministerin, dessen Artikel explosive Anschuldigungen beinhaltet.

(11) Deutlich kritisiert der bekannte Redakteur im Magazin die Präsidentin, dessen Publikation viele Leser erreicht.

(12) Oftmals bewundert der exzentrische Designer in Fotoshootings die Fotografin, dessen Label nachhaltige Mode produziert.

(13) Bereitwillig fördert der bekannte Künstler durch Illustrationen die Autorin, dessen Zeichnung surreale Qualitäten besitzt.

(14) Mehrmals ärgert der berühmte Schauspieler durch Improvisationen die Regisseurin, dessen Rolle subtiles Schauspielern erfordert.

(15) Natürlich informiert der sorgfältige Handwerker bei Unfällen die Bauleiterin, dessen Mannschaft gefährliche Dacharbeiten erledigt.

(16) Sofort beauftragt der erfahrene Taxifahrer für Reparaturen die Expertin, dessen Auto einige Mängel aufzeigt.

(17) Fortan vertritt die junge Lehrerin im Notfall den Direktor, deren Klasse exzellente Leistungen erzielt.

(18) Vertragsgemäß benachrichtigt die neue Anwohnerin vor Reparaturarbeiten den Vermieter, deren Wohnung einige Wasserschäden hat.

(19) Überraschend besiegt die begabte Tänzerin im Turnier den Favoriten, deren Technik viel Begeisterung hervorruft.

(20) Nochmals kontaktiert die erfahrene Chefin zur Rekrutierung den Studenten, deren Firma gute Programmierer sucht.

(21) Sogleich beeindruckt die nervöse Bewerberin im Gespräch den Interviewer, deren Lebenslauf einige Lücken aufzeigt.

(22) Erstmalig wechselt die bekannte Rapperin nach Jahren den Manager, deren Karriere neue Größen erreicht.

(23) Häufig fragt die sorgfältige Soziologin bei Recherchen den Historiker, deren Forschung soziale Minderheiten betrifft.

(24) Sorgsam überwacht die neue Pflegerin im Nachtdienst den Patienten, deren Dienstplan wöchentliche Schichtwechsel vorsieht.

(25) Sogleich entzückt die erfahrene Köchin im Testessen den Kritiker, deren Restaurant mehrere Auszeichnungen besitzt.

(26) Derzeit benötigt die erfolgreiche Sportlerin zur Unterstützung den Trainer, deren Verletzung große Probleme bereitet.

(27) Zumeist ehrt die gute Sängerin nach Auftritten den Dirigenten, deren Solo finale Höhepunkte bildet.

(28) Abermals überrascht die talentierte Musikerin in Probeaufnahmen den Produzenten, deren Lied persönliche Erlebnisse widerspiegelt.

(29) Gerne berät die fleißige Verkäuferin bei Unklarheiten den Kunden, deren Laden echte Delikatessen anbietet.

(30) Oft ruft die betagte Rentnerin für Baumarbeiten den Gärtner, deren Anwesen viel Arbeit macht.

(31) Morgens trifft die strenge Managerin zur Besprechung den Angestellten, deren Firma viel Potential hat.

(32) Meistens konsultiert die kritische Investorin vor Aktienkäufen den Bankier, deren Anlage viel Rendite bringt.

Appendix B: Analyses of the Pre-Critical Regions

In self-paced reading, effects are often observed with delayed impact or with prolonged duration. Thus, an effect originating at one sentence region can affect the reading times at the subsequent so-called spillover regions. For the experiments reported in the present paper, we observed effects at the pre-critical region (CR-1) that may have spilled over into the critical region (CR). Our models for the CR therefore included the log RT from the pre-critical region as a predictor. Nevertheless, we hereby report a set of analyses conducted for the CR-1 in both experiments. This is intended to provide full transparency to the reader on any effects that may have affected the results at the CR itself.

All analyses were performed using the data and analysis procedures outlined in the Results sections of Experiments 1 and 2, respectively.

B.1 Experiment 1

The model for the CR-1 yielded a Determiner × Locality interaction, $\hat{\beta} = -0.36$, CrI = [-0.44, 0.27], $P(\beta < 0) = 1$, and main effects of determiner, $\hat{\beta} = 0.14$, CrI = [0.10, 0.19], $P(\beta > 0) = 1$, and locality, $\hat{\beta} = -0.19$, CrI = [-0.29, -0.09], $P(\beta < 0) = 0.99$. The posterior was consistent with a substantial RT increase at this region for the 'complex-close' condition (where the region contained the complex DP *denjenigen Mitarbeiter,* 'the-one employee') compared to the 'bare-close' condition (*den Mitarbeiter,* 'the employee'), $\hat{\beta} = 0.32$, CrI = [0.27, 0.38], $P(\beta > 0) = 1$. This is also reflected in the comparisons between the two 'complex' conditions, $\hat{\beta} = 0.37$, CrI = [0.26, 0.47], $P(\beta > 0) = 1$, such that the analysis supports that the 'complex-close' (*denjenigen Mitarbeiter,* 'the-one employee') condition was read more slowly than the 'complex-distant' (*gesehen,* 'seen') one.

The observed RT contrast can be accounted for by processing difficulties incurred through the morphological complexity and the lower frequency of the complex determiner,

relative to the bare determiner. The *Wortschatz* corpora (http://corpora.uni-leipzig.de) of Leipzig University show that the determiner *derjenige* has the frequency class 12[8] (compared to the bare determiner *der* with the frequency class 0).

B.2 Experiment 2

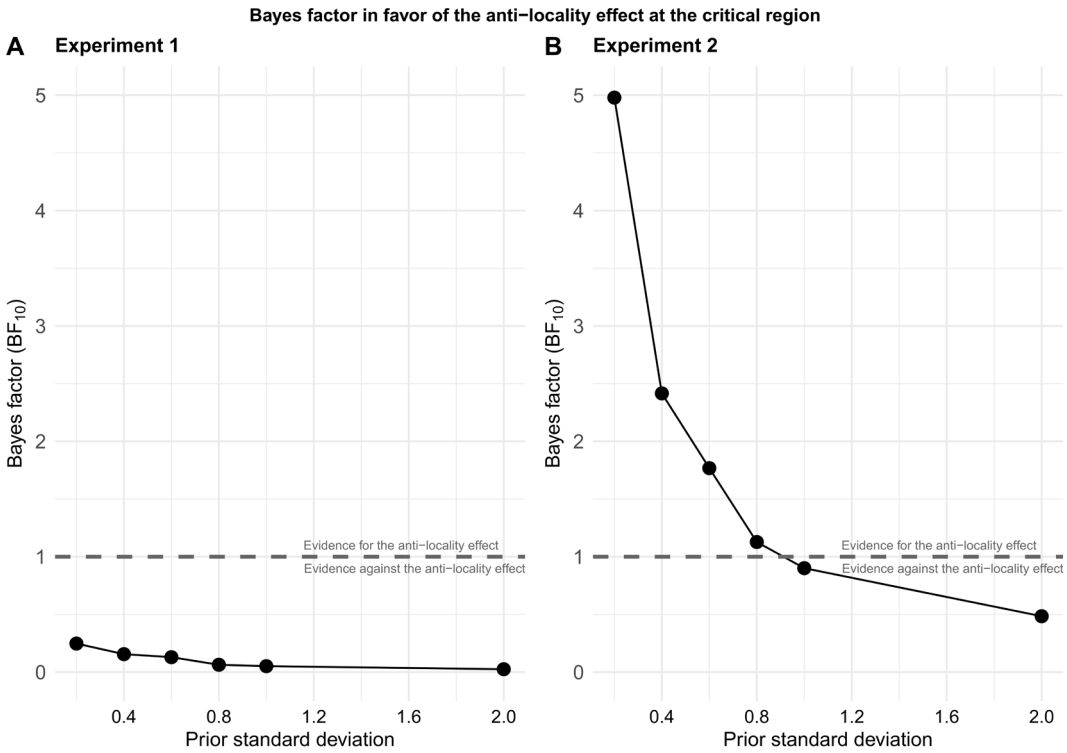
The model for the CR-1 yielded a Determiner × Locality interaction, $\hat{\beta} = -0.07$, CrI = [-0.10, -0.03], $P(\beta < 0) = 1$, and a main effect of determiner, $\hat{\beta} = 0.04$, CrI = [0.03, 0.06], $P(\beta > 0) = 0.99$. Despite the addition of an extra region between the determiner and the noun, the posterior was still consistent with increased processing times at this region for the 'complex-close' condition (where the region was preceded by the complex determiner *d-jenig* and a pre-nominal adjective) compared to the "bare-close" condition, $\hat{\beta} = 0.08$, CrI = [0.06, 0.10], $P(\beta > 0) = 1$. As for Experiment 1, this effect can be explained by the low frequency of *d-jenig*. In contrast to Experiment 1, however, the effect is now much smaller (compare the estimate -0.07 to the previous estimate 0.32) and clear RT differences are indicated only in the comparison between the two 'close' conditions. In fact, the posterior is only weakly supportive of RT differences between the 'complex-close' and 'complex-distant' conditions, $\hat{\beta} = -0.02$, CrI = [-0.06, 0.01], $P(\beta < 0) = 0.92$.

[8] The frequency class of a word FC(w) is calculated based on the occurrence frequency of the word "f(w)" in comparison to the frequency of the most frequent word "fmax," in this case, *der* 'the$_{MASC}$,' and is defined as $FC(w)=[\log[2](f_{max}/f(w))]$. The higher the frequency class, the rarer the word.

Appendix C: Additional Figures

Figure AC.1

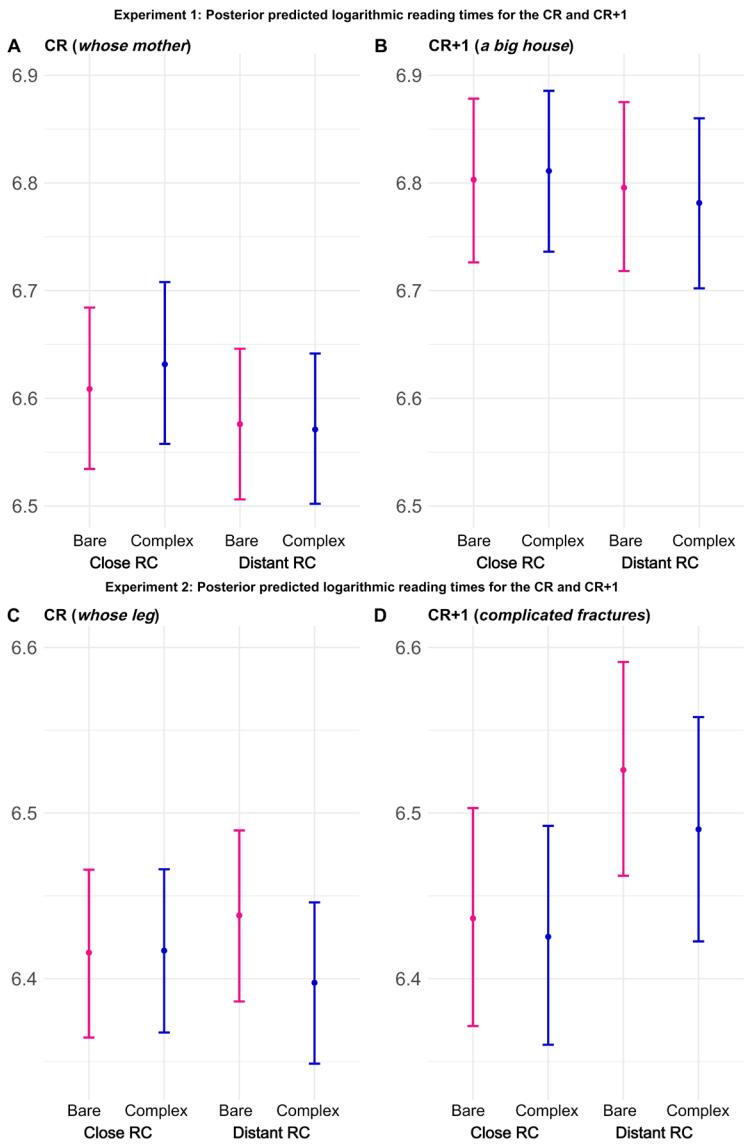
Bayes factor analyses.



Note. Bayes factor analyses for the Determiner x Locality interaction (i.e., the anti-locality effect) at the critical region of Experiments 1 and 2. The plot shows the Bayes factor in favor of the alternative hypothesis versus the null hypothesis (BF_{10}), which was computed for a range of increasingly informative priors, see the Results section of Experiment 2 for details.

Figure AC.2

Posterior predicted logarithmic reading times for the critical and post-critical regions.



Note. Posterior predicted logarithmic reading times for the critical and post-critical regions of Experiments 1 and 2. Error bars indicate the bounds of the 95 % credible interval.