ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 4
https://doi.org/10.18148/zs/2026-2013

Lisa Schäfer (corresponding author), Saarland University, Germany, lisa.schaefer@uni-saarland.de
Ingo Reich, Saarland University, Germany, ingo.reich@uni-saarland.de

The identity condition of right node raising in German

Experimental evidence

Abstract: This article examines the identity condition of right node raising (RNR) in German, focusing on how morpho-phonological mismatches in the verbal domain between the target and its correlate affect the acceptability and processing of RNR. We relate experimental evidence from a cloze study, three rating studies, and a reading time study to the predictions of the central theoretical approaches to RNR mismatches that postulate (gradual) phonological, lexeme, or stem identity. Comparing RNR to gapping and full forms as baselines shows that while complete identity between target and correlate yields the highest ellipsis rates, mismatches do not result in acceptability or processing penalties. However, speakers are less likely to use RNR in the cloze task as the mismatch becomes increasingly morphologically marked, whereas for gapping, we do not observe such a staircase-like pattern. While the rating and processing data of RNR align best with lexeme or stem identity, the cloze data are more in line with modified gradual phonological identity. Our findings suggest that RNR in German is more flexible than previously assumed, with acceptability and processing being tolerant of mismatches, but the cloze data showing greater sensitivity.

Keywords: ellipsis, right node raising, gapping, mismatches

Cite as: Schäfer, Lisa & Ingo Reich. 2026. The identity condition of right node raising in German: Experimental evidence. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2013.
Received: 16 October 2025; **accepted:** 27 August 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

While discussing a schema to derive constituent coordination from the coordination of two sentences, later famously known as "conjunction reduction," Chomsky (1957: 35–36) mentions in a footnote that structures like (1) cannot be derived by this rule, simply because the seemingly coordinated expressions *John enjoyed* and *my friend liked* are not syntactic constituents. In fact, he even considers the possibility that structures like (1) should be regarded as semi-grammatical, since they show properties like long pauses (between *liked* and *the*) and contrastive stress that "normally mark the reading of non-grammatical strings" (Chomsky 1957: 36).

(1) *John enjoyed and my friend liked the play.* (Chomsky 1957: 36)

However, it did not take long for a broad consensus to emerge that structures like (1) need to be accounted for in grammar in one way or another. One of the first accounts to this effect came from Ross (1986 [1967]: 108) who proposes to derive (1) from the coordination in (2) in terms of across-the-board (ATB)-movement: since both sentences contain identical copies of the object *the play*, the object can be moved out of both conjuncts at the same time, and adjoins as a single copy to the right of the coordination; see (2a). Since in this case ATB-movement only affects the rightmost (major) constituent and targets a right-peripheral position, it is called right node raising (RNR) in Postal (1974) and has been known by this name ever since (see also Sabbagh 2007 for a more recent proposal in terms of ATB-movement.).

(2) *John enjoyed the play, and my friend liked the play.*
   a. [[*John enjoyed* t_1] *and* [*my friend liked* t_1]] *the play*$_1$ (ATB-movement)
   b. [*John enjoyed* ~~*the play*~~] *and* [*my friend liked the play*] (ellipsis)
   c. [[*John enjoyed* 1] *and* [*my friend liked* 1]], 1 = *the play* (multi-dominance)
   d. [[*John enjoyed*] *and* [*my friend liked*]] *the play* (coordination)

Subsequent research has produced convincing arguments that RNR is not the result of movement: RNR is not subject to any relevant island constraints, nor does it respect constituency; see, e.g., Hartmann (2000), for relevant discussion. This is why many researchers have argued that structures like (1) are not the result of movement, but ellipsis. And to better reflect the analysis, other terms have been introduced like "leftward deletion" or "backward conjunction reduction," emphasizing the directionality of RNR in coordination, or "right-peripheral ellipsis," highlighting that RNR only targets strings at the right periphery of non-final conjuncts; see, e.g., Klein (1981), Wilder (1997), Hartmann (2000), for early proposals, and (2b) for illustration (strikeouts indicate ellipsis). More recently, observations that relate to the interpretation of adverbials and agreement properties have motivated multi-dominance (MD) analyses, which start from the assumption that, in principle, one and the same node may be dominated by two different nodes in syntax. With respect to structures like (1), it is argued that the object *the play* is in fact jointly dominated by the verb phrase (VP) of the first and the VP of the second

sentence; see, e.g., McCawley (1982), Wilder (1999), and de Vries (2009), among many others. In other words, the two conjuncts in (2) do not contain identical, albeit distinct copies of *the play* but the very same object. This is indicated in (2c) with the marker [1] that refers to the same object in all occurrences, namely *the play*. This marker is set in gray in the first conjunct to clarify that the object is only spelled out in the final conjunct. Last but not least, psycholinguistic considerations (see Philips 2003), more flexible frameworks (see Steedman 2000), and pragmatic considerations (see Ott 2025) open up analyses in terms of syntactic or pragmatic coordination see (2d), thus closing the circle that began with Chomsky.[1]

The aim of this contribution is not to add another proposal to the already existing accounts, nor to give a thorough discussion of their merits. The primary objective is empirical in nature: to provide crucial empirical data to underpin and support the discussion of identity constraints in German RNR. Existing accounts of RNR make sometimes more, sometimes less clear predictions concerning the nature of the gap in the first conjunct: ATB-movement accounts require morpho-phonological identity of the expression to be moved. Coordination-based approaches make, in principle, similar predictions, since it is one and the same expression that relates to the two conjuncts in the coordinate structure. In MD approaches, the picture is somewhat less clear: on the one hand, the node related to the gap is the same node as the node spelled out in the second conjunct. On the other hand, the phenomenon of "summative agreement" reported in Yatabe (2003), Kluck (2009), Grosz (2015), and others shows that the form of the finite verb may change from singular to plural; see (3) from Shen (2019).

(3) *John's glad that Mary, and Bill's happy that Sue, has/have been to Cameroon.* (Shen 2019: 2)

Still, all three approaches predict—without further assumptions—strict identity conditions for RNR. The deletion approach is more flexible in this respect: the preconditions for deletion could be stated, in principle, at any of the levels of phonology, morphology, syntax, semantics, or pragmatics, or any combination of these levels, depending on the specific empirical behavior of the ellipsis at hand.

In this paper, we will present experimental evidence from a cloze study, three rating studies, and a reading time study on verbal RNR in German that aim to test the identity conditions imposed on the target (gap) and its overt correlate (pivot). German is a great testing ground in this respect, since in German-dependent clauses, the finite verb typically ends up in sentence-final position and is thus always part of the characteristic gap in RNR structures; see (4).

[1] It should be noted that there are also mixed approaches to RNR that argue that RNR has two different potential sources: ellipsis and MD; see, e.g., Barros & Vicente (2011) and Belk et al. (2023). Where RNR is derived by ellipsis, the predictions are those of the ellipsis approach. And where RNR is derived by MD, the predictions are those of the MD approach. Since it is not fully clear (independently of the observed properties) when ellipsis and when MD applies, we do not discuss these approaches in any further detail here.

(4) *Er glaubt, dass Bernd in Berlin* (*arbeitet*) *und Anne in London arbeitet.*
he thinks that Bernd in Berlin (works) and Anne in London works
"He thinks that Bernd works in Berlin and Anne in London."

This allows us to systematically vary person and number features of the target and the correlate, thereby generating stimuli with different degrees of mismatches. But before we go into more detail about the logic behind the experiments, we will give a (probably too) brief overview of positions and discussions in the literature on identity constraints in RNR.

2 Identity constraints in RNR: a brief overview

In an early squib, Eisenberg (1973) argues that RNR tolerates mismatches in syntactic features like person and number if the target and the correlate are identical in phonological form, i.e., if they are syncretic (5a). If, however, a mismatch in the relevant features results in different forms, RNR is judged as ungrammatical (5b).

(5) a. *weil wir das Haus (kaufen.*1PL*) und die Müllers den Garten kaufen.*3PL
since we the house (buy.1PL) and the Millers the garden buy.3PL
(Eisenberg 1973: 418)

b. **weil Franz das Haus (kauft.*3SG*) und ich den Garten kaufe.*1SG
since Franz the house (buy.3SG) and I the garden buy.1SG
(Eisenberg 1973: 418)

From this contrast, Eisenberg concludes that phonological identity is the relevant constraint in RNR structures, as opposed to identity in morphosyntactic features. It should be noted, however, that he explicitly excludes case from this generalization, since "most people can also say" (Eisenberg 1973: 419) *mit (Kindern.*DAT*) und ohne Kinder.*ACC "with (children.DAT) and without children.ACC." In the following, we will refer to this approach as the *phonological identity (PI) approach.*

Klein (1993: 774) essentially follows Eisenberg's generalization but adds that the target and the correlate need to mean essentially the same (in a sense still to be specified). A similar point is made in Chaves (2014: 869), who argues, based on data from the literature, that RNR is blocked if phonologically identical expressions differ in meaning (6).

(6) a. **Mary fed* (*the lamb*) *and Tom enjoyed the lamb.* (Chaves 2014: 869)
b. **Robin swung* (*an unusual bat*) *and Leslie tamed an unusual bat.* (Chaves 2014: 869)

Like Eisenberg (1973), Chaves (2014) accepts mismatches in features such as case, if the target and the correlate are identical in form (i.e., syncretic). However, in contrast to Eisenberg (1973), he also accepts mismatches in number that result in different verb forms (7a), though his judgments seem to depend on the ordering of target and correlate (7b).

(7) a. *Tom thinks that the shoes* (*are …*)*, but Mary thinks that the coat is too expensive.*
Chaves (2014: 871)

b. ??*Tom thinks that the coat* (*is …*)*, but Mary thinks that the shoes are too expensive.*
Chaves (2014: 871)

Chaves (2014: 870) concludes from this data that "the identity conditions imposed by backward periphery deletion pertain to morph stems." While in some cases identity in phonological form is still taken to be a prerequisite for RNR (e.g., in case mismatches, cf. **er findet (Männer.*ACC*) und hilft Männern.*DAT "he finds men.ACC and helps men.DAT;" our judgment), this is not generally assumed, as the number mismatch in (7a) illustrates. Chaves (2014) speculates that the acceptability of form mismatches might depend on the strength of the semantic contribution of the feature (weaker semantic contributions resulting in increased acceptability). Thus, in a way, Chaves (2014) still assumes some kind of phonological identity but also considers a semantically based mechanism that allows for some degree of flexibility in phonological form. We will refer to this as the *morph stem form identity (MSI) approach.*

Like Chaves (2014), Christ (2011) also aims at weakening the phonological identity requirement, but he does so in a very different way. Christ (2011: 383) argues that the acceptability of RNR decreases the more the phonological form of the correlate deviates from that of the target. He illustrates this with RNR that targets nouns (8); the judgments are his:

(8) a. *Hans hat einen* (*Hund.*SG) *und Peter hat mehrere Hunde.*PL (Christ 2011: 383)
Hans has a (dog.SG) and Peter has several dogs.PL

b. ??*Hans hat eine* (*Maus.*SG) *und Peter hat mehrere Mäuse.*PL (Christ 2011: 383)
Hans has a (mouse.SG) and Peter has several mice.PL

c. **Hans hat einen* (*Atlas.*SG) *und Peter hat mehrere Atlanten.*PL (Christ 2011: 383)
Hans has a (atlas.SG) and Peter has several atlases.PL

In (8a), the plural correlate only differs in one respect, namely the plural suffix *-e*. According to Christ, RNR is still fine, i.e., RNR seems to tolerate a single deviation; this is also taken to be true of Umlaut as in *Vögel* "birds" vs. *Vogel* "bird." But if the plural shows Umlaut plus a plural suffix like *-e* as in *Mäuse* "mice" vs. *Maus* "mouse," see (8b), the two deviations from the singular seem to add up, and the resulting RNR structure is judged as significantly less acceptable. And finally, if the plural form even alters the stem of the target, as is the case in *Atlanten* "atlases" vs. *Atlas* "atlas," see (8c), RNR is judged as fully ungrammatical. This approach therefore assumes a gradual, even staircase-like decline of acceptability as a function of phonological deviation. We will refer to this as the *gradual phonological identity (GPI) approach.*

It is surprising that the discussion of identity constraints in RNR is almost exclusively based on introspective data, even though judgments are not always very clear. A notable exception is Shiraïshi et al. (2019). They report on two acceptability studies, one on French and the other on English, that test the acceptability of full matches and of syncretic and non-syncretic forms with a temporal mismatch (which is compared to *do so*-anaphora as a baseline). Given the above discussion on identity constraints, their results are unexpected: in English, both syncretic and non-syncretic mismatches are judged significantly better than the corresponding matches. This is particularly surprising given that Chaves (2014: 870) judges tense mismatches as ungrammatical. In French, mismatches are only marginally worse than matches. A comparison with the grammatical and non-grammatical controls shows that matches as well as mismatches need to be considered as grammatical. Shiraïshi et al. (2019) conclude that variation in both morphological and phonological form is generally tolerated in RNR and that identity is only required on the level of lexemes.[2] In the following, we will call this the *lexeme identity (LI) approach*.

Like Shiraïshi et al. (2019), we will also test full matches against morphological mismatches with syncretic and non-syncretic forms. Unlike Shiraïshi et al. (2019), however, we will use mismatches in number and person since they easily allow for a stepwise increase in morphological deviation required to test the GPI approach: in the first mismatch condition, target and correlate are syncretic forms, in the second condition, they differ in their ending, and in the third they differ in addition in the realization of the vowel within the stem. These four conditions are schematically illustrated in Table 1 using *lesen* "to read" as an example.

Table 1: Design of the four conditions: *identical*, *syncretic*, *ending*, and *ending+vowel*

			mismatch in		mismatch in		
	TARGET	CORRELATE	PERSON	NUMBER	ENDING	VOWEL	CONDITION
1	*(er) liest*	*(sie) liest*	–	–	–	–	identical
2	*(wir) lesen*	*(sie) lesen*	+	–	–	–	syncretic
3	*(ich) lese*	*(sie) lesen*	+	+	+	–	ending
4	*(ich) lese*	*(er) liest*	+	–	+	+	ending+vowel

[2] It is worth mentioning here that this seems to be also true of the proposal presented in Hartmann (2000), which essentially assumes that RNR needs to meet a contrastive focus condition. This focus condition is implemented within alternative semantics, and thus—quite like the lexeme/stem approaches—semantic in nature. Hartmann's (2000) proposal seems to be more general, though, since it concerns not only single lexemes but any sequence of lexemes that is supposed to complete the first conjunct and is located at its right edge. We leave it to future research whether or where exactly the contrastive focus approach and the lexeme/stem approaches diverge in predictions.

Without further assumptions, ATB-movement, MD, and coordination (ATB-MD-C) approaches all predict that RNR should only be grammatical, if target and correlate are identical, simply because one and the same object relates to two different positions. This is, of course, a somewhat simplified representation. As already mentioned, there is the phenomenon of summative agreement that has been dealt with within MD analyses; and there is the phenomenon of conjunct agreement, where agreement is confined to the closest conjunct; see, e.g., Aoun et al. (1994), and more recently Himmelreich et al. (2024). Himmelreich et al. (2024: 77) explicitly state that syncretic forms improve person mismatches in last conjunct agreement in German. This is why we would like to leave open what exactly the predictions of these approaches are regarding condition 2. Conditions 3 and 4 are, for sure, predicted to be ungrammatical, since mismatches in person are usually taken to be ungrammatical in German; see Fuß (2018). The PI approach predicts that only conditions 1 and 2 are acceptable in German. While the PI approach predicts conditions 3 and 4 to be equally unacceptable, the GPI approach predicts a staircase-like decline in acceptability. With respect to the MSI approach, predictions are, again, not fully clear. The judgments concerning (7) and the remarks concerning number mismatches suggest that in the MSI approach the possibility and acceptability of conditions 3 and 4 is at least not ruled out. The LI approach, on the other hand, clearly predicts conditions 3 and 4 to be as acceptable as conditions 1 and 2 are. The predictions of the above-discussed approaches are summarized in Table 2.

Table 2: Predictions of the above-discussed approaches with respect to conditions 1–4

	C1: IDENTICAL	C2: SYNCRETIC	C3: ENDING	C4: ENDING+VOWEL
ATB–MD–C	✓	(*)	*	*
PI approach	✓	✓	*	*
GPI approach	✓	✓	✓	??
MSI approach	✓	✓	✓	✓
LI approach	✓	✓	✓	✓

To determine whether differences related to morphological deviations are specific to RNR, we need a reference point to compare RNR to. Gapping is ideal for this role. In verb-final structures like those we used to test the RNR mismatches, see (4), we can straightforwardly create analogous gapping examples like (9) where the gap occurs at the end of the second conjunct. In this case, gapping is superior to a full form with two overt verb forms because the full forms might be degraded for being redundant (an effect that we found in previous work, e.g., Schäfer et al. 2021). More importantly, there is a broad consensus in the literature that gapping tolerates mismatches and that all conditions presented in Table 1 are fully acceptable under gapping; see Eisenberg (1973), Klein (1993), Hartmann (2000), and many others. As a welcome side-effect of this investigation, these intuitions on gapping will be corroborated by empirical data.

(9) *Er glaubt, dass Bernd in Berlin arbeitet und Anne in London (arbeitet).*
he thinks that Bernd in Berlin works and Anne in London (works)
"He thinks that Bernd works in Berlin and Anne in London."

In the following sections, we will present empirical data on the production, perception, and processing of RNR mismatches. In Experiment 1, we begin by using a cloze study to investigate whether speakers of German accept RNR mismatches and whether the degree of morphological deviation has an influence on the ellipsis rate. In Experiments 2 and 3, we explore the perception of mismatches by collecting acceptability judgments for all four conditions, once with gapping, once with the full form as baseline. In Experiment 4, we focus on the most extreme contrast (between condition 1 and condition 4 in Table 1) in a third rating study. Finally, Experiment 5 is a reading time study in which we investigate the processing of RNR mismatches.

3 Experiment 1: production study RNR and gapping

As we saw above, RNR mismatches have been widely discussed in the theoretical literature, but the existing work has focused mostly on introspective judgments. We aim at expanding the sparse experimental work with a cloze study, which offers a controlled way of assessing how acceptable RNR mismatches are to speakers. By giving participants the option to either fill in the gap resulting from RNR or leave it blank, we can assess whether mismatches are treated as natural and retained or avoided in favor of full forms. In this way, Experiment 1 provides a production-oriented empirical perspective that is currently missing from the literature.

Our cloze study has the form of a 2 × 4 design, crossing EllipsisType and Identity. For EllipsisType, the baseline gapping is compared to RNR. Identity has the four levels sketched in Table 1: full identity of the two verb forms, syncretism, deviation in the inflectional ending, and deviation in the ending plus a change of the stem vowel. We measured the ellipsis rate, i.e., the proportion where participants left the gap empty relative to all (reducible) answers provided.

As sketched in Table 2, only the MSI and the LI approaches predict that all four levels of verbal mismatches under RNR are, in principle, fine. This means that under these approaches we expect no difference in the ellipsis rate across the Identity conditions. This is different for the PI approach, under which only phonologically matching conditions are predicted to be acceptable. The ending and ending+vowel conditions consequently should yield a lower ellipsis rate than the identical and syncretic conditions. The GPI approach predicts that the ellipsis rate varies as a function of the degree of morphological deviation: the more the forms differ, the lower the ellipsis rate. Concretely, based on Christ's (2011: 383) judgments for nominal RNR mismatches, the identical and syncretic conditions should essentially pattern with the ending condition, whereas a simultaneous deviation in ending and stem vowel should be degraded and

result in a lower ellipsis rate. However, in Christ's nominal example, adding an ending only represents a number mismatch, whereas in our condition, it is a mismatch in number *and* person. For this reason, the change in ending could represent a greater morphological deviation here, to be reflected in a reduced ellipsis rate compared to phonological identity. The ATB-MD-C approaches predict, without further assumptions, all mismatches to be degraded with the potential exception of syncretic forms. So at least the last two IDENTITY conditions, ending and ending+vowel, should yield lower ellipsis rates according to these approaches.

3.1 Materials

We constructed 24 items in the form of complex sentences, such as (10), that allow for RNR or gapping. They contain an embedded verb-final clause with two verb phrases coordinated by *und* "and." In the potential RNR conditions, we replaced the finite lexical verb at the end of the first conjunct with an empty text box (indicated with underscores in (10)), in the potential gapping conditions, it was the verb at the end of the second conjunct.[3] The subjects in the embedded clause were varied so that under lemma identity, the verb forms of the two conjuncts would either be identical (10a), syncretic (10b), differ in their ending (10c), or differ in ending and their stem vowel (10d).

(10) "For the presentation, it is agreed that…"

a. *Anna den Zeitschriftenartikel (___| liest) und Jan das Kapitel aus dem Buch (liest |___)*
Anna the journal.article read.3SG and Jan the chapter out the book read.3SG (identical)

b. *wir den Zeitschriftenartikel (___| lesen) und Anna und Jan das Kapitel aus dem Buch (lesen |___).*
we the journal.article read.1PL and Anna and Jan the chapter out the book read.3PL (syncretic)

c. *ich den Zeitschriftenartikel (___| lese) und Anna und Jan das Kapitel aus dem Buch (lesen |___).*
I the journal.article read.1SG and Anna and Jan the chapter out the book read.3PL (ending)

d. *ich den Zeitschriftenartikel (___| lese) und Jan das Kapitel aus dem Buch (liest |___).*
I the journal.article read.1SG and Jan the chapter out the book read.3SG (ending+vowel)

[3] To increase readability, we indicate the manipulation of ELLIPSISTYPE by placing the verb in parentheses in both conjuncts.

The items were combined with 72 fillers, all of which were coordination structures using *und* "and." Of these, 24 were from another study and included a context sentence and a target sentence featuring potential RNR structures with nominal antecedents and targets; inserting a word in these cases was optional, as with the critical items. The remaining 48 fillers were designed to present participants with all possible scenarios for filling the gap: optional filling, obligatory filling, and instances where no filling was possible. This should enhance the naturalness of the task and prevent participants from always filling in something or always leaving the gap blank. One third of these fillers contained gaps where insertion was optional, e.g., adding an adjective between article and noun. Another third included gaps that required completion, e.g., providing the object of a transitive verb. The final third comprised gaps where inserting an element would result in an ungrammatical structure, such as between a pronominalized subject and its corresponding finite verb in a verb-second clause. The last two categories also served as attention checks; see below. For each of the 16 fillers in a given category, half were accompanied by a context sentence like the 24 RNR fillers, while the other half did not include a context, aligning with the critical items.

3.2 Procedure

We recruited 241[4] native German speakers between the ages of 18 and 50 from the crowdsourcing platform Prolific (www.prolific.com) and paid them £3.69.[5] The experiment was implemented using the LimeSurvey survey presentation software (Limesurvey GmbH). The participants' task was to read the materials and decide whether to fill in a text box in the critical sentence. In the instructions, they were given examples for optional, obligatory, and impossible filling together with a brief explanation. Each experimental list started with three fillers, one of each category. The remaining fillers and the items were presented in pseudo-randomized order, ensuring that no two trials of the same type immediately followed each other. Each participant saw each token set only once and in one condition and each condition equally often. This way, we collected at least 30 answers per token set and condition.

3.3 Results

We used 16 fillers that required obligatory completion and 16 fillers where completion should not be possible as attention checks. Participants were categorized as inattentive if they did not complete at least 50% of the obligatory fillers, or if they completed 50% or more of the fillers

[4] Since one participant did not claim their reward on Prolific, we have data from 241 instead of the planned 240 participants.

[5] In all experiments reported in this paper, compensation corresponds to the German minimum wage at the time of the experiment, calculated based on the estimated time required to complete the experiment.

intended to remain unfilled due to ungrammaticality.[6] Based on these criteria, two participants who did not fill at least 50% of the mandatory fillers were excluded.

We manually reviewed the remaining 239 participants' answers and pre-processed the 5736 data points.[7] In a first step, we excluded any meaningless responses, e.g., English words like *first*, and responses that did not semantically or grammatically match the context, such as (11), where the singular verb does not agree with the plural subject.

(11) *Für den Kongress in Brüssel ist es praktisch, dass ich Französisch spreche*
for the congress in Brussels is it practical that I French speak
und Vanessa und Lena Niederländisch ***kann****.*
and Vanessa and Lena Dutch can.3SG

This resulted in a loss of 1.88% of the data. Grammatical correctness was required to ensure that participants accurately recognized the IDENTITY manipulation; entering a mismatching verb form could indicate a lack of attention to this aspect. Most agreement errors occurred when participants provided a verb form for a coordinated subject such as *Vanessa und Lena*. In several instances, apparently cursory reading resulted in only the second name, *Lena*, being identified as the subject, leading to the use of the 3rd singular instead of the 3rd person plural, as in (11).[8]

The remaining data were divided into three categories: ellipsis (60.18%), full form (20.43%), and other (19.39%). When the participant left the text box blank or did not enter a verb but only an adverb or a noun modifier,[9] this was formally classified as *ellipsis* (without implying that the gap should be analyzed as an elliptical reduction). A response was classified as *full form* when the participant entered the (correctly inflected) identical verb found in either the antecedent or

[6] Some participants provided meaningful responses even in these cases, which was not anticipated, such as placing lassen "to let" between a lexical and a modal verb (*den neuen Boden verlegen lassen können* "to be able to have the new floor laid"). These responses were not classified as ungrammatical or considered indicative of inattentiveness.

[7] Each data point was independently annotated by two student assistants and one of the authors. In instances where annotations differed, the final category was determined through joint discussion and agreement between both authors. We would like to thank Canan Borhofen and Sarah Galling for their help.

[8] It should be noted that some of these sentences, such as (i), can also yield a grammatical interpretation when the first conjunct's object is animate. Specifically, a part of the second conjunct's coordinated subject, e.g., *Daniel*, may be analyzed as belonging to the first conjunct's object, i.e., *ein paar Freunde und Daniel*, and then being extraposed. As a result, the second conjunct would consist solely of *Tobias einige Arbeitskollegen*, wherein *trifft* is an appropriate continuation. Given that it is uncertain whether participants interpreted these constructions in this manner or generated a mismatch analogous to (11) above, these sentences were excluded from analysis.

(i) *Heute Abend ist geplant, dass ich ein paar Freunde treffe und Daniel und Tobias einige*
today evening is planned that I a few friends meet and Daniel and Tobias some
Arbeitskollegen ***trifft****.*
work.colleagues meets

[9] For example, several participants specified the definite noun *Gewerkschaft* "trade union" by adding the proper name *Verdi*, which denotes the second-biggest German trade union.

postcedent. All other grammatically correct responses were categorized as *other*. This primarily involved alternative verbs, including synonyms such as *schauen* for *sehen* "to look." Additionally, participants occasionally employed the same verb lemma while varying tense or mood or embedding it under a modal verb (0.82% of the data). We did not consider them *full forms* because they cannot be readily omitted without losing this additional meaning aspect.[10]

Figure 1 shows the distribution of the three answer categories for each condition. Both gapping and RNR have the highest ellipsis rates in the identical condition. For gapping, the other three categories each show an ellipsis rate 8 to 10% lower, with no obvious difference between them. In RNR, the ellipsis rates seem to decrease stepwise from syncretic to ending to ending+vowel, while the full form rates increase accordingly. In the ending+vowel condition of RNR, participants produced slightly more full forms than ellipses. The rates of other answers are relatively stable across conditions from 14% to 24%. We therefore focused only on the ellipses and full forms for the following analysis and this way ensure that our reference quantity for ellipsis consists only of reducible sentences.

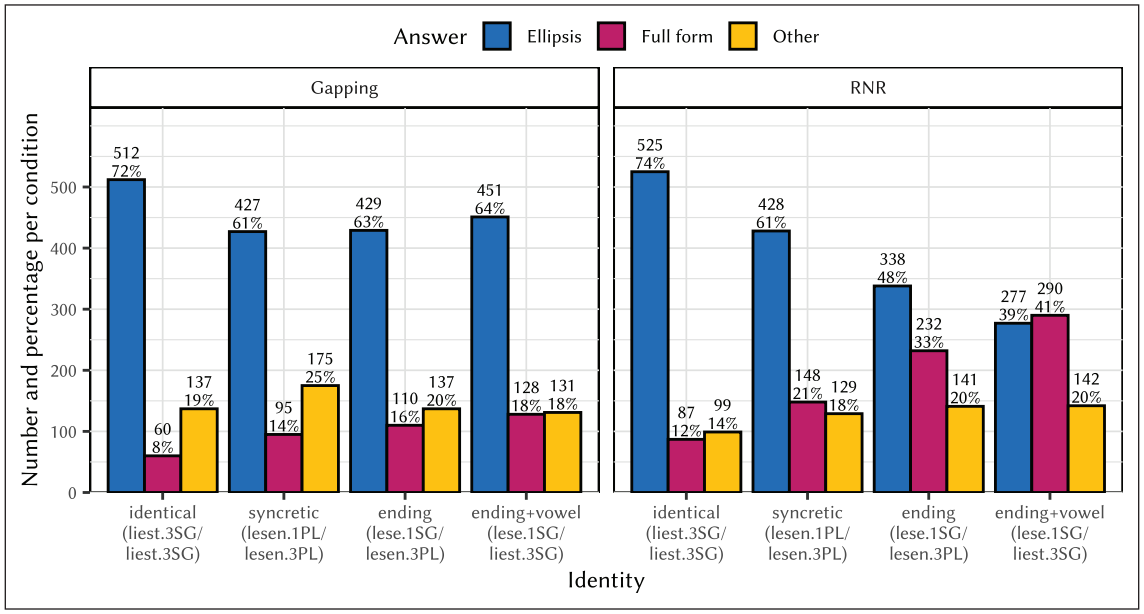


Figure 1: Absolute numbers and percentages of produced answer types per condition in Experiment 1

[10] Cases like (i) are interesting, though, since the participant replaced a potential gapping structure by an RNR construction with an elided modal verb in the first conjunct.

(i) [...] *dass wir Ski fahren und Jana und Niklas Snowboard* ***fahren wollen***.
that we ski drive and Jana and Niklas snowboard drive want

We used a logistic mixed regression model (package lme4, Bates et al. 2015) in R (R Core Team 2025). The model predicted the binary Answer (ellipsis vs. full form) with ellipsis as reference level from the dummy-coded EllipsisType (gapping as baseline with 0, RNR as 1) and from Identity. Identity was forward-coded to allow for finding a potential gradation effect: the first contrast identical.rest compared the identical condition (coded as 0.75) to the other conditions (each coded as −0.25), the contrast maphon.mmphon compared the phonologically matching conditions (identical and syncretic, coded as 0.5) to the phonologically mismatching conditions (ending and ending+vowel, coded as −0.5), and the contrast rest.vowel compared the identical, syncretic and ending conditions (each coded as 0.25) to the ending+vowel condition (coded as −0.75), i.e., to the condition with the highest deviation between antecedent and target. We also included the *z*-transformed Position of the trial in the experiment to control for potential habituation effects and considered the possible interactions between EllipsisType and Identity to account for RNR-specific effects. The model contained random intercepts for participants and items and by-participant and by-item random slopes for EllipsisType.[11] More complex models resulted in singular fits. Table 3 shows the fixed effects of the model.

Table 3: Fixed effects in the logistic mixed regression model of Experiment 1

Fixed effect	Est.	SE	*z*	*p*	
Intercept	−2.79	0.27	−10.30	< 0.001	***
EllipsisType	1.61	0.23	6.93	< 0.001	***
identical.rest	−1.17	0.25	−4.67	< 0.001	***
maphon.mmphon	−0.42	0.22	−1.92	0.055	.
rest.vowel	−0.05	0.20	−0.23	0.81	
Position	−0.30	0.05	−5.93	< 0.001	***
EllipsisType: identical.rest	−0.004	0.32	−0.01	0.99	
EllipsisType: maphon.mmphon	−0.71	0.28	−2.55	< 0.05	*
EllipsisType: rest.vowel	−0.69	0.26	−2.68	< 0.01	**

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

The simple effects of the Identity contrasts indicate how the reference level of EllipsisType, i.e., gapping, behaves with respect to the Identity manipulation: the model contains a significant simple effect of identical.rest (z = -4.67, p < 0.001). It indicates that for gapping, participants produced significantly more ellipses in the identical condition than in all three mis-

[11] The formula was Answer ~ EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) + Position + (1 + EllipsisType | Subject) + (1 + EllipsisType | Item).

match conditions. The effect of MAPHON.MMPHON is marginal (z = −1.92, p = 0.055), suggesting tentatively that the phonologically matching conditions lead to a higher ellipsis rate for gapping than the mismatches. Based on the visual inspection of Figure 1, this tendency seems to be driven by the identical condition. REST.VOWEL has no significant effect, i.e., there is no difference for gapping between the ellipsis rates of the ending+vowel condition compared to the rest.

The interaction terms between the IDENTITY contrasts and ELLIPSISTYPE indicate the deviation of RNR from gapping with respect to the IDENTITY manipulation. While there is no significant difference between gapping and RNR when comparing the identical condition to the rest (z = 0.01, p = 0.99), the ellipsis rate becomes significantly lower for RNR of a verb form with a different ending and a different ending plus a vowel change as compared to gapping (z = 2.55, p < 0.05) and again even lower when comparing the most extreme morphological deviation (ending+vowel) to the rest (z =2.68, p < 0.01).

Due to the forward coding of the predictor IDENTITY, the significant effect of ELLIPSISTYPE cannot be straightforwardly interpreted. While it suggests a statistically stronger preference for ellipsis across all gapping conditions compared to RNR, we refrain from concluding a general preference for ellipsis under gapping because the ellipsis rates in the identical condition are the same for gapping and RNR, and the effect seems to be mainly driven by the interaction between ELLIPSISTYPE and IDENTITY. The significant main effect of POSITION (z = -5.93, p < 0.001) shows that participants provided fewer ellipses as the experiment progressed.

As a final point of our analysis, we would like to mention the notable differences in the ratio of ellipses and other verbs between token sets and participants. Four token sets had a ratio of under 4% of other verbs, while four further reached up to 40% of other verbs. The latter are mainly token sets where participants seemed to dislike the given verb, e.g., replacing *sehen* "to see" with the synonym *schauen*. This suggests a preference to correct the verb over avoiding redundancy by producing ellipsis. There were six participants who consistently used ellipses, whereas another six always used the full form. No participant only filled in other verbs. While the strategy of only providing full forms could be a result of the participants wanting to fulfil the task "correctly", only producing ellipses might indicate that these participants are most tolerant to morphological mismatches, i.e. for them, LI seems to suffice. The statistical results remain comparable when we disregard the participants without variation.

3.4 Discussion

The purpose of our first experiment, a cloze task, was to find out if speakers accept and maintain verbal RNR mismatches or if they correct them by producing the corresponding full form. We compared the ellipsis rates of (potential) RNR structures to those of (potential) gapping structures, which served as our baseline.

The results indicate first that mismatches under gapping and RNR are frequently produced or, more precisely, not avoided by converting the elliptical utterance to a full form. Gapping exhibits an ellipsis rate of over 60% in all four IDENTITY conditions, and even RNR still shows almost 40% ellipsis in the condition with the lowest rate. For our baseline gapping, this result confirms the prevailing research opinion that gapping generally tolerates mismatches in person and number. For RNR, the result challenges the ATB-MD-C approaches, which predict that mismatches are ungrammatical and therefore not produced by speakers.

The second result is that both gapping and RNR exhibit higher ellipsis rates when there is complete identity between correlate and target compared to mismatch conditions. There is a significant mismatch penalty even in the syncretic condition and even for the baseline gapping, which, according to the literature, should tolerate any form of mismatch in person and number. The data show that while speakers produce a lot of mismatches, they display sensitivity to morphosyntactic deviations in this type of task, though they do not distinguish between different degrees of deviation in gapping.

The picture is different for RNR, where a staircase-like pattern emerges. As the morphological difference between target and correlate increases, speakers are more likely to produce the corresponding verb rather than omit it. This pattern is incompatible with the PI approach, which predicts the syncretic condition to behave like the identical condition, and with the MSI and the LI approaches, which predict all mismatches to be equally possible. At first glance, the staircase-like pattern seems to correspond to the predictions of the GPI approach: a gradual decrease in acceptability relative to the degree of morphological deviation. There are, however, two important qualifications: first, Christ (2011) considered syncretic forms and forms only differing in one aspect (like the adding of the nominal plural affix *-e*) to be acceptable as well, whereas our data already show a reduced ellipsis rate for these conditions. Second, Christ marks forms such as our ending+vowel condition with two question marks, i.e., as highly questionable. Nevertheless, the participants accepted these mismatches in almost 40% of the cases in the cloze task and did not change them to the full form. The GPI approach, if we base it literally on Christ's (2011) judgments on nominal RNR mismatches, is thus too restrictive on the one hand and too permissive on the other to explain our data in its entirety.

While the present experiment used a cloze task rather than genuine production, which may have increased participants' sensitivity to inconsistencies, and while the observed differences between items and participants suggest that factors such as verb semantics and the development of individual strategies may also have contributed to the result, we nevertheless find the clear staircase-like pattern for the ellipsis rate of RNR. In the following three experiments, we investigate whether the varying ellipsis rates are also reflected in differences in acceptability. We start by testing the complete pattern exploratorily with four IDENTITY levels, first with gapping as a baseline, then with full forms. Finally, we zoom in and focus on the strongest contrast with more data points per condition.

4 Experiment 2: gradual mismatches in RNR (and gapping)

In Experiment 2, we tested the perception of RNR mismatches to see whether it patterns with the cloze data: do speakers not only produce fewer RNR mismatches when they differ more strongly in their morphological properties, but do they also perceive them as degraded? To answer this question, we conducted an acceptability rating study in the form of a 2 × 4 design that crossed EllipsisType (gapping vs. RNR) with all four Identity conditions (identical vs. syncretic vs. ending vs. ending+vowel; see Table 1).

Experiments 2 and 3 were designed as an exploratory first step to see whether mismatches are degraded at all, and if so, to what degree, and therefore rely on a relatively small sample size given the eight experimental conditions. Although this limits the robustness and statistical power of the analyses, both experiments provide an initial overview of how speakers respond across the full set of conditions. These preliminary patterns help identify trends that warrant further investigation in the reduced Experiment 4. The items of Experiment 2 were included as fillers in another study on topic drop and therefore transformed to instant messaging dialogues (see below).[12]

All approaches discussed in Section 2 consider RNR to be acceptable if the elided form is morpho-phonologically identical to its correlate. With the possible exception of the ATB-MD-C approaches, all approaches predict syncretic forms not to be degraded either. According to the LI and the MSI approaches, this extends to all (considered) morphological mismatches in person and number. This means that we should see no difference between the Identity conditions of RNR as compared to gapping. In contrast, the PI account predicts that all phonological mismatches are degraded. Consequently, there should be a difference between the identical and the syncretic conditions of RNR on the one hand and the ending and the ending+vowel conditions on the other hand. The GPI approach derived from Christ (2011) takes it that one deviation in phonological form is still in the range of acceptable sentences; two deviations in phonological form, however, are predicted to be substantially degraded. A result in line with the GPI predictions would also best correspond to the production pattern of Experiment 1.

4.1 Materials

We used a variant of the 24 items of Experiment 1 where some token sets varied in their wording or were exchanged for reasons of greater naturalness in the different communication form, namely short instant messaging dialogues between two persons. Person A sets the discourse

[12] One reviewer suggested that the messaging format may involve greater tolerance for errors, potentially affecting judgments of RNR mismatches. We cannot fully rule out this possibility, but while certain orthographic slips are indeed more acceptable in text messages, there is no evidence that the grammatical constraints relevant to RNR are relaxed in this form of communication. Moreover, Experiments 3 and 4, which were not in the messaging format, had similar (null) results.

topic and then asks a question (12) to which person B replies (13). B's answer is always an embedded verb-final clause with a coordination using *und* "and" that either contains gapping or RNR. In (13a) the verb forms are identical, in (13b) the forms are syncretic, in (13c) only the ending differs, and in (13d) there is additionally a vowel change.

(12) A: "Again about the presentation preparation. Have you already divided the work with the others, who prepares which texts?"

(13) B: *Noch nicht ganz, aber klar ist, dass…*
B: "Not entirely yet, but it is clear that..."

a. B: … *Jan den Zeitschriftenartikel* (*liest*) *und Sara das Kapitel aus dem Buch* (*liest*). (identical)
Jan the journal.article read.3SG and Sara the chapter from the book read.3SG

b. B: … *wir den Zeitschriftenartikel* (*lesen*) *und Jan und Sara das Kapitel aus dem Buch* (*lesen*). (syncretic)
we the journal.article read.1PL and Jan and Sara the chapter from the book read.3PL

c. B: … *ich den Zeitschriftenartikel* (*lese*) *und Jan und Sara das Kapitel aus dem Buch* (*lesen*). (ending)
I the journal.article read.1SG and Jan and Sara the chapter out the book read.3PL

d. B: … *ich den Zeitschriftenartikel* (*lese*) *und Jan das Kapitel aus dem Buch* (*liest*). (ending+vowel)
I the journal.article read.1SG and Jan the chapter out the book read.3SG

The 24 items were mixed with 84 fillers. Among these, four were variants of the items with the copula verb *sein* "to be," which exhibits suppletive forms (e.g., *bin* "am" vs. *ist* "is"). We included them for exploratory purposes and will not discuss them further. In addition, 24 fillers were items from a study on topic drop. Another seven fillers were ungrammatical catch trials with two finite verbs in a single clause, serving as attention checks. The remaining 49 fillers covered a wide range of acceptability, from clearly acceptable to highly marked sentences.

4.2 Procedure

Forty-eight native German speakers aged 18 to 50 participated in the experiment. They were recruited from the crowdsourcing platform Clickworker (www.clickworker.de) and were paid €2.50. We used the LimeSurvey survey presentation software (Limesurvey GmbH) to host the experiment. The participants rated the naturalness of the target utterance on a 7-point Likert

scale (1 was *vollkommen unnatürlich* "completely unnatural" and 7 *vollkommen natürlich* "completely natural"). Following a Latin square design, the items were distributed across eight lists. Each participant rated each critical item once and in only one condition. The items and the fillers were displayed as lowercase instant messages[13] and presented in pseudo-randomized order to prevent items or fillers of the same type from appearing consecutively.

4.3 Results

We excluded six participants from the analysis because they had rated four or more of the ungrammatical control utterances with a 6 or 7 on the Likert scale, indicating potential inattentiveness. We had to discard six further data points because the ending+vowel condition of one critical item contained a typo. Figure 2 shows the mean ratings and standard errors. All conditions, including the clear mismatches, received relatively high ratings near 5 and above on the 7-point scale. The identical condition was rated higher for both gapping and RNR than the other conditions, which look very similar. For those other conditions, syncretic, ending and ending+vowel, there seems to be a slight degradation of RNR compared to gapping.

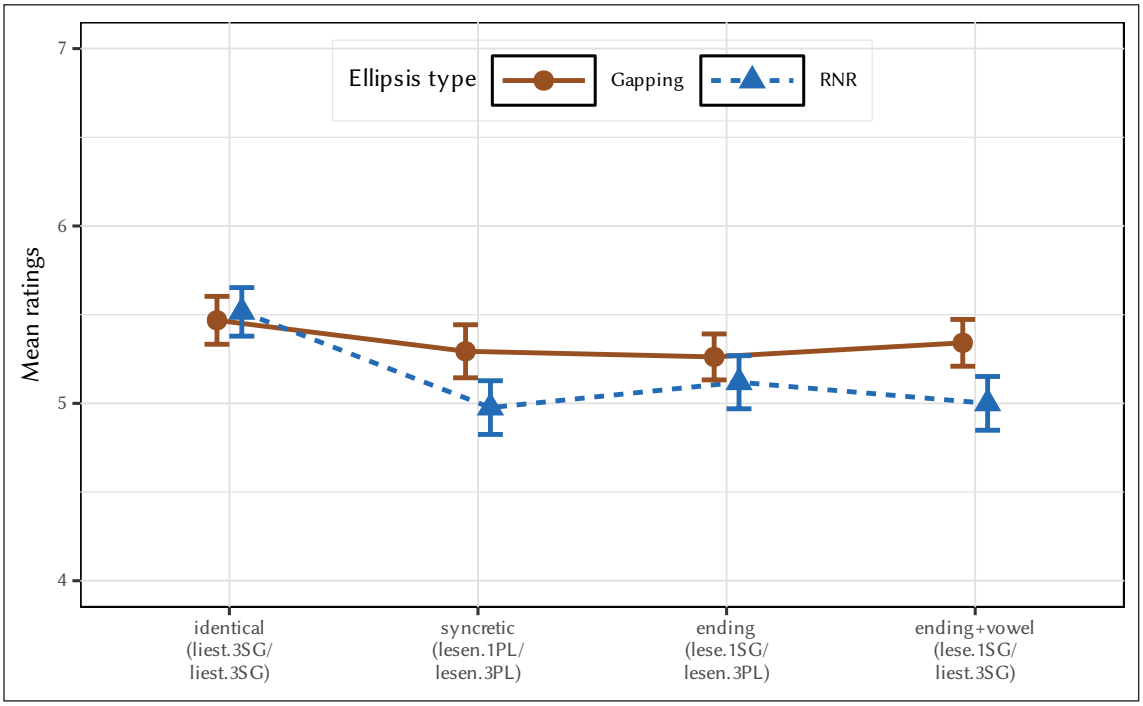


Figure 2: Mean ratings and standard errors per condition for Experiment 2

[13] This ensured that the verb was written consistently in all topic drop conditions.

We analyzed the data with cumulative link mixed models (CLMMs) from the package *ordinal* (Christensen 2023) and coded the independent variables like in Experiment 1. The acceptability scores were modeled as a function of the dummy-coded binary predictor EllipsisType (the baseline gapping coded as 0, RNR as 1) and the forward-coded predictor Identity: identical.rest compared the identical condition (coded as 0.75) to the other conditions (coded as −0.25), maphon.mmphon compared the phonologically matching conditions (identical and syncretic, coded as 0.5) to the phonologically mismatching conditions (ending and ending+vowel, coded as -0.5) and rest.vowel compared the identical, syncretic, and ending conditions (coded as 0.25) to the ending+vowel condition (coded as -0.75). We also included the *z*-transformed Position of the trial in the experiment to control for potential habituation effects and considered the possible interactions between EllipsisType and Identity to account for RNR-specific effects. The model contained random intercepts for participants and items, by-participant and by-item random slopes for the three Identity contrasts and their interactions with EllipsisType, and a by-item random slope for EllipsisType.[14] Table 4 shows the fixed effects of the model with flexible thresholds.

Table 4: Fixed effects in the CLMM of Experiment 2

Fixed effect	Est.	SE	*z*	*p*	
EllipsisType	−0.26	0.15	−1.78	0.07	.
identical.rest	0.25	0.27	0.92	0.36	
maphon.mmphon	0.15	0.26	0.59	0.56	
rest.vowel	−0.12	0.26	−0.48	0.63	
Position	−0.06	0.07	−0.83	0.41	
EllipsisType:identical.rest	0.78	0.51	1.53	0.13	
EllipsisType:maphon.mmphon	−0.48	0.49	−0.98	0.33	
EllipsisType:rest.vowel	0.43	0.42	1.03	0.30	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

We found a marginal main effect of EllipsisType ($z = -0.83$, $p = 0.07$) suggesting a tentative preference for structures with gapping compared to those with RNR. Neither the other main effects nor the interactions were significant. There was no mismatch penalty for the baseline gapping, and the non-significant interaction suggests that RNR did not deviate from this baseline.

[14] The formula was Rating ~ EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) + Position + (1 EllipsisType : (identical.rest + maphon.mmphon + rest.vowel) | Subject) + (1 EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) | Item). When including the by-participant random slope for EllipsisType, the variance-co-variance matrix of the model was undefined. We excluded this slope because it explained the least variance of all random effects.

4.4 Discussion

In Experiment 2, we tested the acceptability of RNR (mis)matches compared to gapping (mis) matches as a baseline. As was to be expected, gapping and RNR pattern alike in the identical condition, which requires identity in morphological features and phonological form. Basically, the participants seem to find matching gapping and matching RNR equally good.

In terms of mismatches, there is a slight drop in ratings from a purely visual perspective, but this concerns both gapping and RNR, with the drop being a bit larger for RNR. This result is in line with Experiment 1, where the ellipsis rate decreased for all mismatches under gapping and RNR, but it is not predicted by any of the RNR-specific identity approaches. However, the rating differences observed are very small and do not result in statistically significant effects. With a larger sample size and more power, it is conceivable that the difference between identical and the mismatches and the differences between gapping and RNR in the identical and the ending+vowel condition might become significant, but most likely not the difference for ending, for which there is not even a descriptive tendency. Moreover, the mismatches were rated very high, with around 5 points on a 7-point scale and therefore can hardly be classified as ungrammatical.

The result of Experiment 2 is inconsistent with the ATB-MD-C approaches, which predict a general mismatch penalty. Moreover, also in RNR, just like in gapping, the mismatch conditions syncretic, ending, and ending+vowel essentially pattern alike. This is consistent with both the MSI and the LI approach, but it is evidence against the PI approach, which predicts a difference between identical and syncretic on the one hand and the two ending conditions on the other, and the GPI approach, which predicts a staircase-like pattern.

In summary, this initial exploratory rating study suggests that RNR behaves predominantly in parallel with gapping, and that RNR mismatches are acceptable and cannot be considered ungrammatical. The data pattern is compatible with the MSI and LI approaches but not with the GPI approach, which best describes the results of Experiment 1. We will return to this point and the possible differences between cloze and rating study in the general discussion. First, however, we look at all four conditions again in the equally exploratory Experiment 3, but this time we compared RNR with the full forms.

5 Experiment 3: comparing RNR structures with full forms

In Experiment 2, we used gapping as a baseline for RNR. In the likewise exploratory Experiment 3, we compared RNR to the full forms of the RNR items. This way, we wanted to ensure that the results of Experiment 2 persist in the presence of a non-elliptical baseline, i.e., we wanted to rule out that they were driven by any ellipsis-specific effects that potentially apply to both gapping and RNR.

The predictions are basically the same as for Experiment 2: the full forms should be fully acceptable, just like the RNR identical condition. Depending on the theoretical approach, the RNR mismatches should not be acceptable at all (ATB-MD-C approaches), only acceptable if the phonological form is the same (PI approach) or does not deviate too much from the correlate (GPI approach), or all be acceptable (MSI and LI approaches), which would also be in line with the result of Experiment 2.

5.1 Materials

In anticipation of the reading time experiment that was planned to test whether mismatches come with an increased cognitive effort, see Experiment 5, we made an adjustment to our materials. Measuring reading times for RNR structures requires that the area of interest, i.e., the correlate in the last conjunct, is kept constant, while the target is varied. Therefore, we consistently used a 3rd person plural verb form in the second conjunct and changed the person/number features of the overt or covert verb in the first conjunct. This modification of the IDENTITY variable is sketched in Table 5 and illustrated in (14).

Table 5: Modified IDENTITY manipulation in Experiment 3

			mismatch in		mismatch in		
	TARGET	CORRELATE	PERSON	NUMBER	ENDING	VOWEL	CONDITION
1	*(sie) lesen*	*(sie) lesen*	–	–	–	–	identical
2	*(wir) lesen*	*(sie) lesen*	+	–	–	–	syncretic
3	*(ich) lese*	*(sie) lesen*	+	+	+	–	ending
4	*(er) liest*	*(sie) lesen*	+	+	+	+	ending+vowel

(14) A: "What is the status of the presentation?"
B: *Es ist abgemacht, dass...*
B: "It is agreed that..."

a. *die Mia und der Ben den Zeitschriftenartikel (lesen) und der*
the Mia and the Ben the journal.article read.3PL and the.NOM
Jonas und die Maike das Kapitel aus dem Buch (lesen) (identical)
Jonas and *the* Maike the chapter out the book read.3PL

b. *wir den Zeitschriftenartikel (lesen) und der Jonas und die Maike das*
we the journal.article read.1PL and the.NOM Jonas and the Maike the
Kapitel aus dem Buch (lesen) (syncretic)
chapter out the book read.3PL

c. *ich den Zeitschriftenartikel* (*lese*) *und der Jonas und die Maike*
I the journal.article read.1SG and the.NOM Jonas and the Maike
das Kapitel aus dem Buch (*lesen*) (ending)
the chapter out the book read.3PL

d. *die Mia den Zeitschriftenartikel* (*liest*) *und der Jonas und die*
the Mia the journal.article read.3SG and the.NOM Jonas and the
Maike das Kapitel aus dem Buch (*lesen*) (ending+vowel)
Maike the chapter out the book read.3PL

We reformulated or exchanged some token sets to improve the overall naturalness of the materials. Additionally, definite articles were inserted before proper names to clearly mark the case of the subject of the second conjunct, which was always a male name or a conjunction of a male and a female name, as nominative.

Along with the 24 items, we tested 68 fillers. Thirty-two of them were items from a study on verbal complexes with three verbs, where the focus constituent was marked by capitalization. Sixteen further fillers were pairs of *wh*-questions with prosodically (again indicated by capitalization) matching or mismatching answers. Twelve fillers contained focused direct objects in the target clause (either article or noun capitalized), and eight fillers were catch trials with severe case mismatches used as attention checks.

5.2 Procedure

Forty-eight native German speakers between the ages of 18 and 40 from Prolific (www.prolific.com) participated in the experiment and were paid £4.43. The experiment was implemented in PCIbex (Zehr & Schwarz 2023). Again, the participants rated the naturalness of the target utterances with a 7-point Likert scale (7 = "completely natural"). Because of the fillers' focus manipulation, the participants were furthermore instructed to read the dialogues out loud before rating them. Like in Experiment 2, the items were distributed across eight lists according to a Latin square design and presented in a pseudo-randomized order together with the fillers.

5.3 Results

Since no participant rated more than half of the catch trials with 5 or better, there were no exclusions. Figure 3 shows the mean ratings and standard errors for all eight conditions. Visually, the full forms and RNRs show a very homogeneous pattern, with all conditions being equally acceptable except for the full forms in the identical conditions, which seem slightly degraded.

Like in Experiment 2, we fitted a CLMM and modeled the ratings as a function of the dummy-coded predictor Form (full form as 0, RNR as 1), the predictor Identity, and the numerical *z*-transformed Position. Identity was again forward-coded: identical.rest compared identical (coded as 0.75) to the rest (−0.25), maphon.mmphon compared identical and syncretic (0.5) to ending and ending+vowel (-0.5) and rest.vowel compared identical, syncretic, and ending (−0.75) to ending+vowel (0.25). We also included all interactions of these contrasts with Form in our model, just like random intercepts for participants and items and by-subject and by-item random slopes for Form, the three Identity contrasts and their interactions.[15] Table 6 shows the fixed effects of the model with flexible thresholds. It contained only a significant main effect of Position ($z = 2.38$, $p < 0.05$), according to which participants gave better ratings as the experiment unfolded.

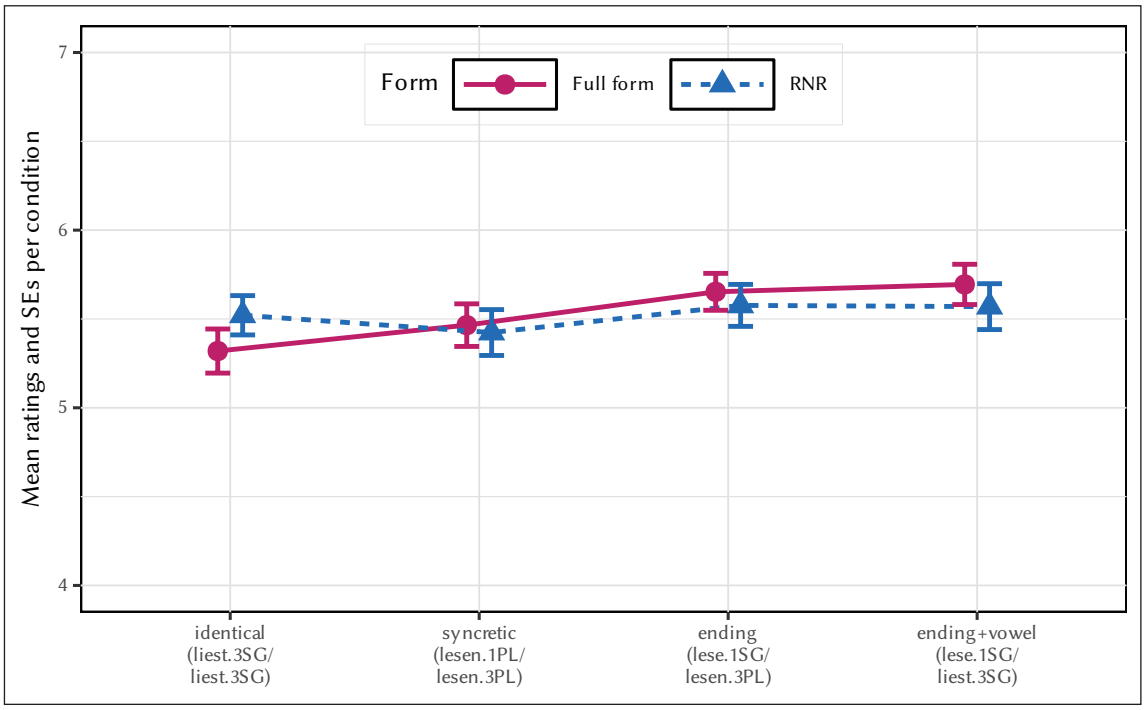


Figure 3: Mean ratings and standard errors for Experiment 3

[15] The formula was Rating ~ EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) + Position + (1 EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) | Subject) + (1 EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) | Item).

Table 6: Fixed effects in the CLMM of Experiment 3

Fixed effect	Est.	SE	*z*	*p*	
Form	0.20	0.23	0.91	0.36	
identical.rest	−0.25	0.27	−0.93	0.35	
maphon.mmphon	−0.31	0.30	−1.01	0.31	
rest.vowel	−0.30	0.26	−1.15	0.25	
Position	0.16	0.07	2.37	0.02	*
Form:identical.rest	0.28	0.38	0.75	0.45	
Form:maphon.mmphon	0.05	0.43	0.12	0.90	
Form:rest.vowel	0.04	0.37	0.10	0.92	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

5.4 Discussion

Experiment 3 suggests that the conditions syncretic, ending, and ending+vowel pattern parallel to their full forms, and thereby provides additional support that all mismatch conditions are rated within the realm of full grammaticality. Moreover, all RNR conditions are essentially rated the same with no significant or clearly visible differences. This result is similar to the result of Experiment 2 and in line with the MSI and the LI approaches, while it runs against the PI and GPI approaches.

While Experiments 2 and 3 must be treated with caution, as they were probably underpowered with 48 subjects and 8 conditions, they consistently indicate that RNR mismatches are considered acceptable and behave similarly to the corresponding baselines gapping and full forms. In two follow-up experiments, we aimed to confirm this result with a decreased number of conditions and, thus, more data points per condition.

6 Experiments 4a and 4b: testing the ends of the scale

To address the potential power problem of Experiments 2 and 3, we reduced the number of conditions in Experiments 4a and 4b. We again used gapping as a more natural baseline and focused on the contrast that the GPI approach predicts to be most extreme, i.e., between the identical condition and the ending+vowel condition, in which the ending and stem vowel change between target and correlate. This gives us a 2 × 2 design with the two binary factors EllipsisType (gapping vs. RNR) and Identity (identical vs. ending+vowel; see Table 7).

Table 7: Design of Experiments 4a and 4b

			mismatch in		mismatch in		
	TARGET	CORRELATE	PERSON	NUMBER	ENDING	VOWEL	CONDITION
1	*(sie) lesen*	*(sie) lesen*	–	–	–	–	identical
2	*(er) liest*	*(sie) lesen*	+	+	+	+	ending+vowel

Given the results of the two previous rating studies, we expect no difference between the identical and the mismatch condition, at least no RNR-specific difference. This would be in line with the LI and the MSI approaches, which predict all (considered) morphological mismatches to be acceptable.

In contrast, the PI and the GPI accounts predict that all phonological mismatches are degraded. Consequently, there should be a difference between the identical condition of RNR on the one hand and the most extreme ending+vowel condition on the other hand. Such an interaction between Identity and EllipsisType would also be tentatively in line with the production data from Experiment 1, where we saw the staircase-like pattern of the ellipsis rate only for RNR and not for gapping. However, if the rating data patterned with the production results, we would additionally expect a (weak) main effect of Identity that shows a general mismatch penalty for both gapping and RNR in line with the reduced ellipsis rates in the cloze study for all mismatches.

We tested two variants of Experiment 4, Experiments 4a and 4b, which only differed in their fillers, which were the experimental stimuli of two studies on sluicing. Since this difference is not relevant for our research on RNR mismatches, we report both experiments in parallel below.

6.1 Materials

We tested basically the same items as in Experiment 3 in the reduced form. We simplified their structure from a dialogue to a complex sentence and removed the definite article again for greater naturalness (15).

(15) *Für das Referat ist abgemacht, dass...*
"For the presentation, it is agreed that..."

a. *Jan und Sara den Zeitschriftenartikel* (*lesen*) *und Julia und Moritz*
Jan and Sara the journal.article read.3PL and Julia and Moritz
das Kapitel aus dem Buch (*lesen*). (identical)
the chapter out the book read.3PL

b. *Jan den Zeitschriftenartikel* (*liest*) *und Julia und Moritz das Kapitel*
Jan the journal.article read.3SG and Julia and Moritz the chapter
aus dem Buch (*lesen*). (ending+vowel)
out the book read.3PL

In both experiments, 72 fillers were included: 24 instances of sluicing, in Experiment 4a with the remnant *wer* "who," in Experiment 4b with *welcher* "which," and 48 complex sentences, with half containing verb-final complement clauses and the other half causal subordinate clauses.

6.2 Procedure

We recruited further native German speakers between 18 and 40. In Experiment 4a, 47 participants[16] from Clickworker (www.clickworker.de) received €4.14 for their participation. In Experiment 4b, 72 subjects from Prolific took part and were paid £2.61. The studies were implemented in PCIbex (Zehr & Schwarz 2023). The participants' task was again to rate the naturalness of given sentences on a 7-point Likert scale (7 = "completely natural"). Additionally, they had to answer a total of eight yes–no comprehension questions asked after fillers. We used these questions as attention checks. The distribution of materials to lists and pseudorandomization was analogous to the previous experiments.

6.3 Results

In Experiment 4a, two participants were excluded because their accuracy on the comprehension questions was below 75%. In Experiment 4b, there were no exclusions. Figure 4 shows the means and standard errors for both experiments next to each other. In both experiments, gapping and RNR received almost identical ratings. From visual inspection, there seems to be a slight degradation of the mismatch condition in Experiment 4a of less than 0.25 points, while in Experiment 4b the two IDENTITY conditions are rated equally acceptable.

We analyzed both experiments again with CLMMs (Christensen 2023) that modeled the ratings as a function of ELLIPSISTYPE and IDENTITY, their interaction, and the *z*-transformed POSITION in the experiment. Both ELLIPSISTYPE and IDENTITY were dummy-coded (gapping and identical as baselines coded as 0, RNR and ending+vowel as 1). The random effect structure consisted of random intercepts for participants and items and of by-item and by-subject random slopes for ELLIPSISTYPE, IDENTITY, and their interaction.[17] The fixed effects of the model for Experiment 4a with flexible thresholds are shown in Table 8, the ones for the model of Experiment 4b with

[16] Due to a platform error, data were collected from 47 rather than the intended 48 participants.
[17] The formula was RATING ~ ELLIPSISTYPE * IDENTITY + POSITION + (1 + ELLIPSISTYPE * IDENTITY | SUBJECT) + (1 + ELLIPSISTYPE * IDENTITY | ITEM).

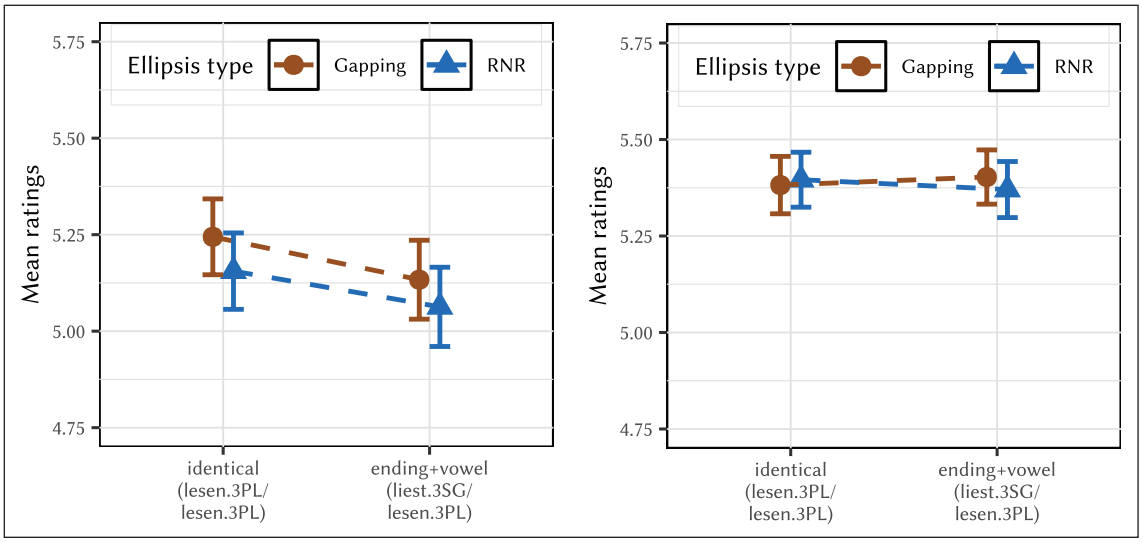


Figure 4: Mean ratings and standard errors for Experiments 4a and 4b

flexible thresholds are in Table 9. In both models, there is only a significant main effect of Position ($z_a = 4.91$, $p_a < 0.001$; $z_b = 3.83$, $p_b < 0.001$), indicating a habituation effect. There is no significant difference between the match (identical) and the mismatch (ending+vowel) condition, either in general or specifically for RNR.

Table 8: Fixed effects in the CLMM of Experiment 4a

Fixed effect	Est.	SE	*z*	*p*	
EllipsisType	−0.12	0.20	−0.62	0.54	
Identity	−0.14	0.20	−0.69	0.49	
Position	0.30	0.06	4.91	< 0.001	***
EllipsisType:Identity	0.02	0.27	0.08	0.94	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

Table 9: Fixed effects in the CLMM of Experiment 4b

Fixed effect	Est.	SE	*z*	*p*	
EllipsisType	−0.08	0.16	−0.44	0.63	
Identity	−0.08	0.19	−0.49	0.66	
Position	0.19	0.05	3.83	< 0.001	***
EllipsisType:Identity	0.07	0.22	0.32	0.75	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

6.4 Discussion

In both Experiments 4a and 4b, we found no difference in acceptability between an RNR match and a clear mismatch with a change in morphological ending and in stem vowel. RNR behaves completely parallel to gapping, where no differences can be seen either. The two Experiments 4a and 4b thus confirm the picture that already emerged in the exploratory Experiments 2 and 3, this time with more data points per condition: mismatches in person and number under verbal RNR are not particularly degraded in comparison to gapping or non-elliptical controls or not degraded at all. Therefore, we conclude that they are acceptable or at least not ungrammatical. Concerning RNR's identity conditions, this contradicts all approaches except for MSI and LI.

However, these approaches do not account for the pattern that we found in the cloze study in Experiment 1 and that is best compatible with (a modified) GPI account. The most extreme production effect, i.e. the bottom step of the staircase-like pattern, was not at all reflected in degraded acceptability in Experiments 4a and 4b. It might be that the task of rating, i.e. assessing acceptability, differs from that of the cloze task, in which participants have the opportunity not only to implicitly evaluate the presented sentences, but also to correct them so that they may become even more acceptable. Another possible explanation for the divergence between the cloze task and the acceptability judgment studies might be processing. The acceptability judgments occur after processing has concluded, so these ratings may not indicate issues encountered during parsing. Such issues, however, could lead speakers to spontaneously specify a full form during the cloze task. To determine whether the discrepancy between cloze and rating data is attributable to challenges in parsing, we conducted a self-paced reading study.

7 Experiment 5: self-paced reading study, RNR vs. full form

Experiment 5 is a self-paced reading study dedicated to investigating whether there is evidence for more complex processing of a mismatching verb form under RNR. Using a 2 × 2 design, the study compared the processing of RNR with that of the corresponding full form (factor Form) when the verbs were either identical or varied in their ending and vowel (factor Identity).

If the findings from the reading time study correspond to those of the cloze study, we anticipate observing processing difficulties in the mismatching ending+vowel condition. Specifically, longer reading times are expected at or after the point where participants detect the mismatch, namely, on the verb in the second conjunct or in the subsequent (spillover) material. Such processing difficulties may explain why more participants avoided ellipsis in the mismatching ending+vowel condition compared to the identical condition in the cloze task.

In our discussion of Experiment 4, we noted that the rating results, which showed no RNR-specific mismatch penalty, may not capture initial processing difficulties, as ratings are provided after parsing is complete. Alternatively, of course, these results might just indicate that participants did not experience such difficulties at all, rather than that they successfully resolved or "repaired" them. If this is in fact the case, we would expect no reading time differences for RNR between the identical and the ending+vowel conditions.

With respect to the full forms, there are in principle two possibilities. In the identical condition, the verb in the second conjunct is the same as in the first and, thus, redundant. This redundancy could either be reflected in faster reading times in the matching compared to the mismatching condition, which is less redundant, indicating easier processing or in a slow-down because the reader would expect an ellipsis but encounters a full form.

7.1 Materials

We used the same 24 items as in Experiment 4 and added a final infinite subordinate clause as spillover after the verb of the second conjunct (16). The clause always began with the conjunction *um* "in order to", a definite article, and a singular noun. These three words and the verb in the second conjunct form the critical region, on which reading times were measured.

(16) *Für das Referat ist abgemacht, dass...*
"For the presentation, it is agreed that...",

a. *Jan und Sara den Zeitschriftenartikel* (*lesen*) *und Julia und Moritz*
Jan and Sara the journal.article read.3PL and Julia and Moritz
das Kapitel aus dem Buch $\underline{lesen}_1$, $\underline{um}_2$ $\underline{den}_3$ $\underline{Aufwand}_4$ *für jeden*
the chapter out the book read.3PL to the effort for every
Einzelnen geringer zu halten. (identical)
individual lesser to keep

b. *Jan den Zeitschriftenartikel* (*liest*) *und Julia und Moritz das Kapitel aus*
Jan the journal.article read.3SG and Julia and Moritz the chapter out
dem Buch $\underline{lesen}_1$, $\underline{um}_2$ $\underline{den}_3$ $\underline{Aufwand}_4$ *für jeden Einzelnen geringer zu halten.*
the book read.3PL to the effort for every individual lesser to keep
(ending+vowel)

We also tested 84 fillers: 24 from a study comparing the present perfect and the preterite, 12 with (potential) gapping structures with parentheses, 12 coordinations where ellipsis was not licensed, and 12 complex sentences with causal subordinate clauses.

7.2 Procedure

48 German native speakers aged 18 to 40 were recruited via Prolific and paid £3.50 to participate in a PCIbex (Zehr & Schwarz 2023) study. Participants read materials word-by-word in a moving window paradigm, advancing with the spacebar, then answered balanced yes-no questions using the *j* or *f* key. Four practice trials preceded the experiment to familiarize participants with the tasks. Items were split across four lists, ensuring each token set appeared only once per participant and that all conditions were equally often represented. Again, the materials were presented in pseudorandomized order.

7.3 Results

The mean accuracy for the comprehension questions was notably high at 95.5%. As the lowest accuracy achieved was 88.1%, no participants were excluded, and data from 48 individuals were analyzed. Figure 5 shows the mean reading times and standard errors per condition for the words in the critical region. Overall, the reading times for all conditions are very similar on each of the words. On words 2 to 4, RNR seems to be read a bit slower than the full forms. On the first word, the critical verb, there seem to be slightly larger differences between conditions, however, also the error bars are larger.

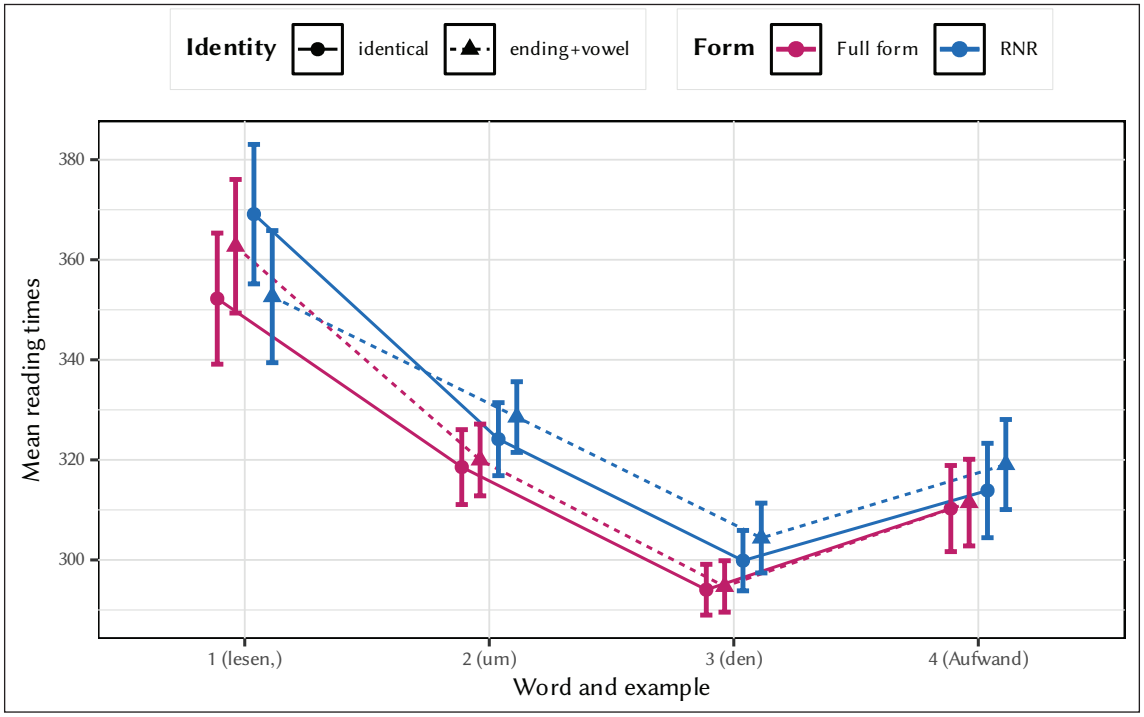


Figure 5: Mean reading times and standard errors per condition for the region of interest for Experiment 5

We conducted an analysis of the reading times for each of the four words within our designated region of interest, as indicated in (13) and in Figure 5, using separate linear mixed effects regression models (LMERs, package lme4, Bates et al. 2015) in R. All models contained the log-transformed reading times as dependent variable and as independent variables the dummy-coded Form (full form 0, RNR 1) and Identity (identical 0, ending+vowel 1), as well as their interaction and the *z*-transformed Position. The analyses of $word_1$ and $word_4$ additionally contained a fixed effect of Length, i.e., the *z*-transformed number of characters of the corresponding word.[18] This predictor was not necessary for $word_2$ and $word_3$ because they were always the subjunction *um* with two and a definite article with three letters. We reduced the random effects structure until the models converged and did not result in singular fit. The models for $word_2$ and $word_3$ contained only random intercepts for items and participants. The model for $word_1$ additionally had a by-subject random slope for Form and a by-item random slope for Identity, the model for $word_4$ had a by-subject random slope for Identity.[19]

All models contained a significant main effect of Position ($p < 0.001$), indicating that participants' reading times decreased over the course of the experiment. Only in the model for analysing $word_1$ were there additionally two marginal simple effects of Form and Identity and a significant interaction between Form and Identity (Table 10). The second verb tends to be read slower in RNR when both verbs are identical ($t = 1.74$, $p = 0.08$) and mismatches tend to be slower than matches in full forms ($t = 1.68$, $p = 0.09$). The interaction ($t = 2.24$, $p < 0.05$) qualifies these effects and suggests that the verb in the second conjunct is significantly slower in RNR structures when antecedent and target were identical compared to the full form with identical verbs. Viewed from the other side, the verb was read faster in case of a mismatch for RNR compared to the full form.

Table 10: Fixed effects in the LMER of $word_1$ (verb in second conjunct) of Experiment 5

Fixed effect	Est.	SE	*t*	*p*	
Intercept	5.71	0.06	99.85	< 0.001	***
Form	0.04	0.02	1.74	0.08	.
Identity	0.04	0.02	1.68	0.09	.
Length	0.04	0.02	1.63	0.12	
Position	−0.09	0.01	−11.86	< 0.001	***
Form:Identity	−0.07	0.03	−2.24	< 0.05	*

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

[18] The model calls were as follows, with the variation in random effects discussed in the following: LogRT ~ Form * Identity + Position (+ Length) + (1 | Participants) + (1 | Items).

[19] For $word_4$, we found an alternative model featuring a distinct random slope that likewise did not produce a singular fit. We compared the two models and selected the one with the lower AIC value. The parameter estimates were consistent to the second decimal point between the two models.

7.4 Discussion

In our self-paced reading study, we investigated the processing of RNR in comparison to the corresponding full form. While we did not find any relevant effects on the spillover, participants read the verb in the second conjunct faster in the mismatching ellipsis condition. This result is unexpected: the mismatch condition, which is predicted to be cognitively more demanding, is processed faster under ellipsis, while the match condition is processed more slowly. This contradicts the hypothesis that processing mismatching correlates requires greater effort.

At this point, we have no explanation for the apparent processing advantage of the mismatch.[20] What we can say, however, is that in any case the mismatch condition is not more difficult to process than corresponding full forms or than the match. Therefore, from a processing perspective, whatever the explanation for the RNR-mismatch condition is, it is not to be expected that the mismatches are rated worse compared to matches and full forms. Accordingly, our rating data from Experiments 2 to 4 and the processing data are consistent. At the same time, the results of the cloze study, i.e., the staircase-like decreasing ellipsis rates for RNR mismatches, apparently cannot be accounted for by processing. There is no evidence that RNR mismatches cause processing problems that lead speakers to more often produce a full form.

8 General discussion

In this paper, we reported on a cloze task, three rating studies, and a reading time study on RNR mismatches. Their aim was first and foremost an empirical one, namely to contribute (much-needed) empirical evidence to the discussion of identity constraints in German RNR. Another objective was to evaluate the different theoretical approaches to RNR presented at the beginning of the paper.

Even though the results differ between the more production-oriented cloze task and the other perception-oriented experiments, we can conclude from the whole picture that verbal mismatches in person and number in German RNR are not ungrammatical. In the two exploratory Experiments 2 and 3 and the two Experiments 4a and 4b, these structures received ratings that were comparable to those of gapping and the full forms, which functioned as baselines. This result is also in line with Shiraïshi et al. (2019), who did not find reduced acceptability for RNR tense mismatches in English and French. Furthermore, the cloze task in Experiment 1 showed that the participants consistently did not correct mismatches in all identity conditions, and that they did so in (almost) all conditions as frequently as they produced full forms.

[20] It is notable though that, in their rating study on English RNR, Shiraïshi et al. (2019: 8) found that "the RNR-MATCH condition is less acceptable than the RNR-MISMATCH condition ($p < 0.05$)," see their Figure 2 on the same page.

The results that the mismatch conditions pattern alike in the rating studies, that they seem not to require additional cognitive effort according to Experiment 5, and thus arguably should be regarded as grammatical are in line with semantic MSI and LI approaches. Under Chaves' (2014) MSI approach, number and person would then be features with a relatively weak semantic contribution, weak enough not to cause a reduced acceptability.

What is problematic for the MSI and the LI approaches, though, is the significantly higher frequency of ellipsis in the match condition in Experiment 1. This result is at first glance more in line with the PI approach; however, this approach would also predict the syncretic conditions to pattern alike with the identical conditions, which they did not in Experiment 1, and an acceptability penalty for the morphologically deviating conditions ending and ending+vowel, which we did not find in any of the rating studies. The GPI approach tentatively predicts the staircase-like pattern of Experiment 1, even though Christ (2011) considered the syncretic and the ending conditions to be fine as well. However, the cloze result is not reflected in any of the rating studies, where the RNR mismatches were not degraded. In other words, it seems that listeners have no problem processing this type of RNR mismatches and consider it acceptable. However, when given the opportunity to avoid such a mismatch by using a non-elliptical form, many of them take advantage of this option. No theoretical approach fully predicts this behavior. We therefore conclude that none of the current approaches can (without further assumptions) account for all the data presented in this paper.

The observed discrepancy between the results of the cloze study and the rating studies concerning RNR mismatches highlights the value of employing multiple methods to investigate apparent mismatch penalties. There are at least two options to account for this difference. The first option is that the cloze task in fact tested production rather than comprehension and that RNR mismatches are an issue for language production. This hypothesis would then have to be tested in a "real," i.e., less restricted and more spontaneous production study. The second option could be a methodological difference between the cloze task on the one hand and the rating task on the other. In the rating task, participants assess a sentence post hoc, presented as complete and attributed to another individual. In contrast, the cloze task exposes participants to a possibly incomplete sentence, granting them limited scope to amend it by inserting information at a specific position and thereby "making it their own." So, while participants may deem sentences in the rating task acceptable as utterances produced by others, the cloze task allows them to make an acceptable sentence even more acceptable (to their preferences) by changing it. The cloze task may therefore demonstrate increased sensitivity to subtle differences within high acceptability areas, at least regarding identity-related issues. These observations remain speculative and require further exploration, especially with respect to other phenomena and domains.

Acknowledgments: We thank the three anonymous reviewers for their constructive feedback on this article. We are also grateful to the two editors for their helpful guidance during the review process. Special thanks go to Robin Lemke for the in-depth discussions on the revised version of this article, which substantially improved it. We would also like to thank the student assistants of project B3, who helped in preparing the experiments and in annotating the cloze data.

Funding: Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 232722074 – SFB 1102 / Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102.

Abbreviations

ATB	across the board
GPI	gradual phonological identity
LI	lexeme identity
MD	multi-dominance
MSI	morph stem form identity
PI	phonological identity
RNR	right node raising
VP	verb phrase

Appendix

The materials, data, and scripts from this article can be found online at the following OSF storage: https://osf.io/f7g5k/files/osfstorage.

References

Aoun, Joseph, Elabbas Benmamoun & Dominique Sportiche. 1994. Agreement, word order, and conjunction in some varieties of Arabic. *Linguistic Inquiry* 25(2). 195–220.

Barros, Matthew & Luis Vicente. 2011. Right node raising requires both ellipsis and multidomination. *University of Pennsylvania Working Papers in Linguistics* 17(1).

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Belk, Zoe, Ad Neeleman & Joy Philip. 2023. What divides, and what unites, right-node raising. *Linguistic Inquiry* 54(4). 685–728. https://doi.org/10.1162/ling_a_00454.

Chaves, Rui P. 2014. On the disunity of Right-Node Raising phenomena: Extraposition, ellipsis, and deletion. *Language* 90(4). 834–886. https://doi.org/10.1353/lan.2014.0081.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
Christ, Rüdiger. 2011. Linkstilgung und phonologische Quasi-Identität. *Linguistische Berichte* 228. 371–411. https://doi.org/10.46771/2366077500228_1.
Christensen, Rune H. B. 2023. *ordinal – regression models for ordinal data*. R package version 2023.12-4.1. https://CRAN.R-project.org/package=ordinal, 15 September 2025.
de Vries, Mark. 2009. On multidominance and linearization. *Biolinguistics* 3(4). 344–403. https://doi.org/10.5964/bioling.8735.
Eisenberg, Peter. 1973. A note on "identity of constituents." *Linguistic Inquiry* 4(3). 417–420.
Fuß, Eric. 2018. Sprachliche variation. In Angelika Wöllstein, Peter Gallmann, Mechthild Habermann & Manfred Krifka (eds.), *Grammatiktheorie und Empirie in der germanistischen Linguistik,* 185–224. Berlin: De Gruyter. https://doi.org/10.1515/9783110490992-007.
Grosz, Patrick. 2015. Movement and agreement in Right-Node Raising constructions. *Syntax* 18(1). 1–38. https://doi.org/10.1111/synt.12024.
Hartmann, Katharina. 2000. *Right node raising and gapping: Interface conditions on prosodic deletion.* Philadelphia: John Benjamins. https://doi.org/10.1075/z.106.
Himmelreich, Anke, Melissa Jeckel & Johannes Mursell. 2024. Agreement patterns of coordination. In Anke Himmelreich, Daniel Hole & Johannes Mursell (eds.), *To the left, to the right, and much in between: A Festschrift for Katharina Hartmann,* 71–87. Frankfurt: Goethe University Frankfurt.
Klein, Wolfgang. 1981. Some rules of regular ellipsis in German. In Wolfgang Klein & Willem Levelt (eds.), *Crossing the boundaries in linguistics*, 51–78. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-8453-0_3.
Klein, Wolfgang. 1993. Ellipse. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung / An international handbook of contemporary research* (Vol. 1), 763–799. Berlin: De Gruyter. https://doi.org/10.1515/9783110095869.1.12.763.
Kluck, Marlies. 2009. Good neighbours or far friends: matching and proximity effects in Dutch Right Node Raising. *Groninger Arbeiten zur germanistischen Linguistik* 48. 115–158.
Limesurvey GmbH. n.d. *LimeSurvey: An Open Source survey tool.* Hamburg, Germany: LimeSurvey GmbH. http://www.limesurvey.org, 16 October 2025.
McCawley, James D. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13. 91–106.
Ott, Dennis. 2025. Rethinking right-node raising. *The Linguistic Review* 42(3). 389–418. https://doi.org/10.1515/tlr-2025-2013.
Philips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34(1). 37–90. https://doi.org/10.1162/002438903763255922.
Postal, Paul M. 1974. *On raising: One rule of English grammar and its theoretical implications*. Cambridge, MA: MIT Press.
R Core Team. 2025. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/, 15 September 2025.
Ross, John Robert. 1986 [1967]. *Infinite Syntax!* Norwood: Ablex Publishing Company. [Originally: *Constraints on variables in syntax.* PhD-Dissertation, MIT.]

Sabbagh, Joseph. 2007. Ordering and linearizing rightward movement. *Natural Language & Linguistic Theory* 25(3). 349–401. https://doi.org/10.1007/s11049-006-9011-8.

Schäfer, Lisa, Robin Lemke, Heiner Drenhaus & Ingo Reich. 2021. The role of UID for the usage of verb phrase ellipsis: Psycholinguistic evidence from length and context effects. *Frontiers in Psychology* 12. 661087. https://doi.org/10.3389/fpsyg.2021.661087.

Shen, Zheng. (2019). The multi-valuation agreement hierarchy. *Glossa* 4(1). 46. https://doi.org/10.5334/gjgl.585.

Shiraïshi, Aoi, Anne Abeillé, Barbara Hemforth & Philip Miller, (2019). Verbal mismatch in right-node raising. *Glossa: A Journal of General Linguistics* 4(1). 114. https://doi.org/10.5334/gjgl.843.

Steedman, Mark. 2000. *The syntactic process*. Boston: MIT Press. https://doi.org/10.7551/mitpress/6591.001.0001.

Wilder, Chris. 1997. Some properties of ellipsis in coordination. In Artemis Alexiadou & Tracy Alan Hall (eds.), *Studies on universal grammar and typological variation*, 59–107. Amsterdam: John Benjamins. https://doi.org/10.1075/la.13.04wil.

Wilder, Chris. 1999. Right node raising and the LCA. In Sonya Bird, Andrew Carnie, Jason D. Haugen & Peter Norquest (eds.), *Proceedings of the 18th West Coast Conference on Formal Linguistics,* 586–598. Somerville: Cascadilla Press.

Yatabe, Shuichi. 2003. A linearization-based theory of summative agreement in peripheral-node raising constructions. In Jong-Bok Kim & Stephen Wechsler (eds.), *Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar,* 391–411. Stanford: CSLI Publications. https://doi.org/10.21248/hpsg.2002.20.

Zehr, Jeremy & Florian Schwarz. 2023. *PennController for Internet Based Experiments (IBEX).* https://doi.org/10.17605/OSF.IO/MD832.