Open Library of Humanities

Processing of verbal versus adjectival agreement: Implications for syntax and psycholinguistics

Zuzanna Fuchs, University of Southern California, US, zfuchs@usc.edu

Olga Parshina, Middlebury College, US, oparshina@middlebury.edu

Irina A. Sekerina, College of Staten Island CUNY, US, irina.sekerina@csi.cuny.edu

Maria Polinsky, University of Maryland, US, polinsky@umd.edu

Linguistic theories distinguish between external and internal agreement (e.g., noun-verb agreement vs. noun-modifier agreement, the latter also known as concord) and model them using different mechanisms. While this distinction has garnered considerable attention in syntactic theory, it remains largely unexplored in experimental work. In an effort to address this gap, we conducted two studies of external/internal agreement in Russian using self-paced reading and eye-tracking while reading. We measured the response to violations generated when native speakers encounter a noun that mismatches the features on an earlier element inflected for agreement (verb, modifying adjective, and predicative adjective). Both experimental studies found strong effects of ungrammaticality: participants were sensitive to agreement mismatches between the agreeing element and the trigger. However, there was no interaction observed between the effect of grammaticality and the type of agreeing element, suggesting that, while participants are sensitive to mismatches, the processing of the mismatches does not differ between external and internal agreement. Despite the cross-methodological replication of the null interaction effect, interpreting this result is necessarily tentative. We discuss possible implications, should the result be further replicated by future high-powered studies. We suggest that this outcome may indicate that differences in real-time processing of internal vs. external agreement may not be observable in time-course measures, or that the lack of such differences constitutes support for analyses of agreement as a two-step process, with one step in syntax, and the other, post-syntactic. We invite future work to test these hypotheses further.

1 Introduction

This paper investigates the contrast between external and internal agreement (also known as agreement proper and concord, respectively). This contrast is commonly found in the theoretical literature; given support for differences between the two types of agreement at the level of the syntactic representation, we explore whether these differences are also reflected during real-time processing. In particular, we draw upon an extensive body of literature that explores the processing of agreement mismatches, i.e., the occurrence of a noun that carries a different feature than the agreeing element, with the aim of investigating potential distinctions.

We build on the tradition of measuring the effects of ungrammaticality in agreement during real-time processing to investigate not only whether individuals are sensitive to (un)grammaticality but also whether they are sensitive to differences between external and internal agreement. The language used in the studies is Russian, in which gender is marked on long and short adjectives as well as on verbs, and which exhibits flexible word order, allowing for direct comparison of effects of ungrammaticality. We thus put theoretical models of external versus internal agreement to the test in experiments that employ self-paced reading (SPR) and eye-tracking.

The paper is organized as follows. Section 2 sets the stage for both the theoretical discussion and for the role of agreement (mis)matches during real-time language processing that the studies discussed in subsequent sections will leverage. Section 3 presents a study on Russian external vs. internal agreement using SPR, and Section 4 presents a study using the same language and materials but using eye-tracking while reading. Section 5 offers a general discussion of our results.

2 Background

2.1 External and internal agreement

Research on sentence processing suggests that linguistic elements are represented as bundles of features, and these features allow comprehenders to establish connections between these elements and linguistic material occurring elsewhere in the speech stream that has matching features. Among abstract features whose comprehension during real-time processing has been investigated, agreement features have received a great deal of attention; such attention is not accidental because agreement is a feature-matching operation *par excellence*. An abundance of evidence has shown that comprehenders track agreement between a linguistic element that carries an agreement marker, i.e., a segment inflected for a feature such as gender or number, and a linguistic element elsewhere in the clause that contains the feature(s) matching the one(s) on the agreement marker.

Defined informally, agreement involves the matching of abstract features of a noun on linguistic elements that co-occur with that noun either in the noun phrase (modifiers, determiners) or in

the clause (predicates). No matter how much the realization of such matching varies in actual languages (Corbett 1991; 2003; 2012), the features are remarkably stable: person, number, and gender.[1] In addition, noun modifiers can match the head noun in case, under the phenomenon known as case concord, though we note that case concord is beyond the scope of the current paper.

In considering agreement, we distinguish the feature(s) on a noun from the matching features on a verb, determiner, or modifier. For verbs, determiners, and modifiers, the features [person], [number] and [gender] are not inherent; rather, these clausal constituents receive [person],[2] [number] or [gender] features from a clause-mate noun. In the linguistic literature, this asymmetrical relation between the feature-bearing category and those categories that receive features is reflected in the distinction between the agreement *trigger,* or *goal* that bears the requisite agreement features (the noun) and agreement *target,* or *probe* (e.g., the verb, the determiner) that copies the features present on the noun.

Another contrast that has long been noted in the domain of grammatical agreement has to do with feature matching in the clausal/verbal domain (external agreement) and nominal domain (internal agreement). To illustrate, consider the following examples from Russian, where (1) shows how the predicate changes form depending on the gender of the subject in the nominative case, and (2) shows how adjectives and possessive pronouns change form to match the gender and number of the head noun (*derevnja, dom*).

(1)   a.   Devuška   dal-a   konfety   rebenku.
            girl(F).SG.NOM gave-F.SG candy.ACC.PL child.DAT
            'The girl gave candy to the child.'

        b.   Mal'čik   dal-Ø   konfety   rebenku.
            boy(M).SG.NOM gave-M.SG candy.ACC.PL child.DAT
            'The boy gave candy to the child.'

(2)   a.   naš-a   star-aja   krasiv-aja   derevnja
            our-F.SG old-F.SG beautiful-F.SG village(F).NOM
            'our beautiful old village'

        b.   naš-Ø   star-yj   krasiv-yj   dom
            our-M.SG old-M.SG beautiful-M.SG home(M).NOM
            'our beautiful old home'

A growing number of researchers have suggested that external and internal agreement follow different syntactic mechanisms; consider Sigurdsson (2009), Chung (2013), Norris (2014; 2017;

---

[1] In what follows, we will be using *gender* as a cover term for gender and noun class (see Corbett 1991; Kramer 2015 for overviews).

[2] See however Danon (2011) and references therein for a different account of [PERSON] features.

2018), Ackema & Neeleman (2020), Grabovac (2022), among others. In such theoretical accounts of agreement, the contrast is formalized in the distinction between the structural relationships involved in the two different types of feature-matching. Under external agreement, a higher probe with unvalued features seeks the closest goal bearing that feature in its c-command domain.[3] Given this structural relationship, the probe can be a determiner heading the DP, the inflectional head of a clause, the verbal head, or a complementizer; in all these instances, the probing relationship is between a head and a phrase. In internal agreement, features are copied from a noun onto its modifier(s) (the actual implementation of such copying differs across analyses, see **Figure 1** for an example); the relationship is between two phrases. Modifiers do not have to c-command the goal or be properly local, and can occur in multiples. It is worth noting that some accounts of internal agreement suggest that it *can* also occur under c-command, but the c-command is not critical, whereas for external agreement, it is the defining criterion. A number of other structural differences suggest that the two processes of feature-matching should receive different syntactic accounts (see Norris 2014; 2018, for a full discussion and a particular implementation).
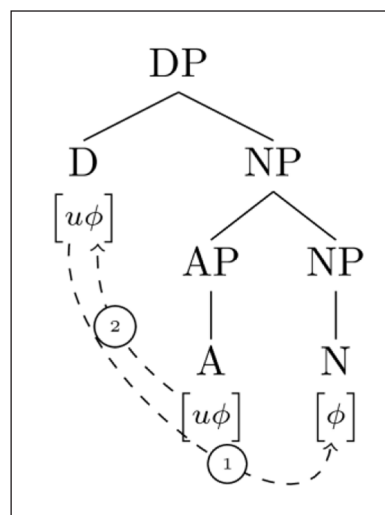

Figure 1: Structural mechanisms involved in feature matching between an NP and D (external agreement) and NP and AP (internal agreement).

The tree in **Figure 1** presents the structural distinction between external agreement (between a D and NP) and internal agreement or concord (between an AP and an NP), illustrating one approach to how features from a noun may be copied onto its modifier. The Agree operation is represented as Step 1 in **Figure 1**, as probing from a head to a phrase (in this example from D to NP). AP must obtain its agreement feature differently, however. Because it does not c-command

---

[3] We set aside theoretical issues concerning the directionality of agreement and multidominance.

the NP, an Agree relation can only obtain between AP and D once D itself has acquired the requisite feature(s). This latter process could happen via a separate instance of Agree, shown as in step 2 in **Figure 1** (Baker 2008; Toosarvandani & van Urk 2013; Ingason & Sigurdsson 2017), or some other mechanism (see Norris 2014; 2017 for details).

While syntactic arguments in favor of distinguishing external and internal agreement have been on the rise, the two phenomena have not yet been directly compared in an experimental setting. In the rich research on the processing of (mis)matches in agreement, differential effects of external vs. internal agreement have received virtually no attention. This work takes a step toward exploring whether such effects can be observed.

2.2 Composite agreement

In the previous subsection we introduced the structural difference between internal and external agreement; we will now discuss the notion of agreement as a two-step operation. Indeed, a number of researchers have proposed that agreement is a two-step operation, with the two steps situated in different modules of grammar. On the theoretical plane, the conceptualization of agreement as occurring in two steps has been motivated by the tension between syntactic and post-syntactic effects implicated in agreement (e.g., Haskell & MacDonald 2005; Benmamoun & Lorimor 2006; Franck et al. 2006; 2010; 2020; Benmamoun et al. 2009; Arregi & Nevins 2012; Bhatt & Walkow 2013; Bruening 2014; Ackema & Neelman 2020; Lyskawa 2021; a.o.).[4] On the one hand, agreement obeys locality, its core phenomena happen in the domain of c-command, and it is sensitive to basic constituency. The crucial role played by these syntactic notions compels one to consider agreement in the syntax. On the other hand, the design of agreement includes phenomena that do not belong to syntax proper. In particular, agreement has morphological exponence and has access to the segmental content of its building blocks. That would be unexpected if agreement were a purely syntactic operation, because the actual exponents are added after the syntactic computation is done (e.g., Embick & Noyer 2001). Further, agreement is case-discriminating (Bobaljik 2008); combined with the assumption that case itself may be post-syntactic (Marantz 1991, a.o.), this also points to the need to locate agreement in the post-syntactic component of grammar.

Practical details and terminologies of two-step agreement models differ. Here, we adopt the implementation by Arregi & Nevins (2012), due to its clarity and the terminology that we find useful (consider also Benmamoun et al. 2009; Franck et al. 2006; 2010; 2020, for a similar approach but different terminology). They define agreement as a two-step process taking place first in the syntax, formalized in the step Agree-Link, and then in post-syntax, a step they refer to as Agree-Copy. Agree-Link amounts to identifying the agreement trigger and its pertinent

---

[4] In a departure from most accounts, Bhatt & Walkow (2013) locate both steps in the syntax.

features in the structure, whereas Agree-Copy consists of sending the featural information about the agreement trigger to the performance system. The performance system, in turn, needs to determine if the featural information is adequately represented.

Empirical motivation for analyzing agreement as a two-step process comes from operations that can intervene between the steps and manipulate the output of the first step of agreement (Agree-Link that was introduced above). Arregi & Nevins (2012) use Basque agreement data to motivate their model. They analyze Basque auxiliaries as a complex of clitics and one non-clitic agreement morpheme. The latter usually expones the phi-features of the absolutive argument, but in some instances it actually expones the features of a dative argument, despite the presence of the absolutive argument elsewhere in the structure. In work which combines syntactic and processing considerations, Franck et al. (2006; 2010; 2020) focus on agreement attraction effects showing that agreement computation can be perturbed by intervening material both when that material c-commands the agreement trigger and when it simply linearly precedes it (see also Bruening 2014).

To summarize the discussion so far, we have identified two contrasts that play a role in grammatical agreement: (i) the structural contrast between external and internal agreement, and (ii) the contrast between establishing the agreement relation (Agree-Link) and the representation of featural information in the linguistic segment (Agree-Copy). We can combine insights from (i) and (ii) to build on the theoretical discussion of the distinction between internal and external agreement: under a composite approach to agreement, internal and external agreement differ at the first step, Agree-Link, but undergo the same process in the second step, Agree-Copy, as captured in **Table 1**.

	Agree-Link: different operations	Agree-Copy: feature-matching
External agreement	c-command; probe is a syntactic head	feature-matching
Internal agreement	no c-command required; probe is phrasal	feature-matching

Table 1: Two dimensions of agreement.

2.3 Processing of agreement (mis)matches

Results from a range of offline and online experimental studies demonstrate that speakers are generally sensitive to feature mismatches in those features that are indexed via agreement in a given language (e.g., Franck et al. 2008 and references therein), though exceptions to this have been noted in certain environments, such as in agreement attraction. Moreover, not all agreement mismatches in agreement are processed equally. Specifically, experimental evidence suggests

that the processing of agreement (mis)matches is modulated by the nature of the feature itself. For instance, experimental work has shown differential processing patterns for singular versus plural number features and, in languages such as Spanish, for agreement in masculine versus feminine (for overviews, see Lago et al. 2015; Beatty-Martínez & Dussias 2019; and references therein).

There has been surprisingly little work directly investigating whether processing of agreement might also be modulated by the lexical category and/or structural position of the agreeing element. For instance, in a language that has gender and number agreement on verbs and modifiers, is agreement processed similarly when it occurs between a noun and an agreeing verbal predicate as compared to an agreeing modifier? If agreement can be modeled differently depending on the structural relationship of the agreeing constituent (cf. Section 2.1), this difference may (but does not have to) translate into processing differences for different agreement targets.

One notable study in this vein is an EEG study by Barber & Carreiras (2005), who among other things compared ERPs elicited by agreement manipulations in Spanish determiner-noun pairs versus noun-adjective pairs. The authors found that agreement errors in these two types of word pairs when presented not in the context of a sentence resulted in different negativity – the noun-adjective errors elicited an N400, while the determiner-noun errors elicited an additional LAN effect. This difference was not observed when the pairs were embedded in sentences.

Further inspiration for this question comes from recent experimental work on German. A three-gender language, German has gender agreement on determiners and on adjectives. In a series of eye-tracking studies using the Visual World Paradigm (Tanenhaus et al. 1995), Hopp & Lemmerth (2018) and Lemmerth & Hopp (2019) investigated how the processing of agreement marking on determiners in German leads the comprehender to actively anticipate a subsequent noun with matching features. As a secondary part of their analysis, the authors compared whether gender marking on the determiner and on the adjective held equal predictive power. They found that monolingual speakers, children and adults alike, processed gender agreement features predictively. However, the predictive effect of gender on the adjective was stronger than the effect on the determiner. A comparison of the relative predictive effect of determiners and adjectives was not part of the experimental set up and emerged as a side effect. Even though an advantage for adjectives appears to hold, Hopp and Lemmerth's studies cannot really isolate these differences, since gender marking occurred only once in the determiner condition, but twice in the adjectival condition (they used two adjectives in a row), possibly resulting in a confounding additive effect. Despite these reservations, these results may still suggest that agreement on the determiner may be processed differently than agreement on the modifying adjective.

The studies presented in Sections 3 and 4 are also concerned with the processing of agreement and agreement mismatches in a sentential context, and so we discuss in broad terms how the

processing of agreement proceeds. Once the comprehender encounters an expression that has an agreement feature, they attempt to match that to an agreement feature on an agreeing linguistic element in an eligible position elsewhere in the clause. If the two elements match in features, the processing of that expression proceeds as usual. However, if the comprehender encounters an expression with a mismatching feature, that should lead to a higher processing load. We propose to leverage this response in investigating the processing of external vs. internal agreement. In SPR, the behavioral paradigm that we use in this paper in Experiment 1, a mismatch in features registered at the trigger item leads to processing difficulty (Smith & Levy 2008), which is reflected in longer reading times in the critical region and sometimes regions following it (spillover regions). Similarly, in reading studies with eye-tracking, a method that we employ in Experiment 2, mismatches in agreement are associated with increases in measures of late stages of processing (see Section 4.4) at the critical and possibly spillover regions (e.g., Keating 2009). Eye-movement recording offers a distinct advantage over other techniques as it enables a more detailed examination of real-time language processing in an environment that closely resembles natural reading.

2.4 Basics of Russian grammatical agreement

We now turn to a brief description of agreement in Russian, the language which we use to explore the relative effect of mismatches in agreement on different agreeing elements. A number of linguistic considerations motivate the choice of Russian for such a study, including the agreement system, the contrast between long and short adjectives, and the availability of flexible word order. We discuss each of these properties and their relevance to the studies in turn.

Russian has three genders: masculine, feminine, and neuter; in the plural, the contrast between the genders is neutralized, so in what follows we will be considering only singular nouns. Numerically, neuter is smaller than the other two genders in terms of its proportion in the lexicon (estimated between 13 percent, cf. Comrie et al. 1996: 109; Akhutina et al. 2001: 296; and 16.7 percent, cf. Slioussar & Samoilova 2014; 2015); neuter nouns are predominantly inanimate. In the study presented here we chose to compare only masculine and feminine nouns. The percentage of masculine and feminine nouns in the disambiguated subcorpus of the Russian National Corpus is 48 percent vs. 35 percent, respectively (Slioussar & Samoilova 2014; 2015).

Unlike some other Slavic languages, Russian has a contrast between long-form (LF) adjectives/ participles and short-form (SF) adjectives/participles (Bailyn 1994; 2012: 68–70); the latter appear only in the predicate position, whereas LF adjectives can have either the modifying or the predicative function. The availability of adjectives in both the modifying and the predicative functions makes way for a direct comparison of internal agreement (in the noun phrase) and external agreement (in the predicate phrase).

LF adjectives agree in gender and number with the head noun; SF adjectives and past tense verbs agree with the subject in the nominative. Agreement on LF adjectives, SF adjectives, and past tense finite verbs is encoded by suffixes. Gender agreement suffixes on LF adjectives are overt for both masculine and feminine; the masculine gender agreement exponent on SF adjectives and on past-tense verbs is null. The paradigms are summarized in **Table 2** and shown both in Cyrillic and in transliteration. The feminine form of SF adjectives and past-tense verbs is usually one letter longer than the corresponding masculine form; note that in Cyrillic, the masculine and feminine forms of the LF adjective are of the same length.

	LF adjective	**SF adjective**	**past tense verb**
	star- 'old'	*star-* 'old'	*beža-* 'run'
Masculine	старый star-yj	стар star-∅	бежал bežal-∅
Feminine	старая star-aja	стара star-a	бежала bežal-a

Table 2: Russian gender agreement.

To directly compare the processing of agreement mismatches on different agreeing elements, we need to keep the order of the agreement trigger and target constant, with the agreeing element consistently preceding the head noun. In the contextually neutral order in the Russian noun phrase, the adjective precedes the noun it modifies, as in (3a), and this order is natural and frequent in Russian noun phrases. In verbal clauses, the order verb-subject (VS) is commonly found in presentational constructions, as in (3b), which roughly correspond to English locative inversion constructions. In clauses with SF-adjectival predicates, as in (3c), this order is commonly found with the locative or dative experiencer in the initial position. Thus, the ordering of nominal and clausal constituents allows us to present all the stimuli in a uniform way, with the target linearly preceding the trigger.

(3)   a. požil-aja   sosedk-a
            aged-F.SG   neighbor(F)-NOM.SG
            'an old neighbor'

       b. K nam   pribeža-l-a   sosedk-a.
            to 1PL.DAT ran-PST-F.SG neighbor(F)-NOM.SG
            'The neighbor came running to us.'

       c. Nam   prijatn-a   sosedk-a.
            1PL.DAT pleasant-F.SG neighbor(F)-NOM.SG
            'We like the neighbor.' (lit.: 'The neighbor is pleasant to us.')

Let us now turn to the syntactic aspects of agreement on adjectives and verbs in Russian, in order to verify that the relevant agreeing elements occur in structural positions that align with external and internal agreement as introduced in Section 2.1.

With respect to modifying adjectives, details aside, formal analyses of the syntax of adjectives fall into two main classes: specifier analyses, according to which the adjective is in the specifier of a dedicated functional projection (Cinque 1994; Carstens 2000; Alexiadou 2001; Bonet et al. 2015), and adjunction analyses according to which adjectives are adjoined to the noun phrase (e.g., Babby 1975; Bošković 2005; Bailyn 2012, for Russian and other Slavic languages; Bošković 2016; Carstens 2016, for general analyses; Radford 1988, Kramer 2009; a.o.). Under both types of analyses, the adjective is phrasal (see Section 2.1 on internal agreement); while analyses that place adjectives in the head position have been proposed (e.g., Abney 1987), they have not received much traction.

Russian SF adjectives have received a number of analyses in the literature. Some researchers propose that the SF adjective is dominated by the predicative head (cf. Bailyn 1994; 2012), others propose a special functional head (Graschenkov 2018, p. 147). Here we assume the analysis by Bailyn (1994; 2012); the crucial takeaway is that the SF adjective stands in the same structural relationship with the subject as a verbal predicate, namely, the structure that entails external agreement.

In finite verb structures, subject-verb agreement is achieved via standard probing from the inflectional head T for agreement features; a crucial condition on this probing is c-command, in line with what we expect for instances of external agreement.

Thus, various linguistic considerations converge on selecting Russian as the language in which to investigate internal vs. external agreement: agreement instantiated on verbs as well as SF and LF adjectives, flexible word order, and evidence from theoretical investigations of the syntax of these constructions that they align with the structural distinctions between external and internal agreement discussed in Section 2.1. We would like to underscore that the study presented here is just the first step, and we hope that other studies, with different languages, will follow.

2.5 The present study

The goal of both studies presented herein is to determine whether agreement mismatches lead to different strengths of response on three types of elements: SF adjectives, LF adjectives, and verbs. The strength of response to a violation is measured using first SPR (Section 3) and then eye-tracking while reading (Section 4). Both methodologies give insight into online processing by measuring the speed of processes associated with reading. SPR is very commonly employed in linguistics as it can be successfully conducted online while still yielding useful results, while eye-

tracking while reading is more resource-intensive but offers a more time-sensitive measure and allows regression to earlier text, potentially identifying effects that may be too fine-grained or localized for SPR to capture (Ferreira & Henderson 1990; Jackson et al. 2012; Witzel et al. 2012).

The key variable in these studies is the response time at the critical region (mismatched agreement-triggering noun) and at spillover regions. If the type of agreement (internal vs. external) – and thus structural position of the agreeing element – matters for the processing of agreement, then we can expect differences between the effect of mismatch in agreement on an LF (modifying) adjective versus on a predicate, be it a verb or SF adjective.

It is also conceivable, particularly given the results in Hopp & Lemmerth (2019) (cf. Section 2.3), that the lexical category of the agreeing element may play a role in processing agreement. In other words, it may be that differences in processing between (modifying) adjectives and verbs are not due to the difference in their structural position but rather non-structural properties of adjectives and verbs. This is where the SF adjectives in Russian are particularly useful. If the lexical category rather than structural position of the agreement target modulates the processing (mis)matches in agreement, then we expect SF and LF adjectives to pattern together to the exclusion of verbs in terms of relative slowdown in processing given a mismatch in agreement features. It is also possible that both the lexical category and the structural position of an agreeing constituent may matter. In that case, the slowdown after a mismatch should exhibit a three-way difference.

To summarize, we can build the following hypotheses[5] concerning potential contrasts in the effect of mismatches in agreement on processing, depending on the lexical category of the agreeing element and/or whether it instantiates internal vs. external agreement (i.e., the structural position of the agreeing element).

Hypothesis 1: Structural position matters in processing of agreement features
LF (modifying) adjective $\neq$ SF (predicative) adjective $=$ verb

Hypothesis 2: Lexical category matters in processing of agreement
LF (modifying) adjective $=$ SF (predicative) adjective $\neq$ verb

Hypothesis 3: Structural position and lexical category matter in processing of agreement
LF (modifying) adjective $\neq$ SF (predicative) adjective $\neq$ verb

---

[5] We note that there are additional possibilities for how results from the three agreeing elements may pattern, including that the modifying adjective and verb pattern together to the exclusion of the predicative adjective. However, these logical possibilities are difficult to link to hypotheses regarding possible syntactic factors, so we do not include them here.

3 Experiment 1: Self-paced reading

3.1 Participants

73 participants in Moscow, Russia, took part in the experiment. We excluded one participant who did not fit the inclusion criterion of age for the study. Thus, we analyzed data from 72 participants between 17 and 40 years old. All participants had high question-answering accuracy (fillers: mean accuracy = 0.98, sd = 0.04; experimental items: mean accuracy = 0.95, sd = 0.08). No participants were excluded from analysis based on accuracy on comprehension questions.[6] Participants were recruited online and were instructed to read at a natural pace and answer questions as accurately as possible.

3.2 Design and Materials

The participants' task was to read the sentences in a phrase-by-phrase manner and answer yes/no comprehension questions that followed 1/3 of them. Their reading times were measured. The design was 2 × 3, with Grammaticality (matching vs. mismatching) crossed with Agreement Type (LF adjective vs. verb vs. SF adjective).

The materials consisted of sentences designed for SPR, with six regions each. The first region was an adverb or a prepositional phrase (or a composite noun in the dative in the SF-adjective condition); Region 2 included the agreeing element (LF adjective, verb, or SF adjective), and the third position included the noun that the word in Region 2 agreed with (the critical region) (**Table 3**). The subsequent material varied depending on clause type, but all the stimuli were uniform in having no commas or dashes.

R1	R2 (agreement target)	R3 (agreement trigger)	R4	R5	R6
PP	LF (modifying) adjective	Noun (masc. or fem.)	**Spillover 1**	**Spillover 2**	
	agreeing verb				
NDAT	SF (predicative) adjective				

Table 3: Structure of SPR stimuli.

There were 48 stimuli sets (16 for each type of agreement target, illustrated in (4) – (6)); the nouns were balanced by animacy, with equal numbers of animate and inanimate nouns. All the stimuli were manipulated to occur with a masculine head noun and a feminine head noun in the critical region, and the nouns were maximally matched for length in every pair. The agreeing

---

[6] One participant had an accuracy of 43% on experimental items, but given that their accuracy on filler items was high (94%), they were not excluded from analysis.

element in Region 2 appeared in the gender-matched (grammatical) and gender-mismatched (ungrammatical) form.

(4) a. **LF (modifying) adjective condition, grammatical**
Na ulice {golodn-aja   bolonka} /   {golodn-yj   doberman}
on street hungry-F.NOM.SG Maltese(F).NOM hungry-M.NOM.SG Doberman(M).NOM
lajet gromk-im laj-em.
barks loud-INS   bark-INS
'The hungry Maltese/Doberman has been barking loud in the street.'

  b. **LF (modifying) adjective condition, ungrammatical**
*Na ulice {golodn-yj   bolonka} /   {golodn-aja   doberman}
on street hungry-M.NOM.SG Maltese(F).NOM hungry-F.NOM.SG Doberman(M).NOM
lajet gromk-im laj-em.
barks loud-INS   bark-INS

(5) a. **verb condition, grammatical**
U babuški {lajal-a   bolonka} /   {lajal-Ø   doberman}
by Grandma barked-F.NOM.SG Maltese(F).NOM barked-M.NOM.SG Doberman(M).NOM
s   samogo utra.
from very   morning
'Grandma's Maltese/Doberman had been barking since early morning.'

  b. **verb condition, ungrammatical**
*U babuški {lajal-Ø   bolonka} /   {lajal-a   doberman}
by Grandma barked-M.NOM.SG Maltese(F).NOM barked-F.NOM.SG Doberman(M).NOM
s   samogo utra.
from very   morning

(6) a. **SF (predicative) adjective condition, grammatical**
Tete Maše   {protivn-a   bolonka}   / {protiven-Ø
Aunt Masha disgusting-F.NOM.SG Maltese(F).NOM   disgusting-M.NOM.SG
doberman}   iz-za   bespreryvnogo laja.
Doberman(M).NOM because.of endless   bark
'The Maltese/Doberman is annoying to Aunt Masha because of its endless barking.'

  b. **SF (predicative) adjective condition, ungrammatical**
*Tete   Maše   {protiven-Ø   bolonka}   / {protivn-a
Aunt.DAT Masha.DAT disgusting-M.NOM.SG Maltese(F).NOM   disgusting-F.NOM.SG
doberman}   iz-za   bespreryvnogolaja.
Doberman(M).NOM   because.of endless   bark

Note that LF adjectives occur only as modifying adjectives in the stimuli. While in principle LF adjectives can be used predicatively (cf. Section 2.4), a number of restrictions on this use prevents us from being able to construct comparable stimuli for the purposes of the present study.[7] Accordingly, we could not test LF adjectives as predicates in this experiment.

The fillers included 90 sentences of comparable length to the stimuli; half of the fillers were grammatical, and half had violations in prepositional case forms.[8] Stimuli were separated into four lists, such that each participant read 2 practice sentences and 48 experimental sentences and 90 fillers, 140 items in total.

3.3 Procedure

The sentences were presented using the PCIbex online platform for behavioral studies (https://farm.pcibex.net/). Each trial began with a sentence in which all words were masked with dashes while spaces and punctuation marks remained intact. Participants pressed the spacebar to reveal a word and re-mask the previous one. Roughly one third of all items (33 experimental sentences and 16 grammatical fillers) were accompanied by forced choice yes/no comprehension questions. Participants pressed 'f' to choose the 'yes' answer and 'j' to choose the 'no' answer. Correct answers were counterbalanced. Participants were not informed in advance that sentences would contain errors. The instructions were presented in Cyrillic.

3.4 Analysis

Statistical analyses were performed using R version 4.1.2 (R Core Team, 2021) and the *lme4* package (Bates et al. 2015). We log-transformed reading times in order to achieve residuals more closely following a normal distribution, to thus meet assumptions for linear mixed effects models (Baayen & Milin 2010; Nicklin & Plonsky 2020).[9] Finally, because not all words in the regions of interest were of the same length, we fit the following linear model to all available data (experimental sentences and fillers) in order to account for effects of word length on reading times:

---

[7] The use of an LF adjective as a predicate requires the use of the verb *be* – in the past and future tense it is overt, and in the present tense it is null. In the present tense, the LF adjective must follow the subject noun, otherwise it is interpreted as a modifier. The order would therefore make this condition quite different from the others.

[8] As an anonymous reviewer points out, 50% of all stimuli (including targets and fillers) were ungrammatical in some way, which is higher than in some self-paced reading studies investigating the processing of agreement errors. Still, robust effects of grammaticality (per Section 3.5) suggest that participants remained sensitive to errors. We refer the reader to Hammerly et al. (2019) for discussion of how the proportion of ungrammatical stimuli and the phrasing of instructions may affect responses to ungrammatical stimuli.

[9] Following Nicklin & Plonsky (2020), we did not remove outliers, given that log-transformation has been shown to resolve issues with skewness of data without the need for removal of potentially meaningful observations.

log_rt ~ word_length + (1 | subj)

Residuals from this model were used as the dependent variable in the main analysis.

We analyzed residualized log-transformed reaction times in critical (R3) and spillover (R4–R5) regions. A maximal model was fitted to the data (Barr et al. 2013) but resulted in a singular fit, so the model was simplified by simplifying the random effect structure (Bates et al. 2015). The final model (Section 3.5.1) predicted residualized logRTs by grammaticality, agreeing element, and their interaction, as well as random intercepts grouped by subject and by item. In models fitted to reading times in R3 and R4, the fitted model also included a random slope for grammaticality grouped by subject; models fitted to data in R5 including a random slope for grammaticality grouped by subject resulted in singular fit. Grammaticality (grammatical, ungrammatical) and agreeing element (LF adjective, SF adjective, verb) were both sum-coded. *P*-values were estimated using the Satterthwaite method as implemented in the R package *lmerTest* (Kuznetsova et al. 2017).

Since our hypotheses included specific oppositions of SF adjectives and verbs versus LF adjectives (structural position; Hypothesis 1) and of LF/SF adjectives versus verbs (lexical category; Hypothesis 2), we fit two additional sets of models, respectively, using the same dependent variable and model fitting procedures. Lexical category and structural position were factors and were also coded using sum coding.[10]

- Resid_log_rts ~ grammaticality * structural_position + (1|subj) + (1|item) (Section 3.5.2)
- Resid_log_rts ~ grammaticality * lexical_category + (1|subj) + (1|item) (Section 3.5.3)

3.5 Results

3.5.1 Grammaticality

We first analyze the data for effects of grammaticality, with a three-way distinction between type of agreeing element, before investigating sub-groupings of these elements that align with our hypotheses (Sections 3.5.2 and 3.5.3). Visual presentation of the results can be found in **Figure 2** (grammaticality results) and **Figure 3** (grammaticality and agreeing element type).

Results of modeling are provided in **Table 4**. Expected effects of grammaticality were found at the critical region and both spillover regions, whereby grammatical sentences were consistently read faster than ungrammatical sentences. There was also a significant effect of agreeing element at the critical region and spillover region 1 that suggests SF adjectives were read significantly slower than the grand mean of reading times. At spillover region 2 this effect was only trending, but there was an additional effect of agreeing element type wherein LF adjective conditions were

---

[10] (Lexical category: adjectives (LF and SF) = 1, verb = –1. Structural position: modifier (LF adjectives) = 1, predicate (verbs and SF adjectives) = –1).

read significantly faster than the grand mean. There were no significant interaction effects at any of the regions.
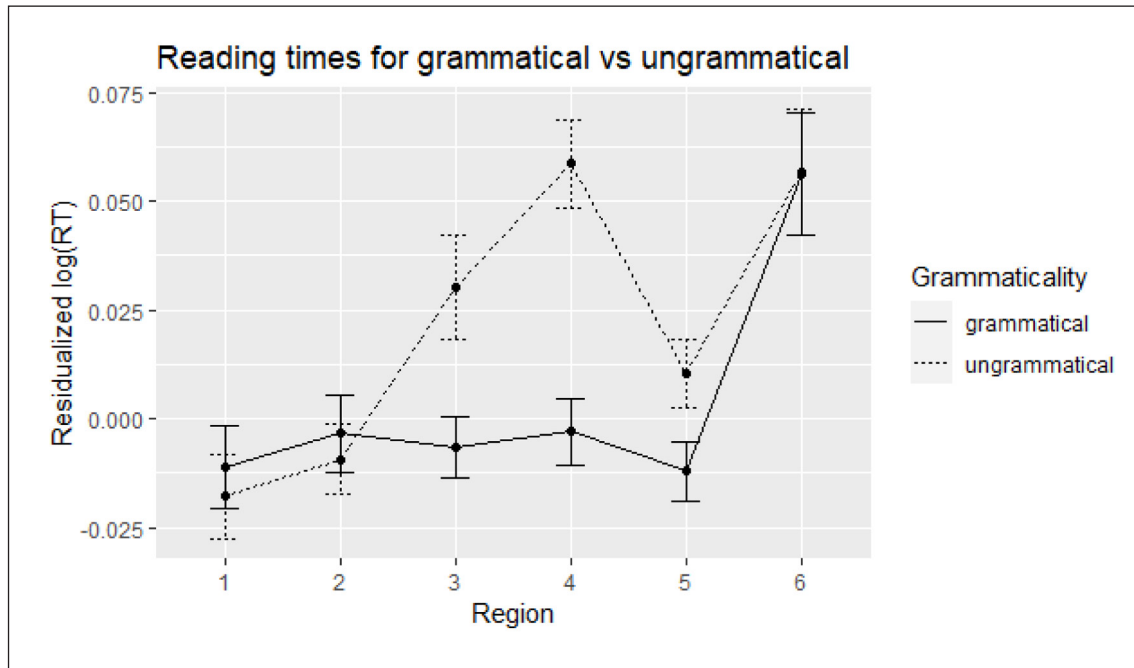

Figure 2: Residualized reading times, grammatical vs ungrammatical conditions.
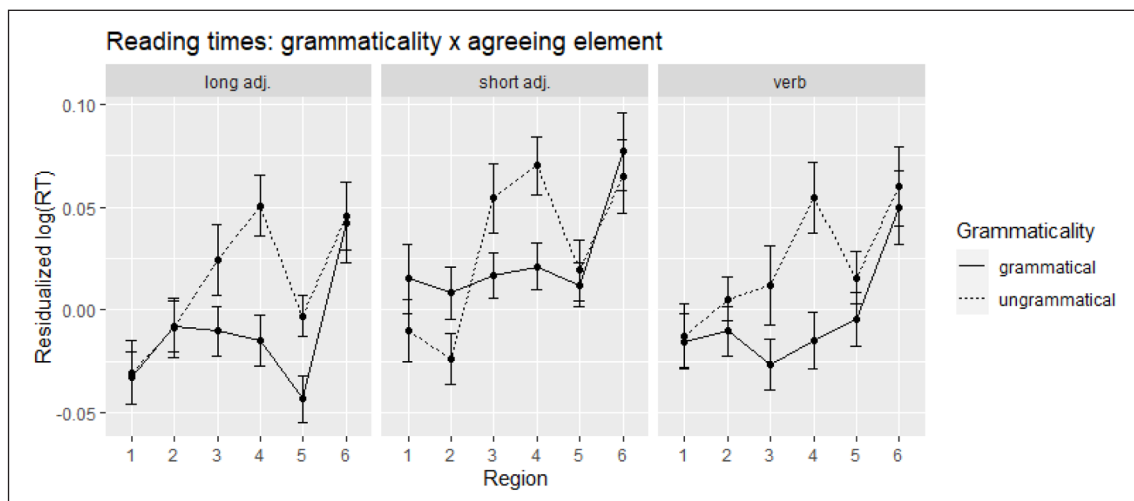

Figure 3: Residualized reading times, grouped by type of agreeing element (LF adjective, SF adjective, verb).

	critical region				spillover region 1				spillover region 2			
	est.	SE	t	p	est.	SE	t	p	est.	SE	t	p
(Intercept)	0.012	0.008	1.49	0.14	0.028	0.008	3.64	<0.001	–.001	0.008	–.10	0.92
Gram	–.018	0.007	–2.58	0.01	–.031	0.005	–5.83	<0.001	–.011	0.004	–2.50	0.01
Agr Type 1	–.005	0.010	–.49	0.63	–.010	0.008	–1.20	0.23	–.022	0.009	–2.48	0.02
Agr Type 2	0.024	0.010	2.41	0.02	0.018	0.008	2.17	0.04	0.016	0.009	1.81	0.08
Gram × Agr Type 1	0.001	0.008	0.17	0.87	–.002	0.007	–.28	0.78	–.009	0.006	–1.43	0.15
Gram × Agr Type 2	–.0002	0.008	–.03	0.98	0.006	0.007	0.86	0.39	0.008	0.006	1.31	0.19

Table 4: Results of statistical modeling: grammaticality × agreeing element type. Significant effects are marked in gray.

3.5.2 Structural position

The fixed effects for the fitted models predicting residualized log-transformed reading times by grammaticality, structural position (modifier versus predicate), and their interaction are presented in **Table 5**. The models revealed expected effects of grammaticality in the critical and spillover regions. There was a significant effect of structural position at the second spillover region (*est.* = –0.017, SE = 0.007, t = –2.50, p = 0.02), wherein sentences with modifier adjectives were read significantly faster at R5 than sentences with a verb or predicative adjective. No other significant effects of structural position or its interaction with grammaticality were observed in the results.

3.5.3 Lexical category

The fixed effects for the fitted models predicting residualized log-transformed reading times by grammaticality and lexical category – LF-adjective and SF-adjective versus verb – are presented in **Table 6**. The models revealed expected effects of grammaticality in the critical and spillover regions. There was only a trending effect of lexical category at the critical region (*est.* = 0.014, SE = 0.008, t = 1.88, p = 0.07). No other significant effects of lexical category or its interaction with grammaticality were observed in the results.

	critical region				spillover region 1				spillover region 2			
	est.	SE	t	p	est.	SE	t	p	est.	SE	t	p
(Intercept)	0.011	0.009	1.22	0.23	0.025	0.008	3.15	0.003	–.006	0.008	–.80	0.43
Gram	–.018	0.007	–2.43	0.02	–.031	0.006	–5.60	<0.001	–.013	0.005	–2.84	0.005
StrucPos	–.004	0.008	–.46	0.65	–.007	0.006	–1.19	0.24	–.017	0.007	–2.50	0.02
Gram × StrucPos	0.001	0.006	0.17	0.86	–.002	0.005	–.28	0.78	–0.007	0.005	–1.43	0.15

Table 5: Results of statistical modeling: grammaticality × structural position. Significant effects are marked in gray.

	critical region				spillover region 1				spillover region 2			
	est.	SE	t	p	est.	SE	t	p	est.	SE	t	p
(Intercept)	0.007	0.008	0.84	0.40	0.026	0.008	3.20	0.002	0.001	0.008	0.10	0.92
Gram	–0.019	0.007	–2.52	0.01	–0.032	0.006	–5.70	<0.001	–0.011	0.005	–2.32	0.02
LexCat	0.014	0.008	1.88	0.07	0.006	0.006	0.93	0.36	–0.005	0.007	–0.64	0.52
Gram × LexCat	0.001	0.006	0.14	0.89	0.003	0.005	0.59	0.56	–0.001	0.005	–0.12	0.90

Table 6: Results of statistical modeling: grammaticality × lexical category. Significant effects are marked in gray.

3.6 Interim summary of results

At this point, the results from the SPR task suggest that overall, participants are sensitive to mismatches in agreement between the probe and the target. Effects of grammaticality show that participants access the gender feature on a gender-inflected agreeing element (verb, LF or SF adjective) and slow down when they subsequently encounter a noun that mismatches this feature, i.e., an ungrammatical condition. However, the lack of an interaction between grammaticality and type of agreeing element shows that the present study failed to observe any difference in strength of the violation generated by the three different agreeing elements tested. Main effects of agreeing element type (Section 3.5.1) or structural position (Section 3.5.2) suggest that certain constructions may take more time to process – e.g., modifier + noun may be overall faster and thus easier than predicate + noun – but in the absence of an interaction effect these do not inform the main research questions.

There is the possibility, of course, that SPR may not be sensitive enough to capture the relevant effects, if they exist. Thus, in the next section, we use the same materials in an

eye-tracking-while-reading study. This methodology offers a more fine-grained measure of processing (or processing slowdowns) revealing not only the time course of the processing slowdowns within the region of interest but also the manner in which the processing slowdowns are manifested (i.e., fixation durations vs. regressive saccades to previous regions).

The general trend in our expectations for all eye-movement measures in the eye-tracking-while-reading version of the study parallels those for the SPR measures. First and foremost, similar to the findings of the SPR experiment, we anticipate identifying an effect of grammaticality, particularly in late eye-movement measures (see Section 4.4), as an indication of morphosyntactic reanalysis. Specifically, for ungrammatical sentences, we expect critical and spillover regions to result in longer total reading times, as well as higher probabilities of regression to the preceding regions.

4 Experiment 2: Eye-tracking while reading

4.1 Participants

Forty one Russian speakers between the ages of 18 and 40 years old took part in the experiment. They were tested at the first author's lab in Los Angeles, California, in the USA; recruitment was by word of mouth and by ads in the local community. We limited participation to Russian speakers who had been in the USA less than three years. We removed data from one participant who was not Russian-dominant. Thus, we analyzed data from 40 participants. All participants included in the analysis had high accuracy on the comprehension questions (at least 93% correct).

4.2 Design and materials

Design and materials for this study were the same as those for the SPR experiment, discussed in Section 3.2.

4.3 Procedure

The sentences were displayed on a 24-inch monitor with the following specifications: resolution of 1920 × 1080 pixels, frame rate of 144 Hz, and a font face of 22-point Courier New. The monitor was controlled by a ThinkStation computer. We used the Eyelink 1000 Plus desktop mount eye-tracker (SR Research) to record eye movements, using a chin rest for head stabilization. The distance between participants and the monitor constituted 35 inches / 90 cm. Only the left eye was tracked, at 1000 Hz rate. The experiment began with a 9-point calibration, which was repeated when eye fixations deviated. Each trial began with a drift correction, where a fixation point appeared to the left of the location of the first letter of the first word. If the drift correction was successful, the experiment progressed to sentence presentation; otherwise, calibration was repeated. Participants indicated the completion of reading a sentence by directing their gaze

to a red dot in the bottom right-hand corner of the screen, after which the trial proceeded to a comprehension question or to the next sentence. Participants clicked on 'yes' or 'no' options to answer the comprehension questions. Again, participants were instructed to read at a pace they were comfortable with and answer questions as accurately as possible. The sentences were presented in a randomized order.

4.4 Analysis

All analyses were conducted using R version 4.2.2 (R Core Team 2022). Linear mixed-effects models were implemented using *lme4* (Bates et al. 2015). Before analysis, all trials that required recalibration were excluded; we also excluded fixations less than 80 ms. We analyzed eye-movement characteristics in pre-critical, critical, and spillover regions.

   We focused on the following *early* fixation-duration measures (i.e., measures that are believed to reflect initial lexical processing – letter-to-sound conversion and initial word recognition) (log-transformed) as our dependent variables:

- single fixation duration (SFD, the region is fixated only once during first-pass reading),
- first fixation duration (FFD, the duration of the very first fixation within the region),
- gaze duration (GD, the sum of the durations of all fixations within the region before eyes move somewhere else out of the region)

Among *late* measures (indicative of post-lexical processing, including morphosyntactic- and semantic-information integration and revision of the information from previous regions) we analyzed:

- total reading time (TT, the sum of all fixations within the region);
- regression-in probability (Rin, the probability of the region receiving regressions from later regions);
- regression-out probability (Rout, the probability of the region being the origin of a saccade to earlier regions, regardless of whether later regions were fixated or not).

For each region, we fitted a series of (generalized) linear mixed-effects models with each of the eye-movement measures as an outcome (all duration measures were log-transformed) and the following fixed effects: grammaticality (match vs. mismatch), type of agreeing element (LF adjective, SF adjective, verb), the interaction between grammaticality and type of agreeing element, and target gender (feminine vs. masculine). All categorical fixed effects were sum-coded. Additionally, the models included continuous fixed effects of frequency (log-transformed),[11] length (centered and scaled), previous and next region length (centered and scaled) and trial

---

[11] Individual word form frequency was obtained from Lyashevskaya & Sharov (2009); the frequency for collocations was obtained from RuSKELL corpus (Apresjan et al. 2016).

number (centered and scaled) to account for practice effects. The full random structure in all of the models included random intercepts by participant and item as well as random slopes for interaction between grammaticality and agreeing element. We simplified the random structure for each model starting with interactions and followed by main effects if the models resulted in a singular fit. All p-values were adjusted for multiple comparisons using Bonferroni correction at an α-level of .004. The full structure of each model and the resulting output is presented in the Supplementary Materials (Tables S1–S4). Descriptive statistics for each eye-movement measure in each region are presented in **Table 7**.

Region		**Grammaticality**	
		Match	**Mismatch**
R2 (Pre-Critical)	SFD	252 (84)	246 (80)
	FFD	234 (87)	234 (85)
	GD	301 (164)	293 (149)
	TT	507 (360)	594 (425)
	Rin	0.29 (0.46)	0.48 (0.50)
	Rout	0.28 (0.45)	0.26 (0.44)
R3 (Critical)	SFD	238 (80)	255 (98)
	FFD	236 (90)	239 (95)
	GD	291 (176)	306 (163)
	TT	437 (343)	521 (374)
	Rin	0.14 (0.35)	0.20 (0.40)
	Rout	0.24 (0.43)	0.44 (0.50)
R4 (Spillover 1)	SFD	248 (79)	254 (85)
	FFD	240 (92)	235 (90)
	GD	345 (210)	343 (192)
	TT	480 (345)	513 (352)
	Rin	0.16 (0.37)	0.17 (0.38)
	Rout	0.15 (0.36)	0.24 (0.42)
R5 (Spillover 2)	SFD	246 (88)	239 (78)
	FFD	232 (83)	233 (88)
	GD	322 (179)	323 (174)
	TT	429 (315)	429 (306)
	Rin	0.17 (0.37)	0.18 (0.38)
	Rout	0.25 (0.43)	0.25 (0.43)

Table 7: Descriptive statistics for all fixation durations (ms) and regression probability by region; standard deviations are in parentheses.

4.5 Results

In all of the regions (R2, R3, R4) except the second spillover region (R5) there was an effect of grammaticality in TT and one or both of the regression probabilities (Rin and Rout), which are late eye-movement measures. These results indicate that, as in the SPR study (Section 3), participants were sensitive to violations in agreement at the stage of post-lexical morphosyntactic-information integration: mismatch condition trials led to longer reading times and more regressions compared to match condition trials (see Tables S1-S4 in Supplementary materials).

4.5.1 Pre-critical region (R2)

There was an effect of agreeing element on TT and Rin measures. To unpack the effect, we performed a set of post-hoc multiple comparison analyses using the estimated marginal means (EMMs) from the *emmeans* package (v1.8.3., Lenth 2021) in R (Bonferroni-corrected for three pairwise comparisons). The post-hoc comparisons showed that compared to verbs, in TT measure, SF adjectives received longer fixation durations (*est.* = 0.257, SE = 0.08, t = 3.39, p = 0.004). Similarly, SF adjectives elicited higher probability of regressions compared to verbs (*est.* = 0.831, SE = 0.21, z = 3.88, p < 0.001) and compared to LF adjectives (*est.* = –0.51, SE = 0.21, z = –2.42, p = 0.047). No other effects reached statistical significance. See Table S1 in Supplementary materials for full results.

4.5.2 Critical region (R3)

Nouns that were preceded by SF adjectives elicited longer total reading times and higher rates of regressions to previous regions compared to nouns that followed verbs (TT: *est.* = 0.265, SE = 0.05, t = 5.1, p < 0.001; Rout: *est.* = 0.789, SE = 0.19, z = 4.08, p < 0.001) or LF adjectives (TT: *est.* = –0.254, SE = 0.06, t = –4.56, p < 0.001; Rout: *est.* = –0.707, SE = 0.21, z = –3.41, p = 0.002). However, there were no differences between LF adjectives and verbs in TT or Rout measures (after Bonferroni correction, all ps = 1.00).

There was also an interaction effect between grammaticality and agreeing element in the early SFD measure (the measure is believed to reflect word recognition time including semantic activation, Juhasz & Rayner 2003): In the mismatch condition, the noun received longer single fixations if the word in R2 was an SF adjective, compared either to LF adjectives (*est.* = –0.147, SE = 0.05, t = –3.11, p = 0.007) or verbs (*est.* = 0.235, SE = 0.04, t = 5.31, p < 0.001) (**Figure 4**). There was no difference in SFD between verbs and LF adjectives (p = 0.148). The full model output is presented in Table S2 in the Supplementary Materials.
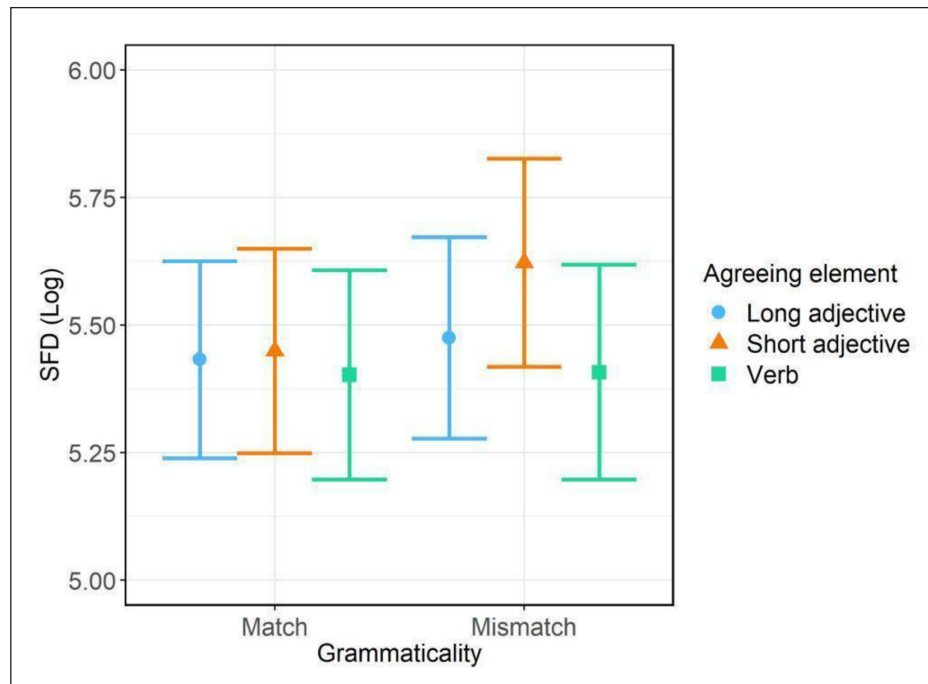
Figure 4: Single Fixation Duration (SFD) as a function of grammaticality and agreeing element in R3. Bars represent 95% CIs.

4.5.3 Spillover regions (R4 and R5)

Tables S3 and S4 in Supplementary materials present the full output for the models for these regions. In the first spillover region (R4), we observed an interaction in the early FFD measure (the measure is believed to reflect early word recognition processes; Juhasz & Rayner 2003) between agreeing element and grammaticality: post hoc analyses indicated that in the match condition, when R2 included SF adjective, R4 received longer first fixation durations compared to when R2 included LF adjectives (*est.* = –0.092, SE = 0.04, t = –2.50, p = 0.045), while all other agreeing elements did not show significant differences either in match or mismatch conditions (ps > 0.318).

In R5, regardless of grammaticality, there were fewer regressions to previous regions (Rout) if R2 was an LF adjective, and the difference was significant compared both to SF adjectives (*est.* = –0.670, SE = 0.23, z = –2.93, p = 0.010) and verbs (*est.* = –0.639, SE = 0.25, z = –2.55, p = 0.033). Given that an increase in the frequency of regressions is believed to reflect the need for morphosyntactic reanalysis and semantic (re)integration processes, these findings indicate that LF adjectives in R2 reduced the frequency of participants needing such revisions while reading the sentences.

4.6 Interim summary of results

Overall, the results of the eye-tracking study suggest that readers were sensitive to violations at the post-lexical stages of processing as indicated by late eye-movement measures in regions R2, R3, and R4. However, we did not observe clear-cut differences in the effects of the lexical category or structural position of the agreeing element. Additionally, we did find effects of SF adjectives on both early and late measures in the pre-critical and critical regions. Specifically, when encountering an SF adjective, readers fixated on it and the words following it for longer durations and regressed more to re-read previous regions. In the critical region (R3), we also observed an interaction wherein nouns in the mismatch condition received longer single fixations when preceded by SF adjectives, although this is typically a measure of early processing stages and is therefore likely not informative about resolving mismatches in syntactic agreement features. Taken together, these findings indicate that SF adjectives posed greater processing difficulty compared to LF adjectives or verbs.

5 Discussion

5.1 Overview

The experimental studies presented above used two different methodologies to investigate the relative strength of violations when readers encounter a noun that mismatches in agreement an earlier adjectival or verbal element inflected for gender. Both the SPR study and the eye-tracking-while-reading study found that participants are sensitive to this agreement marking, and that, when the features on the two linguistic elements are mismatched, participants' reading times significantly slow down, and that in the eye-tracking study they are more likely to have to regress to earlier parts of the phrase. These effects of grammaticality are expected and show that participants are overall sensitive to agreement errors in the experimental paradigms employed here, consistent with previous work on real-time processing of gender agreement (mis)matches in Russian (Slioussar & Malko 2016; Romanova & Gor 2017).

The studies also found some main effects of agreeing element type, suggesting that overall reading times at the agreeing element or later were modulated by the type of agreeing element. Evidence from both the SPR study and the eye-tracking-while-reading study converges to suggest that conditions in which the agreeing element was a predicative (SF) adjective were read overall slower and caused more regressions than verbs and modifier (LF) adjectives. Recall that the frequency of the lexical items themselves was accounted for in the study. Moreover, a search of the Russian National Corpus revealed that – while adjectives were overall much less frequent than agreeing verbs, as expected given that adjectives are optional elements – the frequency of the two adjectival constructions relative to each other was about the same. That is, LF adjectives were about as frequent as SF adjectives preceding a subject (**Table 8**).

Agreement form	Number of occurrences
Agreeing LF adjective, M or F	1,944
Agreeing SF adjective, M or F, predicate-subject order	1,538
Agreeing verb in past tense, M or F, predicate-subject order	25,093

Table 8: Types of agreement forms used in the experiment, number of raw occurrences in the syntax subcorpus of the Russian National Corpus (total number of sentences: 77,105).

One possible explanation of the longer reading times for the SF adjective constructions is that, unlike the experimental conditions with verbs and LF adjectives, these constructions included another nominal phrase (in R1; cf. ex. (6)) that participants may have considered to be a possible agreement trigger. Resolving this may have added additional processing costs. To be more specific, consider example (6b). When participants encountered in R3 a noun that mismatched the SF adjective in R2, they may have had in memory the noun in R1 and proceeded to attempt to resolve the mismatch between R2 and R3 by attempting to match features on the adjective in R2 with the (dative) noun in R1. Indeed, further inspection of results suggests ungrammatical SF adjective conditions were read numerically faster when the noun in R1 matched the SF adjective in gender than when it did not. However, we note that this was not subject to formal statistical analysis and this discussion is thus strictly exploratory. We leave it to future work to investigate why SF adjective constructions may have evoked the additional processing difficulties reported above.

As mentioned in Sections 3.6 and 4.6, despite the observed main effects of grammaticality and agreeing element type, there were crucially no interactions between grammaticality and agreeing element type in the SPR results or in the measures of late processing stages in the eye-tracking results that would point to mismatches on different agreeing elements yielding different degrees of processing difficulties. Such interactions would be most informative for the research questions of this study. Instead, it appears that gender agreement mismatches on verbs, LF adjectives, and SF adjectives all resulted in a similar degree of processing difficulties once participants encountered the mismatched noun. We discuss the implications of this finding in the next section.

5.2 Implications for processing of agreement

The lack of an interaction between an agreeing element and grammaticality constitutes a null result observed in two studies using different methodologies. As is broadly acknowledged, interpreting a null result is notoriously difficult. The null result could, of course, be an instance of a Type II error, or a false negative, whereby in truth there is a difference between the relevant conditions (or in the case of the studies above, a true interaction effect), but the given method and analysis were unable to observe this difference; often such circumstances arise due to lack

of power. Our findings also contrast with the initial results from German (Hopp & Lemmerth 2018; Lemmerth & Hopp 2019) and Spanish (Barber & Carreiras 2005) which suggest that such differences may indeed be observable. It may be, therefore, that a study conducted with higher power may observe the interaction effects that the present studies do not, and that such a study could directly bear on the research questions and hypotheses set out in Section 2.5.

However, given the cross-methodological replication of the results presented here, it is worth exploring the possible implications of a reliable null result, should the outcomes of future work be consistent with the present findings. First, one could suggest that the difference between internal and external agreement is not supported; however, this should be viewed in the context of the broad theoretical literature on agreement. Primary syntactic evidence in favor of the contrast between external and internal agreement is strong; it includes differences in intervention (external agreement is subject to intervention effects, while internal agreement is insensitive to them); long-distance relations (present in external but not internal agreement); closest conjunct effects (present in external but not internal agreement); the availability of omnivorous agreement (Nevins 2011; Preminger 2014) in external but not internal agreement; and the content of features matched (see Norris 2014; 2017; 2018, and further references therein). Further still, there is processing evidence in support of the psychological reality of c-command (e.g., Cunnings at al. 2015; Kush et al. 2015) and locality (e.g., Bartek et al. 2011; Friedmann et al. 2017), the two theoretical notions crucial for the contrast between external and internal agreement.

Should no difference in terms of real-time comprehension of the outcomes of internal versus external agreement processes continue to be observed then, these ought to be aligned with theoretical considerations to differentiate between them. Two logical possibilities arise here. First, as suggested to us by an anonymous reviewer, the structural difference may be real, but it does not translate into a time-course contrast. To use a metaphor from a non-linguistic domain, it may take a person the same time to write a one-page abstract and to prepare a meal, but that does not mean that the two processes have the same underlying mechanism. If we take this possibility seriously, that limits the range of linguistic mechanisms that can be compared on the basis of the time course of processing and may call for different methodologies to capture contrasts undergirded by certain types of theoretical considerations.

A second possibility is that, taken together, theoretical and experimental results point to a critical insight regarding real-time processing of agreement. Specifically, the discrepancy between experimental and theoretical findings may be reconciled by thinking more carefully about the implications of a composite approach to agreement. Recall the two-step model of agreement, which consists of establishing a particular relation between two segments in the syntax (formalized as Agree-Link) and checking if the features on these segments actually match (Agree-Copy); in **Table 9**, we reproduce **Table 1** from Section 2.2, which summarizes internal vs. external agreement under this approach. Under a composite approach to agreement, the structural difference between

internal vs. external agreement can be maintained, if we assume that our experimental toolkits are able to tap into only Agree-Copy (see Franck et al. 2010; 2020, for a similar proposal). At the stage of Agree-Link, different operations are available to yield the featural content on relevant linguistic elements, depending on the structural relationships between these elements. At the stage of Agree-Copy though, the featural representations on the linguistic elements on which the process of agreement operates are equivalent; the nature of the operation that derived them earlier no longer plays a role. Our findings indicate that participants notice mismatches in agreement but that this sensitivity is not modulated by the type of agreeing element. If these results hold up in future work, this would be consistent with the composite approach to agreement where parsing is sensitive to feature-matching at the Agree-Copy stage only.

	Agree-Link: different operations	Agree-Copy: feature-matching
External agreement	c-command necessary	feature-matching
Internal agreement	no c-command necessary	feature-matching

Table 9: Two dimensions of agreement.

As stated previously, this interpretation is necessarily tentative, and it would open the door to further questions. In particular, it would be necessary to investigate what experimental methods could tap into the properly syntactic aspect of agreement (Agree-Link) and therefore more directly test the proposed difference between external and internal agreement. Here, more fine-grained methodologies such as neuroimaging may be promising, given the initial evidence from ERPs in Barber & Carreiras (2005); see also Section 2.3. It would also be important to articulate predictions for other phenomena that could be modeled under a two-step model of agreement and where an absence of real-time processing differences would be expected. Here one possible avenue may be to investigate the resolution of different types of agreement (e.g., resolution versus closest-conjunct agreement) under coordination in languages like Polish, a phenomenon that has been argued to be post-syntactic (e.g., Lyskawa 2021). We leave these and other matters for future work.

6 Conclusion

This study offered an experimental exploration of the hypothesized differences between external and internal agreement. Two studies employing self-paced reading (Section 3) and eye-tracking-while-reading (Section 4) measured processing difficulties when native speakers of Russian encounter an element inflected for agreement that mismatches the features on the noun that determines this agreement. The strength of the violations was measured by observing reading slowdowns and regressions that occurred when participants encountered the relevant noun.

Of particular interest given the research questions was whether these measures of processing difficulty in ungrammatical conditions were modulated by the nature of the agreeing element – verb, predicative adjective, or modifying adjective.

Both studies found effects of (un)grammaticality – participants were sensitive to agreement mismatches between the agreeing element and the trigger. Additionally, participants showed differential processing times across the three agreeing element types, indicating overall more difficulty in processing sentences in conditions in which the agreeing element was an SF (predicative) adjective, which we tentatively attribute to the presence of a competing DP in the dative case in that construction alone. However, there was no interaction observed between grammaticality and the agreeing element type, suggesting that, while participants are sensitive to mismatches, the time-course of processing of these mismatches does not differ between the types of agreeing elements.

We highlight that these results were consistent in the cross-methodological replication presented here. Still, power to detect an interaction effect in both studies was low, which limits our ability to argue that the lack of a difference between agreeing element types is meaningful. We hope that future studies, conducted with higher power, may find differences in line with the hypotheses articulated in Section 2.5, and therefore be able to directly inform the research questions. If, however, such higher-powered studies also observe no differences in the processing of agreement mismatches on different agreeing elements, such a null effect may be more reliably interpreted. We offer an overview of the possible implications, to make clear the contributions of such a result, particularly when viewed in light of the broader theoretical literature on external/internal agreement.

First, rather than taking at face value the absence of the critical interaction effect to suggest that internal and external agreement are two sides of the same coin, it is possible that their differences may not be detectable using the time-course of processing. In other words, the processing of the two types of agreement takes the same amount of time, but qualitative differences can still be expected and will have to be explored using different experimental methodologies. Next, it is possible that the lack of time distinctions constitutes a novel argument in support of agreement as a two-step process. The first step takes place in the syntax and is not detectable by measures of the time-course of processing; this is where differences in the syntax of external and internal agreement are encoded. The second step, one that occurs after the syntactic structure is built, consists of checking if the features are sufficiently matched. This step is the one that is reflected in processing; at this point the featural representations on which agreement operates are not differentiated according to the processes that derived them (see Franck et al. 2008, 2010 for a similar approach), and so internal and external agreement are not distinguished. Having laid out what is at stake in experimental investigations of internal versus external agreement, we hope that future studies will continue to probe this question and tease apart the possibilities explored here.

Abbreviations

ACC = accusative, DAT = dative, GEN = genitive, F = feminine, M = masculine, N = neuter, PL = plural, SG = singular

Data availability

All supplementary files, as well as code used for analysis, are available on OSF at https://osf.io/fvsj2/?view_only=e030f7ad340945a49775af563cb1d5cb.

Ethics and consent

Data collection for the self-paced reading study was approved by the University of Maryland's IRB under protocol #766233-26. Data collection for the eye-tracking study was approved by the University of Southern California's IRB under protocol #UP-22-00158. In both studies, participants read an informed consent form and voluntarily agreed to participate anonymously in the study.

Funding information

Acknowledgments

Competing interests

The authors have no competing interests to declare.

References

Abney, Steven. 1987. *The English noun phrase in its sentential aspect*. Cambridge, MA: MIT dissertation.

Ackema, Peter & Neeleman, Ad. 2020. Unifying nominal and verbal inflection: Agreement and feature realization. In Alexiadou, Artemis & Borer, Hagit (eds.), *Nominalization: 50 years on from Chomsky's "Remarks",* 29–52. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198865544.003.0003

Akhutina, Tatiana & Kurgansky, Andrei & Kurganskaya, Marina & Polinsky, Maria & Polonskaya, Natalya & Larina, Olga & Bates, Elizabeth & Appelbaum, Mark. 2001. Processing of grammatical gender in normal and aphasic speakers of Russian. *Cortex* 37(3). 295–326. DOI: https://doi.org/10.1016/S0010-9452(08)70576-8

Alexiadou, Artemis. 2001. *Functional structure in nominals.* Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.42

Apresjan, Valentina & Baisa, Vít & Buivolova, Olga & Kultepina, Olga & Maloletnjaja, Anna. 2016. RuSkELL: Online Language Learning Tool for Russian Language. *Proceedings of the XVII EURALEX International Congress, Tbilisi, Georgia*, 92–299.

Arregi, Karlos & Nevins, Andrew. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout.* Dordrecht: Springer. DOI: https://doi.org/10.1007/978-94-007-3889-8

Baayen, Harald & Milin, Peter. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28. DOI: https://doi.org/10.21500/20112084.807

Babby, Leonard. 1975. *A transformational grammar of Russian adjectives*. The Hague: Mouton. DOI: https://doi.org/10.1515/9783111356822

Bailyn, John. 1994. The syntax and semantics of Russian Long and Short adjectives: An X'-theoretic account. In Toman, Jindrich (ed.), *Annual Workshop on Formal Approaches to Slavic Linguistics: The Ann Arbor Meeting: Functional Categories in Slavic Syntax*, 1–30. Ann Arbor, Michigan Slavic Publications.

Bailyn, John. 2012. *The syntax of Russian.* Cambridge, UK: Cambridge University Press.

Baker, Mark. 2008. *The syntax of agreement and concord.* Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511619830

Barber, Horacio & Carreiras, Manuel. 2005. Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience* 17(1). 137–153. DOI: https://doi.org/10.1162/0898929052880101

Barr, Dale & Levy, Roger & Scheepers, Christoph & Tily, Harry. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bartek, Brian & Lewis, Richard & Vasishth, Shravan & Smith, Mason. 2011. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(5). 1178–1198. DOI: https://doi.org/10.1037/a0024194

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beatty-Martínez, Anne & Dussias, Paola. 2019. Revisiting masculine and feminine grammatical gender in Spanish: Linguistic, psycholinguistic, and neurolinguistic evidence. *Frontiers in Psychology* 5. DOI: https://doi.org/10.3389/fpsyg.2019.00751

Benmamoun, Elabbas & Bhatia, Archna & Polinsky, Maria. 2009. Closest conjunct agreement in head-final languages. *Linguistic Variation Yearbook* 9. 67–88. DOI: https://doi.org/10.1075/livy.9.02ben

Benmamoun, Elabbas & Lorimor, Heidi. 2006. Featureless expressions: When morphophonological markers are absent. *Linguistic Inquiry* 37. 1–23. DOI: https://doi.org/10.1162/002438906775321157

Bhatt, Rajesh & Walkow, Martin. 2013. Locating agreement in grammar: An argument from agreement in conjunctions. *Natural Language and Linguistic Theory* 31. 951–1013. DOI: https://doi.org/10.1007/s11049-013-9203-y

Bobaljik, Jonathan. 2008. Where's phi? Agreement as a post-syntactic operation. In Harbour, Daniel & Adger, David & Bejar, Susana (eds.), *Phi-Theory: Phi features across interfaces and modules*, 295–328. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780199213764.003.0010

Bonet, Eulàlia & Lloret, Maria-Rosa & Mascaró, Joan. 2015. The prenominal allomorphy syndrome. In Bonet, Eulàlia & Lloret, Maria-Rosa & Mascaró, Joan (eds.), *Understanding allomorphy: Perspectives from optimality theory*, 5–44. London: Equinox. DOI: https://doi.org/10.1558/equinox.25215

Bošković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59(1). 1–45. DOI: https://doi.org/10.1111/j.1467-9582.2005.00118.x

Bošković, Željko. 2016. Getting really edgy: On the edge of the edge. *Linguistic Inquiry* 47(1). 1–33. DOI: https://doi.org/10.1162/LING_a_00203

Bruening, Benjamin. 2014. Defects of defective intervention. *Linguistic Inquiry* 45. 707–719. DOI: https://doi.org/10.1162/LING_a_00171

Carstens, Vicki. 2000. Concord in minimalist theory. *Linguistic Inquiry* 31. 319–355. DOI: https://doi.org/10.1162/002438900554370

Carstens, Vicki. 2016. Delayed valuation: A reanalysis of "upwards" complementizer agreement and the mechanics of Case. *Syntax* 19. 1–42. DOI: https://doi.org/10.1111/synt.12116

Chung, Sandra. 2013. The syntactic relations behind agreement. In Cheng, Lisa & Corver, Norbert (eds.), *Diagnosing syntax*, 251–270. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199602490.003.0012

Cinque, Guglielmo. 1994. On the evidence for partial N-movement in the Romance DP. In Cinque, Guglielmo (ed.), *Paths towards universal grammar,* 85–110. Washington, D.C.: Georgetown University Press.

Comrie, Bernard & Stone, Gerald & Polinsky, Maria. 1996. *The Russian language in the twentieth century*. Oxford: Clarendon. DOI: https://doi.org/10.1093/oso/9780198240662.001.0001

Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville. 2003. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville. 2012. *Features*. Cambridge: Cambridge University Press.

Cunnings, Ian & Patterson, Clare & Felser, Claudia. 2015. Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language* 71. 39–56. DOI: https://doi.org/10.1016/j.jml.2013.10.001

Danon, Gabi. 2011. Agreement and DP-internal feature distribution. *Syntax* 14. 297–317. DOI: https://doi.org/10.1111/j.1467-9612.2011.00154.x

Embick, David & Noyer, Rolf. 2001. Movement operations after syntax. *Linguistic Inquiry* 32. 555–595. DOI: https://doi.org/10.1162/002438901753373005

Ferreira, Fernanda & Henderson, John. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16(4). 555–568. DOI: https://doi.org/10.1037/0278-7393.16.4.555

Franck, Julie & Lassi, Glenda & Frauenfelder, Ulrich & Rizzi, Luigi. 2006. Agreement and movement: a syntactic analysis of attraction. *Cognition* 10(1). 173–216. DOI: https://doi.org/10.1016/j.cognition.2005.10.003

Franck, Julie & Mirdamadi, Farhad & Kahnemuyipour, Arsalan. 2020. Object attraction and the role of structural hierarchy: Evidence from Persian. *Glossa: A Journal of General Linguistics* 5(1). 27. DOI: https://doi.org/10.5334/gjgl.804

Franck, Julie & Soare, Gabriela & Frauenfelder, Ulrich & Rizzi, Luigi. 2010. Object interference: The role of intermediate traces of movement. *Journal of Memory and Language* 62(2). 166–182. DOI: https://doi.org/10.1016/j.jml.2009.11.001

Franck, Julie & Vigliocco, Gabriella & Antón-Méndez, Inés & Collina, Simona & Frauenfelder, Ulrich. 2008. The interplay of syntax and form in sentence production: a crosslinguistic study of form effects on agreement. *Language and Cognitive Processes* 23(3). 329–374. DOI: https://doi.org/10.1080/01690960701467993

Friedmann, Naama & Rizzi, Luigi & Belletti, Adriana. 2017. No case for Case in locality: Case does not help interpretation when intervention blocks A-bar chains. *Glossa: A Journal of General Linguistics* 2(1). 33. DOI: https://doi.org/10.5334/gjgl.165

Grabovac, Anna. 2022. *Maximizing the concord domain: Concord as spell-out in Slavic*. London, UK: University College London dissertation.

Graschenkov, Pavel. 2018. *Grammatika prilagatel'nogo: Teorija ad'jektivnosti i atributivnosti* [The grammar of adjectives: Theory of adjectives and attributives]. Moscow: Izd. YaSK.

Hammerly, Christopher & Staub, Adrian & Dillon, Brian. 2019. The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology* 110. 70–104. DOI: https://doi.org/10.1016/j.cogpsych.2019.01.001

Haskell, Todd & MacDonald, Maryellen. 2005. Constituent structure and linear order in language production: Evidence from subject-verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(5). 891–904. DOI: https://doi.org/10.1037/0278-7393.31.5.891

Hopp, Holger & Lemmerth, Natalia. 2018. Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition* 40(1). 171–199. DOI: https://doi.org/10.1017/S0272263116000437

Ingason, Anton Karl & Sigurdsson, Einar Freyr. 2017. The interaction of adjectival structure, concord, and affixation. *Proceedings of the 47th Meeting of the North East Linguistic Society (NELS 47)*, vol. 2. 89–98.

Jackson, Carrie & Dussias, Paola & Hristova, Adelina. 2012. Using eyetracking to study the on-line processing of case-marking information among intermediate L2 learners of German. *International Review of Applied Linguistics in Language Teaching* 50(2). 101–133. DOI: https://doi.org/10.1515/iral-2012-0005

Juhasz, Barbara & Rayner, Keith. 2003. Investigating the Effects of a Set of Intercorrelated Variables on Eye Fixation Durations in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(6). 1312–1318. DOI: https://doi.org/10.1037/0278-7393.29.6.1312

Keating, Gregory. 2009. Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59(3). 503–535. DOI: https://doi.org/10.1111/j.1467-9922.2009.00516.x

Kramer, Ruth. 2009. *Definite markers, phi-features, and agreement: a morphosyntactic investigation of the Amharic DP*. Santa Cruz, CA: University of California, Santa Cruz, dissertation.

Kramer, Ruth. 2015. *The morphosyntax of gender*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199679935.001.0001

Kush, Dave & Lidz, Jeff & Phillips, Colin. 2015. Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language* 82. 18–40. DOI: https://doi.org/10.1016/j.jml.2015.02.003

Kuznetsova, Alexandra & Brockhoff, Per & Christensen, Rune. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Lago, Sol & Shalom, Diego & Sigman, Mariano & Lau, Ellen & Phillips, Colin. 2015. Agreement attraction in Spanish comprehension. *Journal of Memory and Language* 82. 133–149. DOI: https://doi.org/10.1016/j.jml.2015.02.002

Lemmerth, Natalia & Hopp, Holger. 2019. Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition* 26. 21–45. DOI: https://doi.org/10.1080/10489223.2017.1391815

Lenth, Russell. 2021. Emmeans: Estimated Marginal Means, aka Least-Squares Means. (1.8.3.) [Computer software]. https://CRAN.R–project.org/package=emmeans

Lyashevskaya, Olga & Sharov, Serge. 2009. *Častotnyj slovar' sovremennogo russkogo jazyka (na materiale Natsional'nogo Korpusa Russkogo Jazyka)* [Frequency Dictionary of Modern Russian (based on the materials of the Russian National Corpus)]. Moscow, Russia: Azbukovnik.

Lyskawa, Paulina. 2021. *Coordination without grammar-internal feature resolution*. College Park, MD: University of Maryland dissertation.

Marantz, Alec. 1991. Case and licensing. In Westphal, Germán & Ao, Benjamin & Chae, Hee-Rahk (eds.), *Proceedings of the 8ᵗʰ Eastern States Conference on Linguistics (ESCOL8),* 234–253. Ithaca, NY: CLC Publications.

Nevins, Andrew. 2011. Multiple agree with clitics: person complementarity vs. omnivorous number. *Natural Language & Linguistic Theory* 29(4). 939–971. DOI: https://doi.org/10.1007/s11049-011-9150-4

Nicklin, Christopher & Plonsky, Luke. 2020. Outliers in L2 research: A synthesis and data re-analysis from self-paced reading. *Annual Review of Applied Linguistics* 40. 25–55. DOI: https://doi.org/10.1017/S0267190520000057

Norris, Mark. 2014. *A theory of nominal concord*. Santa Cruz, CA: University of California, Santa Cruz dissertation.

Norris, Mark. 2017. Description and analyses of nominal concord (parts I–II). *Language and Linguistics Compass* 11(11). DOI: https://doi.org/10.1111/lnc3.12267

Norris, Mark. 2018. Nominal structure in a language without articles: The case of Estonian. *Glossa: a Journal of General Linguistics* 3(1). 41. DOI: https://doi.org/10.5334/gjgl.384

Preminger, Omer. 2014. *Agreement and its failures* (Linguistic Inquiry Monographs 68). Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/9780262027403.001.0001

R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radford, Andrew. 1988. *Transformational grammar*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511840425

Romanova, Natalia & Gor, Kira. 2017. Processing of gender and number agreement in Russian as a second language: The devil is in the details. *Studies in Second Language Acquisition* 39(1). 97–128. DOI: https://doi.org/10.1017/S0272263116000012

Sigurdsson, Halldór. 2009. Remarks on features. In Grohman, Kleanthes (ed.), *Explorations of phase theory: Features and arguments*, 21–52. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110213966.21

Slioussar, Natalia & Malko, Anton. 2016. Gender agreement attraction in Russian: Production and Comprehension Evidence. *Frontiers in Psychology* 7. DOI: https://doi.org/10.3389/fpsyg.2016.01651

Slioussar, Natalia & Samoilova, Maria. 2014. A database to estimate frequencies of different grammatical features and inflectional affixes in Russian nouns. *The 9th International Conference on the Mental Lexicon,* 104–105. Niagara-on-the-Lake: Brock University and McMaster University.

Slioussar, Natalia & Samoilova, Maria. 2015. Častotnosti različnyx grammatičeskix xarakeristik i okončanij u suščestvitel'nyx russkogo jazyka [Specifics of various grammatical characteristics and endings of nouns in the Russian language]. *Proceedings of the Conference 'Dialogue 20'.*

Smith, Nathaniel & Levy, Roger. 2008. Optimal processing times in reading: A formal model and empirical investigation. In Love, Bredley & McRae, Ken & Sloutsky, Vladimir (eds.), *Proceedings of the thirtieth annual conference of the Cognitive Science Society*, 595–600. Austin, TX: Cognitive Science Society.

Tanenhaus, Michael & Spivey-Knowlton, Michael & Eberhard, Kathleen & Sedivy, Julie. 1995. Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science* 268(5217). 1632–1634. DOI: https://doi.org/10.1126/science.7777863

Toosarvandani, Maziar & van Urk, Coppe. 2013. The syntax of nominal concord: What ezafe in Zazaki tells us. *Proceedings of the 43rd Meeting of North East Linguistic Society (NELS 43)*, 221–234.

Witzel, Naoko & Witzel, Jeffrey & Forster, Kenneth. 2012. Comparisons of online reading paradigms: Eyetracking, moving-window, and maze. *Journal of Psycholinguistic Research* 41(2). 105–128. DOI: https://doi.org/10.1007/s10936-011-9179-x