Journal of Experimental Psychology: Learning, Memory, and Cognition

Rational Language Comprehension Depends on Priors about Both Meaning and Structure

Moshe Poliak, Saima Malik-Moraleda, and Edward Gibson

Online First Publication, March 31, 2025. https://dx.doi.org/10.1037/xlm0001470

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393
https://doi.org/10.1037/xlm0001470

Rational Language Comprehension Depends on Priors about Both Meaning and Structure

Moshe Poliak[1], Saima Malik-Moraleda[1, 2, 3], and Edward Gibson[1]
[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[2] McGovern Institute for Brain Research, Massachusetts Institute of Technology
[3] The Program in Speech and Hearing Bioscience and Technology, Harvard University

Language comprehension relies on integrating the perceived utterance with prior expectations. Previous investigations of expectations about sentence structure (the structural prior) have found that comprehenders often interpret rare constructions nonliterally. However, this work has mostly relied on analytic languages like English, where word order is the main way to indicate syntactic relations in the sentence. This raises the possibility that the structural prior over word order is not a universal part of the sentence processing toolkit, but rather a tool acquired only by speakers of languages where word order has special importance as the main source of syntactic information in the sentence. Moving away from English to make conclusions about more general cognitive strategies (Blasi et al., 2022), we investigate whether the structural prior over word order is a part of language processing more universally using Hindi and Russian, synthetic languages with flexible word order. We conducted two studies in Hindi (*N*s = 50, 57, the latter preregistered) and three studies with the same materials, translated, in Russian (*N*s = 50, 100, 100, all preregistered), manipulating plausibility and structural frequency. Structural frequency was manipulated by comparing simple clauses with the canonical word order (subject–object–verb in Hindi, subject–verb–object in Russian) to ones with a noncanonical (low frequency) word order (object–subject–verb in Hindi, object–verb–subject in Russian). We found that noncanonical sentences were interpreted nonliterally more often than canonical sentences, even though we used flexible-word-order languages. We conclude that the structural prior over word order is always evaluated in language processing, regardless of language type.

Keywords: psycholinguistics, sentence processing, noisy channel processing, structural prior

The utterances that speakers produce may be corrupted by noise from production, environment, or perception. For example, disfluencies in speech occur in approximately 6% of spoken words (Bortfeld et al., 2001; Fox Tree, 1995; Kasl & Mahl, 1965). Environmental noise, in the case of speech, comes from other sounds that are perceived at the same time as the utterance, which can mask the speech and impede comprehension (Sohoglu et al., 2014; Sumby & Pollack, 1954). Comprehender-level noise can come from lapses of attention or memory limitations. For example, Gibson and Thomas (1999) showed that readers may not notice a major syntactic violation (a missing verb phrase) when faced with a complicated sentence with multiple embeddings. Given the variety and omnipresence of noise, the utterance we perceive may not represent what the speaker intended to convey; to deal with this noisy input, the comprehender relies on expectations regarding meaning and structure to recover the speaker's intention (according to noisy channel processing, reviewed below). In this study, we ask whether expectations regarding structure are used across different types of languages or only in languages where word order has special importance as the main source of syntactic information.

The noisy channel processing framework proposes that we overcome noise in language by merging the perceived utterance with prior knowledge (Gibson et al., 2013; Levy, 2008; Shannon, 1949). Formally, the comprehender tries to find the most probable intended sentence, S_i, given the perceived sentence, S_p. Using Bayes' rule, this value is proportional to the likelihood (the probability of S_p given S_i) times the prior probability of S_i (Equation 1). In a more intuitive sense, the goal of the comprehender is to figure out what the speaker intended to say (S_i) from what they perceived (S_p). To infer the speaker's intended message, the comprehender considers several potential intended sentences and chooses the one that is the most probable. The most probable intended sentence is the one that balances being similar to the perceived utterance (high likelihood) and being likely to be intended a priori (high prior probability). For example, consider a situation where one hears a company representative say "We are going bankrupt for all

Tessa Warren served as action editor.
Moshe Poliak https://orcid.org/0000-0002-5000-6668
The corpora and the analysis can be found on the Open Science Framework at https://osf.io/qyjk6/ (Poliak et al., 2025). This work was supported by the National Science Foundation Award BCS-2121074 (awarded to Edward Gibson).

Moshe Poliak played a lead role in formal analysis, methodology, visualization, and writing–original draft and an equal role in conceptualization and writing–review and editing. Saima Malik-Moraleda played a supporting role in supervision and writing–review and editing and an equal role in investigation. Edward Gibson played a lead role in funding acquisition and supervision and an equal role in conceptualization and writing–review and editing.

Correspondence concerning this article should be addressed to Moshe Poliak, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Building 46, Cambridge, MA 02139, United States. Email: moshepol@mit.edu

intensive purposes." The comprehender notices that "for all intensive purposes" sounds very similar to "for all intents and purposes," but the latter is a more frequent phrase, and it makes more sense in the context, and so it is more likely to be intended a priori. Therefore, the comprehender might infer that the intended sentence was "We are going bankrupt for all *intents and purposes*."

The probability of the intended sentence given the perceived sentence is proportional to the likelihood times the prior (Equation 1).

$$P(S_i|S_p) \propto P(S_p|S_i) \times P(S_i). \quad (1)$$

While the current work is positioned within the Noisy Channel framework, there are other frameworks that try to explain how sentences are interpreted in ways that are not licensed by the grammar. Broadly, these can be seen as different proposals for how rational language comprehension may be implemented. For example, the Good Enough language processing framework proposes that nonliteral interpretations[1] result from a failure to revise an initial, shallow, and inaccurate understanding of the utterance (Christianson et al., 2001; Ferreira, 2003). Other theories that would speak to similar phenomena are memory-retrieval interference (Jäger et al., 2017; Lewis et al., 2006; Van Dyke & McElree, 2006), the competition model (Bates & MacWhinney, 1982, 1989; MacWhinney, 2022; MacWhinney et al., 1984), and the self-organizing model (Kamide & Kukona, 2018; Kukona et al., 2014; Tabor & Hutchins, 2004). Whereas we believe that the field is in need of thorough theoretical work contrasting the approaches above, we will not attempt to do so in this current project: it is out of its scope. Among the approaches that attempt to explain nonliteral comprehension, we use the Noisy Channel framework because it makes concrete predictions regarding the rate of nonliteral interpretations of sentences in relation to their prior probability.

The Meaning Prior

When faced with an implausible sentence, the comprehender often interprets the sentence nonliterally. In an investigation of noisy channel processing, Gibson et al. (2013) presented participants with implausible sentences followed by a yes–no comprehension question. For example, participants might read the implausible sentence "The girl tossed the apple the boy," followed by the question "Did the apple receive something/someone?" Here, a "yes" response would indicate a literal interpretation, while a "no" response would indicate a nonliteral interpretation. Gibson et al. (2013) found that, depending on the condition, as few as 47% of participants responded yes to the question above (i.e., a 47% rate of literal interpretations). The authors propose that the reason for this low rate of literal interpretations is that participants tried to infer the intention of the writer, engaging in noisy channel processing. When encountering such an implausible sentence as the above, participants may consider similar, more plausible sentences like "The girl tossed the apple *to* the boy." If the latter was intended, then the only thing that needed to happen to result in the implausible sentence above was simply the deletion of the preposition "to."

The prior probability of a sentence depends on the context within which it is presented, and, consequently, so does the interpretation of the sentence. Chen et al. (2023) conducted a similar experiment to Gibson et al. (2013), but they manipulated the context in which the sentences appeared, such that the context could either support the plausible meaning or not. For example, a critical implausible sentence like "The girl tossed the apple the boy" could be preceded by either the supportive context "The boy and the girl went apple picking together. The girl picked an apple that the boy wanted" or by the nonsupportive context "The aunt told the nephew she would miss him while he was on vacation. The magician pulled his hat out of the trunk." Their results showed that implausible sentences that appeared in a supportive context were more likely to be interpreted nonliterally (plausibly). Taken together, these results show that language comprehension depends on prior expectations regarding the intention of the speaker.

The Structural Prior

While the meaning prior is concerned with the probability of the semantic content of the utterance, the structural prior is concerned with the *form* of the utterance. Similar meanings can be conveyed using different constructions; these constructions vary in frequency, and more frequent constructions have a higher structural prior probability. For example, a frequent English construction is the transitive construction NP_{agent} $V_{transitive}$ $NP_{patient}$ (where NP stands for noun phrase, and V stands for verb), where an agent performs an action on a patient (Goldberg, 2010). A sample sentence that uses this construction is "The boy threw the trash." Exactly the same event (though with a different focus) can be conveyed using a cleft like "It was the boy who threw the trash." Yet another way to report the event would be using the topicalizing construction, where one of the parts of a typical declarative transitive appears at the front of the sentence, as in "The trash, the boy threw." The topicalizing construction is rare in English as it is used in narrow discourse circumstances, where the content of the clause is old information, and the topicalized element contrasts with what the speaker thinks the listener knows (Birner & Ward, 1998). Following the tradition in the literature, we will call more frequent constructions *canonical* and less frequent constructions *noncanonical*. The present study manipulated word order, which is part of the structural prior. However, it is important to note that the structural prior could consist not only of word order but also of construction frequency more generally: "The boy threw the trash" has a different structure from "It was the boy who threw the trash," even though both follow the subject–verb–object (SVO) word order. Thus, canonical constructions have a higher prior probability than noncanonical constructions.[2]

Generally, noncanonical sentences result in more processing difficulty than canonical sentences, but, in the right context, this difficulty may be mitigated. Less predictable structures are more surprising, and processing difficulty increases linearly with surprise (Shain et al., 2024; Smith & Levy, 2013). However, it has been shown across languages that, in appropriate contexts, where a noncanonical

[1] Throughout this study, we call any interpretation of a sentence that is not licensed by the grammar a "nonliteral interpretation." Other approaches (e.g., Ferreira, 2003) would call those "misinterpretations," and yet others may restrict the meaning of "nonliteral interpretations" to metaphorical interpretations (e.g., "Occam's razor" nonliterally means "Occam's principle of parsimony"). In the face of alternative usages, we still choose to use "nonliteral interpretation," reflecting the ignorance of the comprehender regarding the intended (correct) message.

[2] Historically, some theoreticians argued that noncanonical constructions (less frequent ones) are derived from a canonical counterpart. We will not explore these alternatives here.

construction is not as unexpected, processing difficulty decreases (German: Hörnig et al., 2005; Hindi: Vasishth et al., 2012; Japanese: Yano & Koizumi, 2018; Russian: Slioussar & Harchevnik, 2024). It is currently unclear whether there exists a principle that explains when a noncanonical construction is more predictable in context and what that principle might be. One proposal, however, is that a noncanonical construction may be selected to accommodate a preference for positioning old information before new information in the sentence (Clark & Clark, 1977; cf. Clifton & Frazier, 2004). Whatever properties of the context make noncanonical constructions more probable, in experimental settings, participants are usually not provided context for the sentences they see. Thus, in experiments, noncanonical structures have a lower structural prior probability than in natural settings.

Although the structural prior has been studied in the past, it is currently unclear whether it is a core part of language processing or a language-specific tool used by speakers of analytic languages, where word order has special importance as the main source of syntactic information in the sentence. Because of the special importance of word order in analytic languages, speakers of these languages may be hyperattuned to any deviation from the canonical word order, which is then reflected in how they interpret sentences. Thus, we ask whether sensitivity to the structural frequency of the utterance is a part of language processing more generally. Specifically, we ask whether it is the case that, across different types of languages, comprehenders are sensitive to the probabilities associated with different word orders. Previous work has been done in English (Ferreira, 2003; Poppels & Levy, 2016) and Mandarin (Liu et al., 2020), analytic languages with a strict SVO word order; in Hebrew (Keshev & Meltzer-Asscher, 2021), which has a relatively strict SVO word order; and in a special case of Russian (Poliak et al., 2024), where the nouns in the sentence are unmarked for case; such materials are not representative of typical language processing in Russian. In the present study, we use simple clauses in two synthetic languages with flexible word order, Hindi (Mahajan, 1988; Vasishth, 2004) and Russian (Berdicevskis & Piperski, 2020; Thompson, 1977), which allows us to test the potential influence of the structural prior by alternating two relatively frequent word orders with a minimal manipulation of morphology. If the structural prior is a core part of language processing across language types, we expect to see a main effect of canonicity such that canonical sentences are interpreted literally more often than noncanonical sentences, in both Hindi and Russian. In parallel to the structural prior, we also manipulate plausibility (the meaning prior) as a manipulation check because we strongly predict that plausible sentences are interpreted literally more frequently than implausible sentences. In addition to investigating the ubiquity of the structural prior in language processing, this study adds to the growing body of cross-linguistic evidence for noisy channel processing, testing the framework in Hindi for the first time.

The structural prior has been studied in English, where word order is a critical cue for syntactic information. While nonliteral interpretation of sentences is not a novel finding, Ferreira (2003) was one of the first to study this experimentally and in relation to sentence structure, by presenting sentences auditorily to participants and asking them to identify the agent/patient of the sentence (also see Futrell et al., 2020; Levy, 2011; Levy et al., 2009). The experimental materials varied in several ways, including canonicity, such that some sentences used active voice (canonical: e.g., "the dog bit the man") and others used passive voice (noncanonical: e.g., "the man was bitten by the dog"). Ferreira (2003) found that participants are more likely to misidentify the agent/patient in passive sentences than in active sentences. Similarly, Poppels and Levy (2016) varied plausibility and canonicity, albeit in sentences with two prepositional objects (e.g., "The box fell from the table to the floor"). A noncanonical sentence (low structural prior) would exchange the order of the prepositional phrases, such that the source would appear after the target (e.g., "The box fell *to* the floor *from* the table"). This is a noncanonical sentence because source-after-target is a less frequent construction than target-after-source. Poppels and Levy (2016) found a main effect of plausibility and a small effect of canonicity, such that sentences with the target-after-source structure were interpreted literally more often than sentences with the source-after-target structure. Both Ferreira (2003) and Poppels and Levy (2016) show that, when presented with noncanonical sentences, English speakers are more likely to interpret them nonliterally as compared to similar sentences with a canonical construction. Within the noisy channel processing framework, we take this line of evidence to suggest that English speakers use structural frequency to evaluate the prior probability of sentences as they interpret them. The lower the structural frequency, the lower the prior probability, and the more likely the sentence is to be interpreted nonliterally.

Investigating both English and Mandarin, Liu et al. (2020) manipulated plausibility (the meaning prior; e.g., "the boy threw the trash" vs. "the trash threw the boy") as well as canonicity (the structural prior; e.g., "the boy threw the trash" vs. "The trash, the boy threw"). The authors found main effects and no interaction for plausibility and canonicity, such that plausible and canonical sentences were more likely to be interpreted literally than implausible and noncanonical sentences, respectively. Like Poppels and Levy (2016) and Ferreira (2003), Liu et al. (2020) found evidence that speakers of analytic languages—Mandarin and English, where syntactic information is primarily conveyed using word order—interpret sentences according to both their plausibility and structural frequency.

Additional evidence for the role of the structural prior in sentence processing and production comes from a Hebrew study by Keshev and Meltzer-Asscher (2021). In a production task, participants were asked to complete preambles that started a sentence and a relative clause. For example, one prompt was "We liked the pupil that despite the concerns found …." They manipulated whether or not the verb (e.g., *found*) agreed with the noun that was modified by the relative clause (e.g., *pupil*). When *found* agreed with *pupil*, the relative clause could be completed as a subject-extracted relative clause, with the canonical Modern Hebrew SVO word order. However, when *found* did not agree with *pupil*, the only grammatical way to complete the sentence would be to see the relative clause as an object-extracted relative clause with an object–verb–subject (OVS) word order, which is noncanonical in Hebrew. Keshev and Meltzer-Asscher (2021) found that native speakers of Hebrew would often complete the preambles as subject-extracted relative clauses with SVO word order, even when this was grammatically impossible given the verb agreement in the preamble. They interpreted their findings as evidence that sometimes canonicity is even more important to how natural a sentence is than grammaticality. This falls in line with the findings in English and Mandarin discussed above: In languages with a relatively strict SVO word order, speakers find other word orders unnatural and may sometimes complete sentences or

interpret them in ways that are not licensed by the grammar to arrive at the canonical SVO word order.

Unlike previous studies, Poliak et al. (2024) studied Russian, a synthetic language with a flexible word order, which we also investigate in the present study. We will summarize their experiment and then explain why it cannot be used to infer that the speakers of synthetic languages are sensitive to the structural prior. In Poliak et al. (2024), participants were presented with sentences in Russian in either the canonical SVO word order or a noncanonical OVS word order. Both of these word orders are grammatical because Russian has a flexible word order in the sense that any ordering of subject, verb, and object is grammatical. Sentences with SVO word order differed from sentences with OVS word order only in verb agreement. For example, the sentence "Rachel lifted Joe" was considered SVO if "lifted" had a feminine conjugation and OVS if "lifted" had a masculine conjugation. After each sentence, participants were asked a yes–no question to assess their interpretation of the sentence (e.g., "Did Joe lift someone?"). The study revealed that SVO sentences were almost always interpreted literally, whereas OVS sentences were often interpreted nonliterally. In other words, even though OVS sentences are grammatical and not infrequent in Russian when asked about the subject in OVS sentences, participants often replied that the first noun is the subject, interpreting the sentence nonliterally, like an SVO sentence.

However, the study cannot be used to infer whether speakers of synthetic languages such as Russian use the structural prior in sentence comprehension. Although OVS sentences are common in Russian, Poliak et al. (2024) created sentences using names that cannot be conjugated for grammatical case. In Russian, there are several cues for subjecthood: subject–verb agreement, word order (SVO), case markings, and context. A known phenomenon is *word order freezing*: When a sentence lacks case marking, the canonical word order (SVO) becomes more frequent and all other word orders become infrequent (Berdicevskis & Piperski, 2020; Bouma, 2011; Jakobson, 1971). However, in Poliak et al. (2024), participants were presented with sentences that lacked both grammatical case and context, thus removing two out of four cues for subjecthood. As a result, participants often preferred to interpret sentences nonliterally, ignoring subject–verb agreement. In other words, contextless, caseless OVS sentences are less acceptable than cased OVS sentences and are not representative of OVS sentences more generally (Bloom, 1999; Jakobson, 1971; Mahowald, 2011). Moreover, we conducted an acceptability study (see Appendix A), showing that uncased OVS sentences are less acceptable than cased OVS sentences, whereas the acceptability of SVO sentences (both grammatical and ungrammatical) is not affected by the presence or absence of case marking. Poliak et al. (2024) showed that the structural prior plays a role in language processing in a synthetic language but they did so under highly unnatural conditions. Thus, participants may have interpreted sentences nonliterally not because their structure was infrequent but because the sentences, which were morphosyntactically ambiguous and lacked context, were unnatural in Russian.

The Present Study

In the current investigation, we ask whether the structural prior is used across languages, regardless of whether their word order is strict or not. We operationalize the structural prior as word order specifically, even though the structural prior is a broader term, encompassing the frequency of all constructions in the language, not just word order alternations. Previous work that has studied the structural prior either relied on languages with a relatively strict SVO word order (Ferreira, 2003; Keshev & Meltzer-Asscher, 2021; Liu et al., 2020; Poppels & Levy, 2016) or did so using a construction that borders on being ungrammatical (Poliak et al., 2024). These past results add to a broader pattern, where English-centric research has been limiting the generalizability of cognitive science (Blasi et al., 2022). To pursue the question of the universality of the structural prior in language processing, we move away from English and manipulate canonicity and plausibility in Hindi and Russian, where syntactic relations in the sentence are conveyed with case markings, while word order is allowed to vary. If we find that canonical word orders (SOV in Hindi, SVO in Russian) are interpreted literally more often than noncanonical word orders (OSV in Hindi, OVS in Russian), we will infer that the structural prior over word orders is used in language comprehension regardless of whether the word order of the language is strict or flexible. Additionally, if there is any effect of prior such that noncanonical or implausible sentences are interpreted nonliterally more often than canonical or plausible sentences, we will conclude that we have detected evidence for noisy channel processing in Hindi, an SOV language that is understudied in the psycholinguistics literature.

Hindi is an Indo-European language, which is the most spoken language in India (378,000,000 speakers; Eberhard et al., 2024). As mentioned above, it is an SOV word-order language with a relatively flexible word order. Hindi has a split ergativity system, which is important to the present study because, in the perfective aspect, the agent (subject) of the sentence is marked for the ergative case and, optionally, the patient (object) is marked for the accusative case. Case marking is expressed using monosyllabic words (free morphemes) immediately after the modified noun. Usually, the verb agrees with the object, but if it is marked explicitly for accusative case, then the verb remains in its base form and is not inflected for agreement with the object. Therefore, to keep all verb forms the same, we always use sentences in the past tense with explicit case markings. Russian is also an Indo-European language, and it is the official language of Russia. Russian is an SVO word-order language with a highly flexible word order. Unlike in Hindi, Russian case marking is expressed with bound morphemes that interact with the stem of the word (e.g., dog in nominative case is пёс [pʲos], but in accusative and genitive case, it is пса [psa]). Generally, Russian has subject–verb agreement in gender and number, but in imperfective aspect (roughly equivalent to present tense in English), the verb is not marked for gender. Therefore, to keep verb forms constant across Russian stimuli, we used the imperfective aspect for all the sentences.

We conducted five experiments: Experiment 1 (Hindi); Experiment 2 (Hindi Replication; preregistered); Experiment 3 (Russian; preregistered); Experiment 4 (Russian Replication; preregistered); and Experiment 5 (Russian with noise in filler items; preregistered). In each experiment, we manipulated the structural prior (canonicity) and the meaning prior (plausibility). To try to match the meaning priors in all studies, we initially wrote the Hindi materials and then translated them into Russian, trying to change as little as possible while keeping sentences grammatical. In all plausible sentences, an animate agent performs an action on an inanimate patient (e.g., "the boy opened the box"),

whereas in all implausible sentences, an inanimate agent performs an action on an animate patient (e.g., "the box opened the boy"). The canonical sentences used the most frequent word order in the language (SOV for Hindi and SVO for Russian), while noncanonical sentences used an alternative word order (OSV for Hindi and OVS for Russian). To develop an understanding of the frequency of the different word orders in Hindi and Russian, we conducted a corpus analysis (see below), which verified the intuitions described above.

Corpus Analysis

We investigated the frequency of different word orders in Hindi and Russian by conducting a corpus analysis. The corpora and the analysis can be found on the Open Science Framework (OSF) at https://osf.io/qyjk6/ (Poliak et al., 2025).

Method

We used universal dependency corpora from Deep Universal Dependencies 2.8 (Zeman et al., 2023) and analyzed them using the package udpipe in R (Wijffels, 2023). For Hindi, we used the corpus Hindi Dependency Treebank (Bhat et al., 2017; Palmer et al., 2009), concatenating the development, training, and testing data, resulting in 16,649 sentences (the authors do not specify the sources of the sentences, but Palmer et al., 2009 stated that they have a goal of annotating 400,000 words from "newswire": Therefore, we suspect that the source of the corpus includes news articles). For Russian, we used the SynTagRus corpus (Droganova et al., 2018), concatenating the development, training, and testing data, resulting in 87,336 sentences from a variety of genres (e.g., news, fiction, science). For both languages, we extracted all clauses that had a main verb (avoiding copulas in Hindi), an explicit subject, and an explicit object. Hindi and Russian are both pronoun-dropping languages (Bizzarri, 2015; Butt, 2001), which means that subject and object pronouns are often not stated explicitly in the sentence. Therefore, in Hindi, out of 15,477 verbal clauses, 3,459 clauses had explicit subject, verb, and object (22.3% of verbal clauses), while in Russian out of 79,763 verbal clauses, 13,971 clauses had explicit subject, verb, and object (17.5% of verbal clauses).

Results and Discussion

In Hindi, out of 3,459 verbal clauses with overt subjects and objects, 2,656 had SOV word order and 198 had OSV word order (76.8% and 5.7%, respectively). SOV was the most frequent word order (76.8%), followed by SVO (17.3%), and followed by OSV (5.7%). In Russian, out of the 13,971 clauses with overt subjects and objects, 10,858 sentences had SVO word order and 1,171 sentences had OVS word order (77.7% and 8.4%, respectively). OVS was the second most frequent word order after SVO, yet SVO was markedly more frequent than OVS. Overall, the short corpus analysis shows that, indeed, SOV and SVO are the canonical (most frequent) word orders for Hindi and Russian and that OSV and OVS are noncanonical. It is worth noting that OSV is the third most frequent word order in Hindi, while OVS is the second most frequent word order in Russian, and OSV is relatively rarer in Hindi than OVS is in Russian (5.7% and 8.4%, respectively). Overall, the frequency patterns for the canonical and noncanonical word orders seem comparable and useful for the present study. However, we do not know whether the difference in the relative frequency between OVS and OSV in Russian and Hindi, respectively, is reliable, and, if yes, whether it meaningfully affects sentence processing. Moreover, we also do not know whether the ranking of the frequency of OVS (the third most frequent in Russian) and OSV (the second most common in Hindi) is meaningful for sentence processing. See further discussion of these differences given the present study in the Discussion section.

Experiment 1 (Hindi)

Data and analysis for all of the experiments can be found on the OSF at https://osf.io/qyjk6/.

Method

Materials

Each participant received a questionnaire entirely in Hindi with 60 items (20 critical, 40 fillers). In a 2 × 2 design, all critical items involved an animate noun, an inanimate noun, and a transitive verb in the past tense. Each item was presented only once to each participant, such that the item could be canonical (SOV) or noncanonical (OSV) and plausible or implausible (Table 1). Each participant saw five items in each condition, and we randomized on a participant level, which items were presented in which condition. We used two types of filler items, all of which were plausible. One half of the filler items (20) were identical to the construction of the critical stimuli but allowed more flexibility in word selection, such that the nouns and verbs could consist of several words (like "news anchor" or compound verbs). Half of these fillers were presented in canonical word order, and half were presented in noncanonical word order, to avoid a confound wherein all noncanonical sentences are critical items. The rest of the filler items (20) were canonical and plausible sentences that varied more widely in structure (e.g., "The pilot flew the plane despite the storm"). Each item was followed by a "yes/no" question, like "Did the boy open something/someone?"[3] The questions differed between items such that, for one-half of the items, replying "yes" would indicate a literal interpretation regardless of condition (e.g., Table 1), while for the rest of the items replying "no" would indicate a literal interpretation.

Participants

We collected 57 participants on Prolific who identified their first language to be Hindi, of which seven were excluded because they interpreted filler sentences nonliterally more than 20% of the time.

[3] Note that the phrasing of the question ("something/*someone*") suggests that it is possible that the object of the sentence is animate, which potentially increases the prior probability of implausible utterances. We chose this phrasing because past research that employed a similar paradigm has used the same phrasing (e.g., Gibson et al., 2013). Moreover, previous work has shown that implausible comprehension questions increase the rates of nonliteral interpretations (Bian et al., 2020). Critically, however, there is no evidence or theory that the construction of the question interacts with the experimental manipulation (which in our case concerns the plausibility of the critical sentence and, importantly, its word order).

Table 1
A Critical Item From Experiments 1 and 2 (Hindi, Hindi Replication)

Plausibility	Canonicity	Sample item
Plausible	Canonical	लड़के ने डब्बा को खोला ladke ne dabba ko khola Boy ERGATIVE box accusative opened The boy opened the box
Plausible	Noncanonical	डब्बा को लड़के ने खोला dabba ko ladke ne khola Box accusative boy ERGATIVE opened The boy opened the box
Implausible	Canonical	डब्बा ने लड़के को खोला dabba ne ladke ko khola Box ERGATIVE boy accusative opened The box opened the boy
Implausible	Noncanonical	लड़के को डब्बा ने खोला ladke ko dabba ne khola Boy accusative box ERGATIVE opened The box opened the boy
Question for plausible sentences		क्या लड़के ने कुछ या किसी को खोला? kya ladke ne kuchh ya kisee ko khola? INT boy ERGATIVE something or someone opened? Did the boy open something or someone?
Question for implausible sentences		क्या डब्बा ने कुछ या किसी को खोला? kya dabba ne kuchh ya kisee ko khola? INT box ERGATIVE something or someone opened? Did the box open something or someone?

Note. INT = Interrogative.

Procedure

Participants were asked to provide their Prolific ID and to report whether they are native speakers of Hindi, followed by written instructions to read each sentence and respond to its corresponding yes/no comprehension questions. All questions were presented on a single page at the same time, and participants scrolled through the page responding to them. No additional information was collected (e.g., reaction times, eye movements, mouse cursor patterns). Participants were allowed to change their responses as much as they wanted and to take as long as they desired, and once they clicked on the final "submit" button, they were redirected back to Prolific and compensated with $3.30 (a rate of $15 per hour).

Results

We wrangled and visualized the data in R (R Core Team, 2024) using the tidyverse (Wickham et al., 2024) and ggthemes (Arnold, 2021), fit our model using Bayesian regression with the brms package (Bürkner, 2017, 2018), and extracted posterior distributions from the models using the bayestestR and tidybayes package (Kay, 2023; Makowski et al., 2019). Descriptive results are summarized in Table 2 and visualized in Figure 1 alongside the inferred 95% credible intervals (CrIs). For our statistical analysis, we set literal interpretation (literal = 1, nonliteral = 0) as the dependent variable and plausibility (coded as implausible = −0.5, plausible = 0.5) and canonicity (coded as noncanonical = −0.5, canonical = 0.5) as predictors (main effects andinteraction). We specified the full random effects structure justified by the design, with random intercepts for participants and items, random slopes for the main effects, and the interaction effect of plausibility and canonicity as slopes within participants and items. This resulted in the formula in Equation 2.

$$\begin{aligned}\text{Literal Interpretation} \sim\ & \text{Plausibility} \times \text{Canonicity} \\ & + (\text{Plausibility} \times \text{Canonicity}|\text{Participant}) \\ & + (\text{Plausibility} \times \text{Canonicity}|\text{Item})\end{aligned} \quad . \quad (2)$$

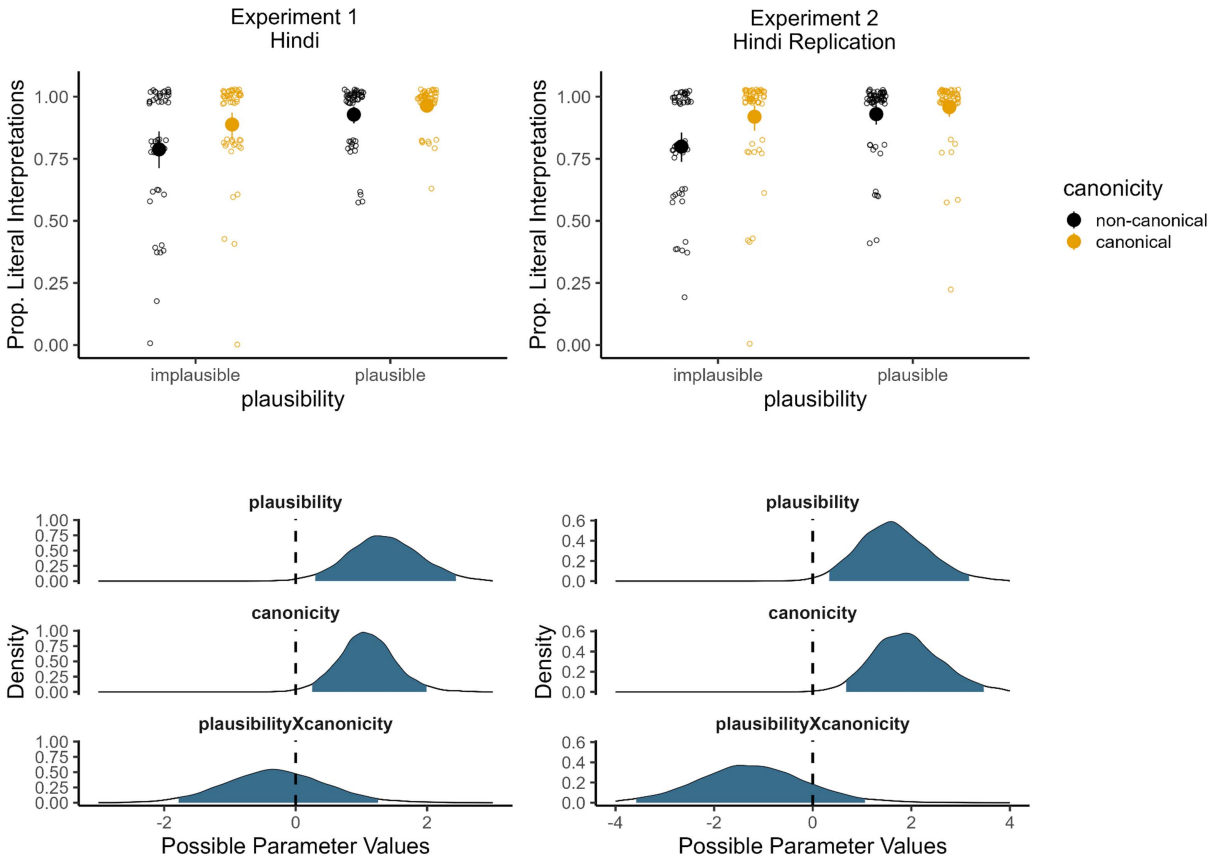
The model used the Bernoulli distribution family with the logit link function, 4,000 iterations (2,000 of which were burn-in transitions), four chains, and the default flat priors for fixed effects from the brms package. The model resulted in no divergent transitions and *R*-hat values of 1.00 for all fixed effects parameters. Visual inspection showed that the chains mixed well and that in a posterior predictive check, the simulated data closely followed the observed data. The posterior mean for the intercept was 3.35 (CrI [2.67, 4.16]). Plausible sentences were interpreted literally more often than implausible sentences (M = 1.35, CrI [0.29, 2.44]), and canonical sentences were interpreted literally more often than noncanonical sentences (M = 1.09, CrI [0.25, 2.00]). There seems to be no evidence for aninteractionbetween canonicity and plausibility (M = −0.30, CrI [−1.79, 1.26]).

Table 2
Results for Experiments 1 (Hindi) and 2 (Hindi Replication)

		Literal interpretation	
Plausibility	Canonicity	Experiment 1	Experiment 2
Plausible	Canonical	96.4%	95.8%
Plausible	Noncanonical	92.8%	93.0%
Implausible	Canonical	88.8%	91.9%
Implausible	Noncanonical	78.8%	79.9%

Figure 1

Results and Inference of Experiment 1 (Hindi) and Experiment 2 (Hindi Replication)



Note. The top plots represent the proportion of literal interpretations for critical sentences, split across canonicity and plausibility. Error bars are bootstrapped 95% confidence intervals over participant means. Unfilled circles represent participant means. The bottom plots represent the posterior distribution from the models over plausibility, canonicity, and their interaction. Shaded regions are 95% equal-tailed credible intervals. Prop. = proportion. See the online article for the color version of this figure.

Discussion

Experiment 1 (Hindi) showed that, in Hindi, both plausibility and canonicity influence whether sentences are interpreted literally. Specifically, plausible sentences are interpreted literally more often than implausible sentences, and canonical sentences are interpreted literally more often than noncanonical sentences.

Experiment 2 (Hindi Replication)

Method

Methods were identical to Experiment 1 (Hindi), except that we made a small improvement in randomization such that the first three filler items for each participant were filler items (our experience shows that participants may exhibit unusual behavior if the first few sentences they encounter in the study are unusual).[4] This experiment was preregistered prior to data collection (https://osf.io/qyjk6/).

Participants

We preregistered an intention to sample 80 participants on Prolific who indicate that their first language to be Hindi. However, out of awareness of the limited pool of participants who speak Hindi on Prolific, we preregistered a stopping criterion by which we commit to terminate data collection after 2 weeks even if we fail to sample 80 participants by then. Eventually, 61 participants completed the study, of which four were excluded for interpreting more than 20% of filler sentences nonliterally (an exclusion criterion that we preregistered). This resulted in a final number of 57 participants. We did not prevent participants from Experiment 1 from participating in Experiment 2 because almost 5 months separated data collection for both experiments. Out of 57 participants, 35 have participated in Experiment 1 too. We compared returning participants to naive participants in an exploratory analysis (Appendix B) and found no difference between them and no change in inference.

Procedure

The procedure was identical to Experiment 1.

[4] As seen descriptively and inferentially, this small change to the experimental design did not result in an observable difference.

Results

Preregistered Analyses

The analysis was identical to Experiment 1 (Hindi), with the only difference being that we preregistered a model with 4,000 iterations (out of which 2,000 were burn-in transitions). The model resulted in no divergent transitions, *R*-hat values of 1.00, and, on visual inspection, the chains were well-mixed. Descriptive results are summarized in Table 2 and visualized in Figure 1 alongside the inferred 95% CrIs. The posterior mean for the intercept was 4.08 (CrI [3.25, 5.14]). Plausible sentences were interpreted literally more often than implausible sentences ($M = 1.64$, CrI [0.33, 3.18]), and canonical sentences were interpreted literally more often than noncanonical sentences ($M = 1.93$, CrI [0.67, 3.48]). There seems to be no evidence for aninteractionbetween canonicity and plausibility ($M = -1.27$, CrI [−3.58, 1.07]).

Discussion

Similar to the results of Experiment 1 (Hindi), Experiment 2 (Hindi Replication) showed two main effects in the predicted direction of plausibility and canonicity with nointeractioneffect. In other words, this is additional evidence that, in Hindi, plausible sentences are more likely to be interpreted literally than implausible sentences, and canonical sentences are more likely to be interpreted literally than noncanonical sentences.

Experiment 3 (Russian)

To investigate whether the results of Experiments 1 and 2 (Hindi, Hindi Replication) may generalize to another language, we extended them to Russian, trying to make as few changes as possible, resulting in Experiments 3 and 4 (Russian, Russian Replication). Experiment 3 (Russian) was preregistered on the OSF prior to conducting the study (https://osf.io/qyjk6/). We predicted that Experiment 3 will show the same pattern of results as Experiment 1: that plausible sentences will be interpreted more often than implausible sentences, that canonical sentences will be interpreted literally more often than noncanonical sentences, and that there will be no interaction between plausibility and canonicity.

Method

Materials

The materials from Experiments 1 and 2 (Hindi, Hindi Replication) were translated to Russian, retaining the same 2 × 2 design with 60 items, of which 20 were critical and 40 were fillers (half of which were manipulated to be either canonical or noncanonical). The Russian stimuli differed in three ways from the Hindi stimuli. First, to avoid subject–verb agreement in gender, all the verbs were translated to be in the present tense and imperfective aspect (roughly equivalent to present progressive in English). Second, agent nouns were inflected for the nominative case (not ergative) because Russian has a nominative–accusative alignment. Third, some nouns and verbs were changed as necessary to make sure that the nouns could be inflected for nominative and accusative cases (some nouns in Russian do not inflect for accusative case, while some ergative verbs in Hindi cannot be directly translated to transitive verbs in Russian). An example of the stimuli can be found in Table 3.

Participants

We recruited 50 participants from Prolific, filtering for participants who identified as (a) being native speakers of Russian, (b) speaking Russian fluently, and (c) using Russian as their primary language in everyday life. All participants interpreted more than 80% of the filler sentences literally, and all indicated on our questionnaire that Russian is their native language (except one participant who indicated their native language to be both Russian and Ukrainian).

Procedure

The procedure was identical to Experiment 3, except we added demographic questions about the contexts in which participants have learned and are using Russians.

Results

Preregistered Analyses

The analyses for Experiment 3 (Russian) were preregistered prior to conducting the study, and they were identical to the analyses in Experiment 2 (Hindi Replication). Descriptive results are summarized in Table 4 and Figure 2. We used a step size of 0.9, and the model resulted in no divergent transitions, *R*-hat values of 1.00, and, on visual inspection, the chains were well-mixed. The posterior mean for the intercept was 5.36 (CrI [4.05, 7.17]). Plausible sentences

Table 3
A Critical Item From Experiments 3 and 4 (Russian, Russian Replication)

Plausibility	Canonicity	Sample item
Plausible	Canonical	Мальчик открывает коробку malʲchik otkryvaet korobku Boy.NOM opens box-ACC The boy opens the box
Plausible	Noncanonical	Коробку открывает мальчик korobku otkryvaet malʲchik Box-ACC opens boy.NOM The boy opens the box
Implausible	Canonical	Коробка открывает мальчика korobka otkryvaet malʲchika Box.NOM opens boy-ACC The box opens the boy
Implausible	Noncanonical	Мальчика открывает коробка malʲchika otkryvaet korobka Boy-ACC opens box.NOM The box opens the boy
Question for plausible sentences		Открывает ли мальчик что-то/кого-то? Otkryvaet li malʲchik chto-to/kogo-to? Opens INT boy.NOM something/someone? Did the boy open something or someone?
Question for implausible sentences		Открывает ли коробка что-то/кого-то? Otkryvaet li korobka chto-to/kogo-to? Opens INT box.NOM something/someone? Did the box open something or someone?

Note. NOM = Nominative; ACC = Accusative; INT = Interrogative.

Table 4
Results for Experiments 3 (Russian) and 4 (Russian Replication)

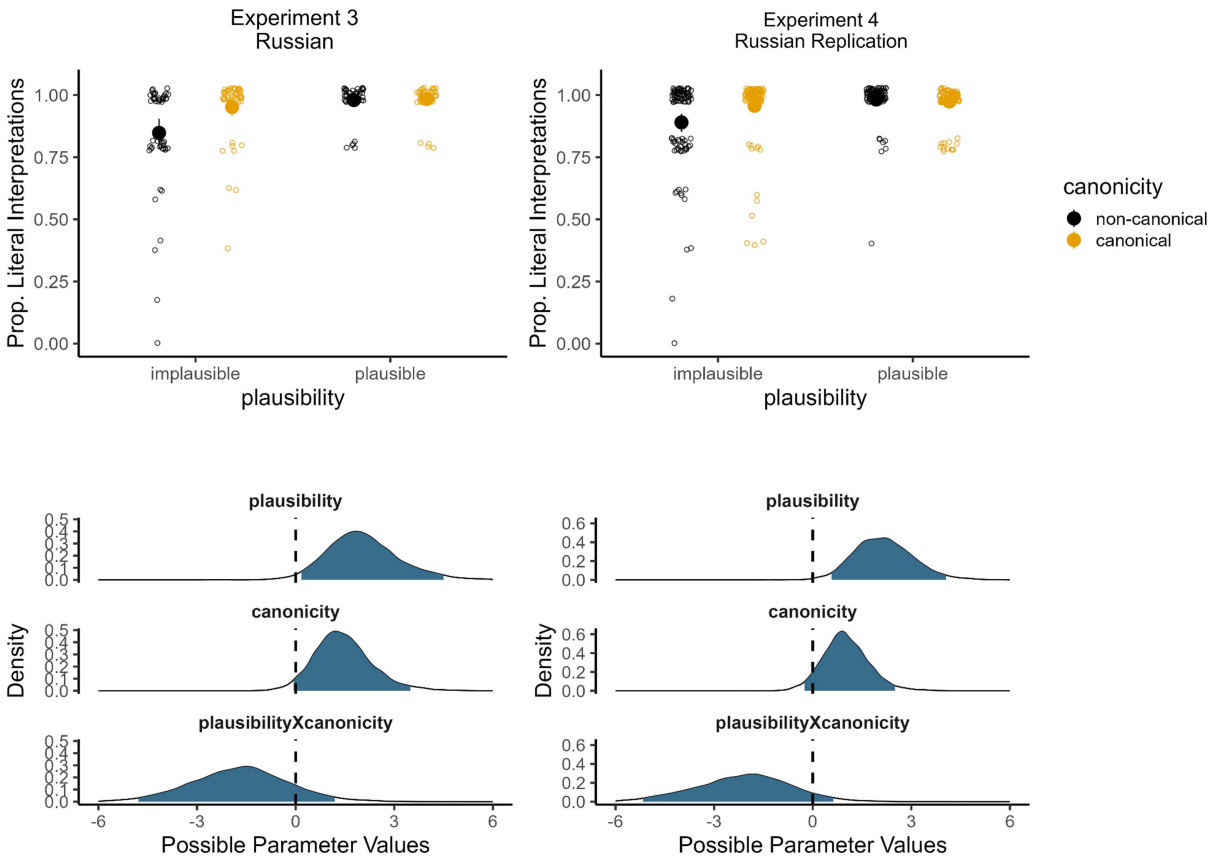
		Literal interpretation		
Plausibility	Canonicity	Experiment 3	Experiment 4	Experiment 5
Plausible	Canonical	98.4%	97.4%	97.8%
Plausible	Noncanonical	98.0%	98.2%	98.2%
Implausible	Canonical	95.2%	95.8%	90.6%
Implausible	Noncanonical	84.8%	89%	80.8%

were interpreted literally more often than implausible sentences (M = 2.06, CrI [0.16, 4.52]). Canonical sentences were not interpreted literally more often than noncanonical sentences (M = 1.46, CrI [−0.08, 3.49]). There was nointeractionbetween plausibility and canonicity (M = −1.73, CrI [−4.78, 1.20]). The posterior distributions over the fixed effects are represented in Figure 2.

Exploratory Analyses

On the surface, we did not find substantial evidence for an effect of canonicity in the preregistered analysis. Upon deeper investigation, converting the logits in the model output to probabilities, we see that the estimates of the model severely mismatch the observed means in the data: The estimates that the model produces for the four conditions (implausible + noncanonical, implausible + canonical, plausible + noncanonical, plausible + canonical) are .960, .996, .998, and .999, respectively, whereas the population means are .848, .952, .980, and .984, respectively. This behavior is consistent with logistic regression in the presence of a ceiling effect: when nearly all observations in a condition are 1 (in our case, 1 represents literal interpretation). For example, the difference between 1 and 0 on a logit scale is .23 on the probability scale; in

Figure 2
Results and Inference of Experiment 3 (Russian) and Experiment 4 (Russian Replication)



Note. The top plots represent the proportion of literal interpretations for critical sentences, split across canonicity and plausibility. Error bars are bootstrapped 95% confidence intervals over participant means. Unfilled circles represent participant means. The bottom plots represent the posterior distribution from the models over plausibility, canonicity, and their interaction. Shaded regions are 95% equal-tailed credible intervals. Prop. = proportion. See the online article for the color version of this figure.

contrast, the difference between a 5 and 4 on the logit scale (again a difference of 1) is .01 on the probability scale (an order of magnitude smaller). Therefore, if the other estimates are high on the logit scale (e.g., an intercept of 5.36 like in our data), the logistic regression may not detect sizeable effects. Additionally, visually inspecting Figure 2, it seems that there is a meaningful difference between canonical and noncanonical implausible sentences. Therefore, we opted to fit investigate the difference between canonical and noncanonical implausible sentences. We kept the model the same as above, but we changed the contrast coding to treatment coding (setting implausible, noncanonical as the reference level), which directly estimates the effect of canonicity within implausible items. This model resulted in an intercept of 2.86 (CrI [1.85, 4.16]), a substantial positive effect of plausibility within noncanonical sentences ($M = 3.90$, CrI [1.48, 8.05]), a substantial positive effect of canonicity within implausible items ($M = 3.15$, CrI [1.17, 6.40]), and no evidence for an interaction ($M = 0.06$, CrI [−4.11, 7.01]).

Discussion

Experiment 3 (Russian) sought to extend Experiments 1 and 2 (Hindi, Hindi Replication) from Hindi to Russian, investigating the effect of plausibility and canonicity of sentences on how frequently they are interpreted literally. The preregistered analysis detected only a main effect of plausibility, such that plausible sentences are interpreted literally more often than nonliteral sentences. Based on an inspection of the model and on visual inspection of the data, we saw that the sum-coded logistic regression failed to capture a simple effect of canonicity within implausible items and conducted a secondary analysis using treatment coding, setting implausible, and noncanonical sentences as baseline. The secondary analysis found two main effects, such that plausible noncanonical sentences were interpreted literally more often than implausible noncanonical sentences, and canonical implausible sentences were interpreted literally more often than noncanonical implausible sentences, with no interaction. Because our predictions were only partially realized in the preregistered model, we suspected that Experiment 3 (Russian) was underpowered and therefore doubled the number of participants in Experiment 4 (Russian Replication).

Experiment 4 (Russian Replication)

After completing Experiment 3 (Russian), we suspected that the effect of canonicity was not detected despite our prediction due to natural variation in sampling and a small sample size. We conducted a power analysis with effect sizes of 75% of the observed effect sizes and discovered that, to have 80% power to detect an effect of canonicity, we would need 100 participants. Therefore, we preregistered a replication of Experiment 3 with 100 participants (https://osf.io/qyjk6/). Experiment 4 was identical to Experiment 3 in all ways except for the following: (a) we changed the sample size (Experiment 3 $N = 50$, Experiment 4 $N = 100$); (b) we made a small improvement in randomization such that the first three filler items for each participant were filler items (our experience shows that participants may exhibit unusual behavior if the first few sentences they encounter in the study are unusual); (c) we changed inclusion criteria to increase the pool of available participants, requiring only that participants identify as native speakers of Russian (not all Prolific participants complete the screening questionnaire in its entirety, so having three filters like in Experiment 3 artificially decreased the participant pool).

Method

Materials

Materials in Experiment 4 were identical to those in Experiment 3 (see Table 3).

Participants

We recruited 100 participants from Prolific, only allowing participants to take the study if they indicated that their native language is Russian. On our questionnaire, two participants indicated that their native language is Ukrainian, and one participant indicated that their native language is Hebrew. Following our preregistered inclusion criteria (having a Russian-speaking environment in childhood and adolescence), we excluded one Ukrainian speaker and one Hebrew speaker. We left one native speaker of Ukrainian in the sample because they indicated speaking Russian with their parents, in preschool, and in school. Then, we recruited two more participants (who self-identified as native speakers of Russian) instead of the two we excluded, resulting in 100 participants.

Procedure

The procedure was identical to Experiment 3.

Results

Preregistered Analyses

The analyses for Experiment 4 were preregistered prior to conducting the study, and they were identical to those for Experiment 3. Descriptive results are summarized in Table 4 and Figure 2. The model resulted in no divergent transitions, *R*-hat values of 1.00, and the chains were well-mixed on visual inspection. The posterior mean for the intercept was 5.34 (CrI [4.34, 6.63]). Plausible sentences were interpreted literally more often than implausible sentences ($M = 2.17$, CrI [0.57, 4.06]). Canonical sentences were not interpreted literally substantially more often than noncanonical sentences ($M = 1.01$, CrI [−0.27, 2.52]). Like in Experiment 3, there is nointeractionbetween plausibility and canonicity ($M = -2.13$, CrI [−5.16, 0.63]). The posterior distributions over the main effects are represented in Figure 2.

Exploratory Analyses

Using the same reasoning as in Experiment 3 (the Exploratory Analyses section), we again fit a new model using treatment coding, setting noncanonical implausible sentences as baseline. The estimated intercept was 2.97 (CrI [2.27, 3.82]), with a substantial positive effect of plausibility within noncanonical sentences ($M = 3.65$, CrI [1.69, 7.01]), a substantial positive effect of canonicity within implausible items ($M = 2.30$, CrI [0.98, 4.18]), and no evidence for an interaction ($M = -0.01$, CrI [−4.12, 6.35]).

Discussion

Experiment 4 (Russian Replication) was a replication of Experiment 3 (Russian) with twice as many participants ($N = 100$). The results of this experiment show roughly the same results as Experiment 3, both inferentially and descriptively. Inferentially, the preregistered analyses detected only a main effect of plausibility, whereas an exploratory, treatment-coded model revealed a main effect of canonicity within implausible items.

Experiment 5: An Attempt to Lower the Rate of Literal Interpretations in Russian

In Experiments 3 and 4 (Russian), all plausible stimuli were interpreted literally almost all of the time, not displaying an effect of canonicity within plausible items. This is consistent with a ceiling effect: The prior probability for plausible items was above the threshold for making a nonliteral interpretation, such that even if the prior probability differed across canonicity, it was not evident in the binary task. The purpose of the current experiment was to uniformly increase the rate of nonliteral interpretations, such that we could see whether there is an effect of canonicity within plausible stimuli in Russian. Gibson et al. (2013) and Poliak et al. (2024) introduced grammatical errors into their filler stimuli, finding that participants who were exposed to these ungrammatical fillers were more likely to interpret sentences nonliterally across the board. They explained that this reflected more noise in the environment, prompting participants to rely more on their priors than on the observed sentences.

Method

Materials

Materials in Experiment 5 were identical to Experiment 4 except we introduced grammatical errors in agreement or case in 20 out of the 40 filler stimuli, expecting to see fewer grammatical interpretations. The experiment was preregistered on the OSF, where we also uploaded the materials, and it can be found at https://osf.io/qyjk6/.

Participants

We recruited 100 participants from Prolific, *only* allowing participants to take the study if they indicated that their native language is Russian. On our questionnaire, one participant indicated that their native language is Ukrainian and another indicated that their native language is Mongolian. Following our preregistered inclusion criteria (having a Russian-speaking environment in childhood and adolescence), we excluded the native speaker of Mongolian and recruited an additional participant, who self-identified as a native speaker of Russian, resulting in 100 participants after exclusions.

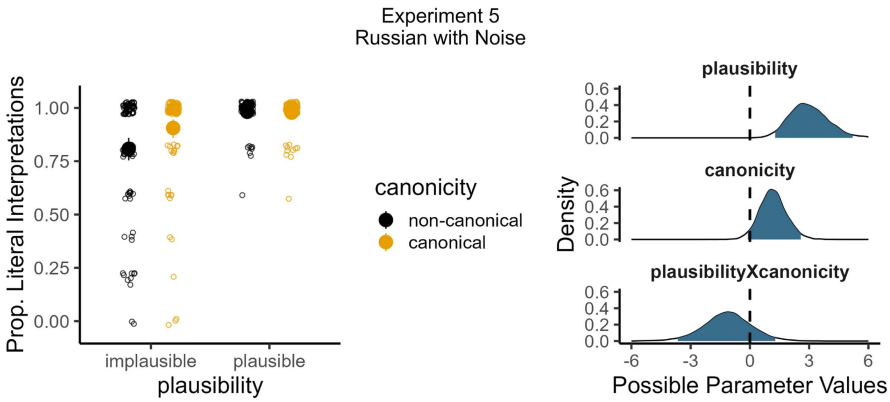
Procedure

The procedure was identical to Experiment 3.

Results

The results and model output are visualized in Figure 3. The Bayesian logistic regression converged, resulting in no divergent transitions and R-hat values of 1.00 for all fixed effects parameters. The preregistered model estimated the posterior mean for the intercept to be 5.07 (CrI [4.12, 6.36]). The model detected a main effect of plausibility ($M = 3.00$, CrI [1.26, 5.22]) but not a main effect of canonicity ($M = 1.17$, CrI [−0.12, 2.58]) nor an interaction ($M = -1.14$, CrI [−3.67, 1.30]). As for Experiments 3 and 4, we fit an exploratory, unpreregistered model, with treatment coding such that noncanonical implausible sentences were the reference level. The estimated intercept was 2.53 (CrI [1.77, 3.41]),

Figure 3
Results From Experiment 5: Russian With Noise



Note. The left plot represents the proportion of literal interpretations for critical sentences, split across canonicity and plausibility. Error bars are bootstrapped 95% confidence intervals over participant means. Unfilled circles represent participant means. The right plot represents the posterior distribution from the models over plausibility, canonicity, and their interaction. Shaded regions are 95% equal-tailed credible intervals. Prop. = proportion. See the online article for the color version of this figure.

with a substantial positive effect of plausibility within noncanonical sentences ($M = 4.57$, CrI [1.46, 2.22]), a substantial positive effect of canonicity within implausible items ($M = 1.96$, CrI [0.68, 0.79]), and no evidence for an interaction ($M = -0.25$, CrI [−3.23, 3.95]).

Discussion

Experiment 5 extended Experiments 3 and 4 by including ungrammatical sentences in the filler materials. Previous work (e.g., Gibson et al., 2013) has found that this manipulation decreases the rate of nonliteral interpretations on the critical stimuli. However, in the case of the current experiment, the results were consistent with Experiments 3 and 4. With sum coding, a substantial main effect of plausibility, but not canonicity, emerged. With treatment coding, directly comparing canonical and noncanonical implausible sentences, the model detected substantial effects of plausibility and canonicity, with no interaction.

General Discussion

In the present study, we investigated whether the structural prior over word orders is a core part of language processing or a tool that is only acquired by speakers of languages where the structure of the sentence is the main source of syntactic information, like English and Mandarin. To investigate this, we conducted two studies in Hindi (pilot and replication, the latter preregistered) and three studies in Russian (pilot and replication, both preregistered). We manipulated plausibility and canonicity, asking participants binary comprehension questions that could indicate a literal or a nonliteral interpretation of the sentence. Like in previous studies, we found that plausible sentences were interpreted literally more often than implausible sentences in all experiments. However, we show for the first time a similar role for canonicity using everyday sentences in synthetic languages: Canonical sentences were interpreted literally more often than noncanonical sentences. This pattern held as a main effect across levels of plausibility in Hindi (Experiments 1 and 2), and in Russian, too, albeit only within implausible stimuli (Experiments 3–5).

These results shed light on a basic fact about language processing: prior expectations about sentence structure affect how the sentence is interpreted. The role of the structural prior has been studied previously, but these investigations focused on analytic languages, which use rigid word order as the main source of syntactic information. This reflects a broader issue in cognitive science: Research subjects are usually from industrialized, rich cultures (Blasi et al., 2022; Henrich et al., 2010), which in psycholinguistics often manifests as Anglo-centric research. However, this poses a major limitation for some of the fundamental questions of cognitive science, like what properties of human cognition are innate. If a cognitive faculty is innate, it must be present across all humans. However, without cross-linguistic research, culture is difficult to separate from cognition. For example, number cognition was once considered an innate cognitive ability independent of language (at least in part because all rich industrialized societies use number). However, evidence was gathered that not all cultures use number words, which led to the current view that number is a cognitive technology, developed within a culture, and not an innate property of human cognition (Frank et al., 2008). In relation to the present study, finding evidence for the structural prior (and, specifically, for the prior over word orders) in analytic languages is not in itself convincing evidence for the universal use of the structural prior across the world's languages: It is possible that the previously observed effects of the structural prior do not reflect a universal processing mechanism but a language-specific tool that speakers of analytic languages use. It may be the case that, because word order is critical for syntactically parsing the sentence in analytic languages, speakers of analytic languages are hyperattuned to any deviation from the canonical word order, and this informs how they interpret sentences. In the present study, we showed that this is not the case and that even speakers of Hindi and Russian, synthetic languages with flexible word order, use the structural prior to interpret the sentences they read. This is a first step toward investigating whether the structural prior influences sentence understanding universally, stepping away from the Anglo-centric bias that has been hindering the generalizability of the cognitive study of language.

We have observed that, in Russian, the effect of canonicity only emerged within implausible sentences. Although we did not predict this pattern of results, it is in line with the noisy channel processing framework. In Russian, the noncanonical word order, OVS, was present in 8.4% of sentences, while, in Hindi, the noncanonical word order, OSV, was present in only 5.7% of sentences. Therefore, the prior probability of noncanonical sentences in Russian is higher than in Hindi. This resulted in a posterior probability of literal interpretations of Russian plausible, noncanonical sentences that were too high to trigger nonliteral interpretations. Seeking to investigate the effect of canonicity within plausible sentences, we tried to lower the overall rate of nonliteral interpretations. To do so, we introduced ungrammatical filler sentences into Experiment 5, following Gibson et al. (2013) and Poliak et al. (2024), who both found more nonliteral interpretations in the presence of noise. However, the manipulation resulted in the same descriptive and inferential patterns as before (except that, numerically, implausible items are interpreted literally slightly less often than in Experiments 3 and 4).

Another potential point of difference between Russian and Hindi is the likelihood function involved. In the experiments reported here, we remained agnostic with regard to what edits are required to turn a noncanonical or an implausible sentence into a canonical and plausible one. However, the fact of the matter is that these edits are different: an implausible, noncanonical sentence can be "corrected" (interpreted as a plausible, canonical sentence) by assuming an exchange of case markings across the nouns; an implausible, canonical sentence can be corrected by exchanging the nouns but leaving the case markings in place; a plausible, noncanonical sentence can be corrected by exchanging the noun phrases in their entirety, noun and case. Is there a difference in the probabilities of an exchange of case, noun, or noun phrase? We currently do not know, but this can be investigated further using corpus analyses and experiments. This could form a potential explanation for why the effect of canonicity does not emerge in Russian plausible sentences: They require an exchange of full NPs, rather than just the nouns or case markings. Moreover, the morphology of case markings is different between Hindi, where case markings are realized as free morphemes that are written with a space after the noun, and Russian, where case markings are realized as bound morphemes at the ends of words and are written without spaces. Russian case markings sometimes trigger a change in the stem

(e.g., dog in nominative case is пёс [pʲos], but in accusative and genitive case, it is пса [psa]). Again, we cannot be certain about whether the difference in the morphology of Hindi and Russian is why the effects of canonicity did not emerge for plausible items in Russian. However, we would like to point out that Poliak et al. (2024) already found robust evidence for noisy channel processing with bound morphemes in Russian, discovering differences in the rates of nonliteral interpretations of sentences based on the underlying corruption of the agreement morpheme of the verb (deletion, insertion, and substitution). That is to say, it seems that the morphological difference between Russian and Hindi did not play a role in why a main effect of canonicity was only observed in Hindi, whereas in Russian, an effect of canonicity was only observed within implausible sentences.

Throughout the study, we have mentioned several times that our models did not detect an interaction between plausibility and canonicity. However, visual inspection seems to suggest that there might be a superadditive effect between plausibility and canonicity. Especially in Russian, we saw using a post hoc analysis that, in both experiments, canonicity had an effect on the probability of literal interpretations within implausible stimuli but not within plausible stimuli. We interpret these results in Russian as a ceiling effect, although we do not have a way of ruling out an underlying interaction. Regardless, we are not theoretically motivated to test for an interaction in these experiments; detecting an interaction would mean that the structural and meaning priors are not independent. This is possible: perhaps, comprehenders do not expect sentences to be both implausible and noncanonical. However, the current experiment does not provide enough information to investigate this hypothesis. In sum, the present study has revealed that the structural prior is used for sentence processing across language types, and it paves the way toward investigating whether the structural prior is used in language processing universally.

References

Arnold, J. B. (2021). *ggthemes: Extra themes, scales and geoms for "ggplot2"* [Computer software]. https://CRAN.R-project.org/package=ggthemes

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In L. R. Gleitman (Ed.), *Language acquisition: The state of the art* (pp. 173–218).

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (Vol. *3*, pp. 73–112).

Berdicevskis, A., & Piperski, A. (2020). Corpus evidence for word order freezing in Russian and German. *Proceedings of the fourth workshop on universal dependencies*. Association for Computational Linguistics.

Bhat, R. A., Bhatt, R., Farudi, A., Klassen, P., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., Vaidya, A., Ramagurumurthy Vishnu, S., & Xia, F. (2017). The Hindi/Urdu Treebank project. In N. Ide & J. Pustejovsky (Eds.), *Handbook of linguistic annotation* (pp. 659–697). Springer.

Bian, Y., Gibson, E., & Levy, R. (2020). *Noisy channel processing of questions* [Poster presentation]. CUNY Sentence Processing Conference, University of Massachusetts Amherst, Amherst, MA, United States. https://websites.umass.edu/cuny2020/

Birner, B. J., & Ward, G. L. (1998). *Information status and noncanonical word order in English*. John Benjamins Publishing. https://doi.org/10.1075/slcs.40

Bizzarri, C. (2015). Russian as a partial pro-drop language. *Annali di Ca'Foscari. Serie Occidentale*, *49*, 335–362. https://doi.org/10.14277/2385-3034/AnnOc-49-15-18

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Bloom, D. B. (1999). *Case syncretism and word order freezing in the Russian language* [Unpublished master's thesis]. Stanford University.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Bouma, G. (2011). Production and comprehension in context. In A. Benz & J. Mattausch (Eds.), *Bidirectional optimality theory* (pp. 169–189). John Benjamins Publishing. https://doi.org/10.1075/la.180.07bou

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

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

Butt, M. (2001). Case, agreement, pronoun incorporation and pro-drop in South Asian languages. *Workshop on the role of agreement in argument structure* (pp. 1–20). Utrecht University. https://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/utrecht01-hnd.pdf

Chen, S., Nathaniel, S., Ryskin, R., & Gibson, E. (2023). The effect of context on noisy-channel sentence comprehension. *Cognition*, *238*, Article 105503. https://doi.org/10.1016/j.cognition.2023.105503

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. Harcourt Brace Jovanovich.

Clifton, C., Jr., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & Cognition*, *32*(6), 886–895. https://doi.org/10.3758/BF03196867

Droganova, K., Lyashevskaya, O., & Zeman, D. (2018). Data conversion and consistency of monolingual corpora: Russian UD treebanks. *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)* (Vol. *155*, pp. 53–66).

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2024). *Ethnologue: Languages of the world* (27th ed.). SIL International. https://www.ethnologue.com

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738. https://doi.org/10.1006/jmla.1995.1032

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819–824. https://doi.org/10.1016/j.cognition.2008.04.007

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), Article e12814. https://doi.org/10.1111/cogs.12814

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical.

Language and Cognitive Processes, *14*(3), 225–248. https://doi.org/10.1080/016909699386293

Goldberg, A. E. (2010). Verbs, frames and constructions. In M. Rappaport Hovav, E. Doron, & I. Sichel (Eds.), *Syntax, lexical semantics and event structure* (pp. 39–58). Oxford University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hörnig, R., Oberauer, K., & Weidenfeld, A. (2005). Two principles of premise integration in spatial reasoning. *Memory & Cognition*, *33*(1), 131–139. https://doi.org/10.3758/BF03195303

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. https://doi.org/10.1016/j.jml.2017.01.004

Jakobson, R. (1971). *Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus (Vol. II)* [Contribution to the general theory of case: The overall meanings of the russian cases (Vol. II)]. De Gruyter Mouton. https://doi.org/10.1515/9783110873269.23

Kamide, Y., & Kukona, A. (2018). The influence of globally ungrammatical local syntactic constraints on real-time sentence comprehension: Evidence from the visual world paradigm and reading. *Cognitive Science*, *42*(8), 2976–2998. https://doi.org/10.1111/cogs.12694

Kasl, S. V., & Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, *1*(5), 425–433. https://doi.org/10.1037/h0021918

Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.1308151

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject–verb agreement to form more probable linguistic structures. *Cognitive Psychology*, *124*, Article 101359. https://doi.org/10.1016/j.cogpsych.2020.101359

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 326–347. https://doi.org/10.1037/a0034903

Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 234–243). Association for Computational Linguistics. https://doi.org/10.3115/1613715.1613749

Levy, R. (2011). Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human language technologies* (pp. 1055–1065). Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21086–21090. https://doi.org/10.1073/pnas.0907664106

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

Liu, Y., Ryskin, R., Futrell, R., & Gibson, E. (2020, September). *Structural frequency effects in comprehenders' noisy-channel inferences* [Conference session]. 26th Conference on Architectures and Mechanisms for Language Processing (AMLaP).

MacWhinney, B. (2022). The competition model: Past and future. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A life in cognition: Studies in cognitive science in honor of Csaba Pléh* (pp. 3–16). Springer.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Mahajan, A. K. (1988). *Word order and negation in Hindi*. Massachusetts Institute of Technology.

Mahowald, K. (2011). An LFG approach to word order freezing. In M. Butt & T. H. King (Eds.), *Proceedings of LFG11* (pp. 381–398). Clinical and Laboratory Standards Institute.

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), Article 1541. https://doi.org/10.21105/joss.01541

Palmer, M., Bhatt, R., Narasimhan, B., Rambow, O., Sharma, D. M., & Xia, F. (2009). Hindi syntax: Annotating dependency, lexical predicate-argument structure, and phrase structure. *The 7th International Conference on Natural Language Processing* (pp. 14–17).

Poliak, M., Malik-Moraleda, S., & Gibson, E. (2025, January 8). *Noisy_Channel_Free_Word_Order*. Open Science Framework. https://osf.io/qyjk6/

Poliak, M., Ryskin, R., Braginsky, M., & Gibson, E. (2024). It is not what you say but how you say it: Evidence from Russian shows robust effects of the structural prior on noisy channel inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(4), 637–649. https://doi.org/10.1037/xlm0001244

Poppels, T., & Levy, R. P. (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 378–383).

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(10), Article e2307876121. https://doi.org/10.1073/pnas.2307876121

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, *37*(1), 10–21. https://doi.org/10.1109/JRPROC.1949.232969

Slioussar, N., & Harchevnik, M. (2024). Word order and context in sentence processing: Evidence from L1 and L2 Russian. *Frontiers in Psychology*, *15*, Article 1344366. https://doi.org/10.3389/fpsyg.2024.1344366

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 186–199. https://doi.org/10.1037/a0033206

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 431–450. https://doi.org/10.1037/0278-7393.30.2.431

Thompson, I. (1977). Russian word order: A comparative study. *The Slavic and East European Journal*, *21*(1), 88–103. https://doi.org/10.2307/306737

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

Vasishth, S. (2004). Work order, negation, and negative polarity in Hindi. *Research on Language and Computation*, *2*(1), 127–146. https://doi.org/10.1023/A:1025480005462

Vasishth, S., Shaher, R. F., & Srinivasan, N. (2012). The role of clefting, word order and given-new ordering in sentence comprehension: Evidence from Hindi. *Journal of South Asian Linguistics*, *5*, 35–56. https://ojs.ub.uni-konstanz.de/jsal/index.php/jsal/article/view/27

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2024). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wijffels, J. (2023). *udpipe: Tokenization, parts of speech tagging, lemmatization and dependency parsing with the "UDPipe" "NLP" toolkit* [Computer software]. https://CRAN.R-project.org/package=udpipe

Yano, M., & Koizumi, M. (2018). Processing of non-canonical word orders in (in)felicitous contexts: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *33*(10), 1340–1354. https://doi.org/10.1080/23273798.2018.1489066

Zeman, D., Nivre, J., Abrams, M., Ackermann, E., Aepli, N., Aghaei, H., Agić, Ž., Ahmadi, A., Ahrenberg, L., Ajede, C. K., Akkurt, S. F., Aleksandravičiūtė, G., Alfina, I., Algom, A., Alnajjar, K., Alzetta, C., Andersen, E., Antonsen, L., Aoyama, T., … Ziane, R. (2023). *Universal dependencies 2.12*. https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-5150

Appendix A

An Acceptability Study in Russian

In the current project, we claim that past work by Poliak et al. (2024) is unrepresentative of sentence processing in Russian. We make this claim because Poliak et al. (2024) investigated how noncanonical sentences are understood in Russian sentences where all NPs are names that cannot be inflected for case, resulting in unusual conditions called "word-order freezing" (Berdicevskis & Piperski, 2020; Bouma, 2011; Jakobson, 1971). To investigate whether it is indeed the case we conducted an acceptability study, varying whether the nouns (names) in the sentence were marked for case (cased/uncased) and the type of sentence resulting from manipulating verb agreement (SVO/OVS/ungrammatical). See Table A1 for a sample item.

We preregistered the study on the Open Science Framework[A1] (https://osf.io/qyjk6/), and all the materials, data, and analyses are publicly available on the project page, together with the materials from the main experiment. We presented 30 items to each of 49 native speakers of Russian (we recruited 50 and excluded one for reporting their first language to be English and no sufficient other contexts for using Russian in daily life, as per the preregistration). See Figure A1 for a visualization of the results.

Table A1
A Sample Item From Appendix A (Russian Acceptability)

Sentence type	Case marking	Sample item
SVO	Case marked	Вера увидела Арсения Vera.NOM saw-FEM Arsenij.ACC Vera saw Arsenij
OVS	Case marked	Веру увидел Арсений Vera.ACC saw-MASC Arsenij.NOM Arsenij saw Vera
Ungrammatical	Case marked	Вера увидели Арсения Vera.NOM saw-PL Arsenij.ACC Vera saw Arsenij (ungrammatical)
SVO	Case unmarked	Рейчел увидела Чарли Rachel saw-FEM Charlie Rachel saw Charlie
OVS	Case unmarked	Рейчел увидел Чарли Rachel saw-MASC Charlie Charlie saw Rachel
Ungrammatical	Case unmarked	Рейчел увидели Чарли Rachel saw-PL Charlie Rachel saw Charlie (ungrammatical)

Note. SVO = subject–verb–object; NOM = Nominative; FEM = feminine; ACC = accusative; OVS = object–verb–subject; MASC = masculine; PL = plural.

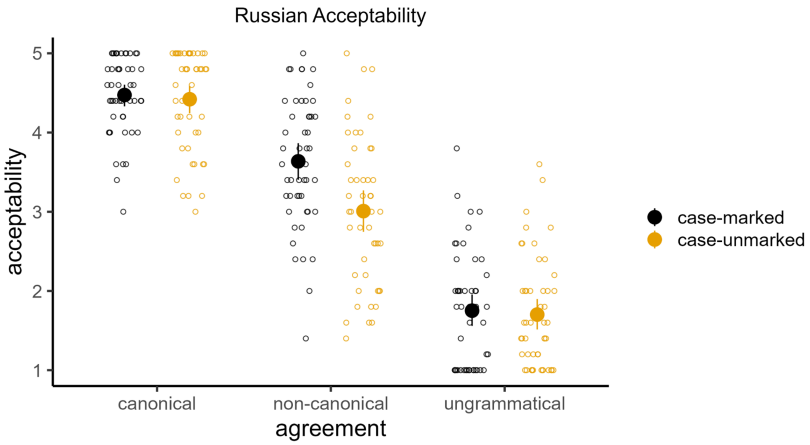
To analyze the data, we fit an ordinal Bayesian regression with main effects for case marking (treatment coded, reference level = case-marked) and verb agreement (treatment coded, reference level = OVS), with the full random effects structure (random intercepts for participants and items, and random slopes for both conditions and their interaction within participants and items). We used the default priors of the brms package, 4,000 iterations (2,000 warmup), observed no divergent transitions, and all parameters had $\hat{R}$ = 1.00 (except one correlation between the random effect with $\hat{R}$ = 1.01). We detected a main effect of case marking, such that case-unmarked sentences were substantially less acceptable than case-marked sentences within OVS sentences ($M = -1.23$, CrI [−1.82, −0.63]). We also found substantial evidence for main effects of verb agreement, such that SVO sentences were more acceptable than OVS sentences ($M = 2.30$, CrI [1.61, 3.04]), and ungrammatical sentences were less acceptable than OVS sentences ($M = -4.68$, CrI [−5.76, −3.68]). We also found positive interactions between case marking and verb agreement, such that case-unmarked sentences were more acceptable than would be expected from the main effects only in the SVO and ungrammatical conditions (SVO: $M = 1.25$, CrI [0.44, 2.11]; ungrammatical: $M = 1.23$, CrI [0.36, 2.10]). This suggests that OVS (noncanonical) sentences become uniquely less acceptable when the sentence lacks case marking. This supports our claim that OVS sentences without case markings are not processed like other regular sentences in Russian, and therefore, the results of Poliak et al. (2024) should not be generalized to all of Russian.

[A1] We made a mistake and forgot to click the final button for submitting the preregistration. As a result, it is officially preregistered after the data had been analyzed, although we have not changed it from the time we wrote it, before collecting any data.

(*Appendices continue*)

Figure A1

Mean Acceptability Rating Split Across Verb Agreement and Case Marking



Note. Error bars are bootstrapped 95% confidence intervals over participant means. Unfilled circles represent participant means. See the online article for the color version of this figure.

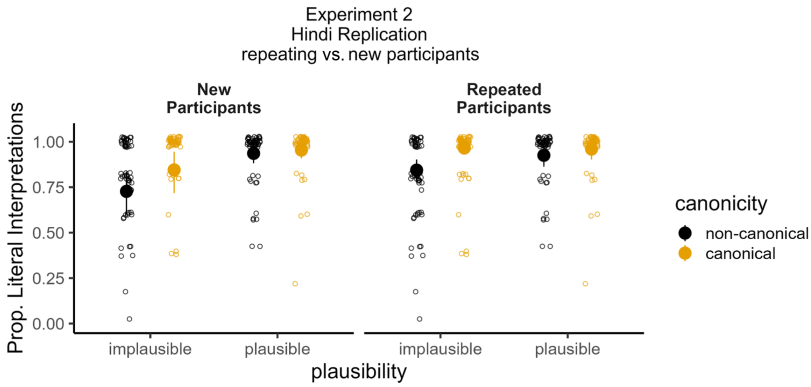
Appendix B

Repeated Participants in Experiment 2 (Hindi Replication)

We wished to investigate whether the results of Experiment 2 were driven by returning participants. Therefore, we fit the same model as in the previous section adjusted to account for main and interaction effects of returning participants (coded as −0.5 = naive, 0.5 = returning), resulting in the following formula for fixed effects: Plausibility × Canonicity × Returning. The inference for the critical effects remained the same: the 95% credible intervals for intercept, plausibility, and canonicity were all positive and did not contain 0, while all other credible intervals in the model contained 0, suggesting no substantial evidence for a difference between naive and returning participants. See Figure B1 for a visualization of the results for new and repeated participants.

Figure B1

Results of Experiment 2 (Hindi Replication)



Note. The figure represents the proportion of literal interpretations for critical sentences, split across canonicity and plausibility. Error bars are bootstrapped 95% confidence intervals over participant means. Unfilled circles represent participant means. Prop. = proportion. See the online article for the color version of this figure.

Received July 25, 2024
Revision received December 30, 2024
Accepted January 6, 2025 ■