Reliance on semantic and structural heuristics in sentence comprehension across the lifespan

Anastasiya Lopukhina[a,b,*], Anna Laurinavichyute[a,c], Svetlana Malyutina[a], Galina Ryazanskaya[c], Elena Savinova[d], Aleksandra Simdianova[e], Anastasia Antonova[a], Irina Korkina[a]

[a] HSE University, Russia

[b] Vinogradov Institute of the Russian Language, Russia

[c] University of Potsdam, Germany

[d] Utrecht University, The Netherlands

[e] University of Trento, Italy

* Corresponding author:

Anastasiya Lopukhina, Center for Language and Brain, HSE University, Room 306, Krivokolenny pereulok, 3, Moscow, Russia, 101000.

Email: alopukhina@hse.ru

DOI: 10.1177/17470218211053263

Abstract

People sometimes misinterpret the sentences that they read. One possible reason suggested in the literature is a race between slow bottom-up algorithmic processing and "fast and frugal" top-down heuristic processing that serves to support fast-paced communication but sometimes results in incorrect representations. Heuristic processing can be both semantic, relying on world knowledge and semantic relations between words, and structural, relying on structural economy. Scattered experimental evidence suggests that reliance on heuristics may change from greater reliance on syntactic information in younger people to greater reliance on semantic information in older people. We tested whether the reliance on structural and semantic heuristics changes with age in 137 Russian-speaking adolescents, 135 young adults, and 77 older adults. In a self-paced reading task with comprehension questions, participants read unambiguous high- vs. low-attachment sentences that were either semantically plausible or implausible: i.e., the syntactic structure either matched or contradicted the semantic relations between words. We found that the use of top-down heuristics in comprehension increased across the lifespan. Adolescents did not rely on structural heuristics, in contrast to young and older adults. At the same time, older adults relied on semantic heuristics more than young adults and adolescents. Importantly, we found that top-down heuristic processing was faster than bottom-up algorithmic processing: slower reading times were associated with greater accuracy specifically in implausible sentences.

Keywords: semantic heuristics; structural heuristics; good-enough processing; adolescence; language and ageing; Russian

DOI: 10.1177/17470218211053263

Introduction

People's interpretation of sentences does not always accurately reflect the actual content of the sentences (Barton & Sanford, 1993; Erickson & Mattson, 1981; Fillenbaum, 1971, 1974; Kamas et al., 1996; Wellwood et al., 2018). This might happen for a number of reasons, one of which is a race between grammar-driven incremental bottom-up processing and "fast and frugal" top-down heuristic processing that serves to support fast-paced communication but sometimes causes misinterpretations (Christianson et al., 2001; Christianson, 2016; Ferreira et al., 2002; Ferreira, 2003; Ferreira & Yang, 2019; Kuperberg, 2007). Top-down heuristic processing may be based on semantic and structural heuristics.

Semantic top-down heuristics are guided by semantic relations between words and by the comprehender's world knowledge. According to the online equilibrium hypothesis of the good-enough processing theory (Christianson et al., 2001; Christianson et al., 2010; Ferreira et al., 2002; Karimi & Ferreira, 2016; Koornneef & Reuland, 2016), heuristic-based representations are computed faster than full syntactically-based representations. Top-down semantic heuristic and bottom-up algorithmic processing launch simultaneously, and top-down heuristic processing results in an interim representation. If this representation is plausible and can be incorporated in the existing knowledge structure, the language processing system can abandon bottom-up algorithmic processing. In that case, the final representation is defined by the output of semantic heuristic processing and can differ from the representation encoded in the sentence. For example, '*The dog was bitten by the man*' can be misinterpreted as 'the dog bit the man' (Ferreira, 2003), indicating that strong semantic heuristics can completely overrule contradicting grammatical information.

Structural top-down heuristics rely on the principle of structural economy: simpler structures (for example, those with fewer syntactic nodes, or those that involve attachment to the closest possible antecedent) become available faster than more complex ones. Several

DOI: 10.1177/17470218211053263

heuristics have been proposed in the literature. For example, the late closure heuristic stipulates adding new material to the clause currently being processed. In the sentence '*Emily said that Lucy went shopping yesterday'*, comprehenders prefer to attach '*yesterday*' to the subordinate clause '*Lucy went…*' rather than to the main clause '*Emily said…*' (Frazier, 1987). Another example is the minimal attachment heuristic that relies on building the structure with the fewest syntactic nodes. In the sentence '*The girl hit the boy with the book*', comprehenders prefer to attach '*with the book*' to '*hit*' rather than to '*the boy*' (although this may differ cross-linguistically, Sekerina et al., 2003). Structural heuristics can promote parses that are ruled out within the global sentence context, such as the subject-verb parse of a string '*the player tossed*' in '*The coach smiled at the player tossed a frisbee by the opposing team'*, where *'tossed'* requires a low-frequency past participle parse (Tabor et al., 2004). When a parse resulting from applying structural heuristics does not match the actual syntax of the sentence, as in the example above, a misinterpretation can arise, similarly to when semantic heuristics are applied.

Scattered experimental evidence suggests that reliance on top-down heuristics may change from greater reliance on syntactic information in younger people to greater reliance on semantic information in older people. Several studies showed that children and pre-adolescents relied on syntactic information and structural heuristics while disregarding semantic and contextual information (Clahsen & Felser, 2006; Felser et al., 2003; Trueswell et al., 1999). For example, in two self-paced reading experiments with 8-to-12-year-old children, Traxler (2002) found that when reading temporarily ambiguous sentences like '*When Sue tripped the table fell over and the vase was broken*', children relied on the structural heuristic of minimal attachment and were not sensitive to semantic implausibility cues. Unlike adults, children favoured structurally simpler analysis of the sentences no matter whether the final representation was semantically plausible or not. One of the explanations, suggested by the author, referred to working memory limitations: possibly, children had difficulty in re-analysis

DOI: 10.1177/17470218211053263

of previously processed information. Additionally, in an eyetracking-while-reading experiment with adults and 7-to-12-year-old children, Joseph and colleagues (2008) showed that children were slower than adults in detecting implausibility in sentences like '*Robert used a hook to catch the horrible mouse that was very scared*'. The delayed implausibility detection in children could stem from slower access to semantic information and less efficient integration of real-world knowledge into sentence representations.

Importantly, reliance on semantic heuristics seems to increase with age. In an eyetracking-while-reading experiment with 9-to-16-year-old participants, Engelhardt (2014) found that older participants were more likely to rely on semantic information in temporarily ambiguous sentences like '*While the storm blew the boat sat in the shed*' and had higher comprehension accuracy compared to younger participants, who relied on the structural heuristic of minimal attachment and made more comprehension errors. To the best of our knowledge, this is the only developmental study of heuristic and algorithmic processing that included adolescents, even though not as a separate group. Most other data were obtained from children and pre-adolescents, so it remains largely unknown how adolescents who have more semantic knowledge than younger children use heuristic and algorithmic processing, especially compared to adults.

The lifespan trend towards greater reliance on semantic heuristics appears to continue in older adulthood. Older adults show continuing growth in lexicon (Verhaeghen, 2003) and world knowledge (Ramscar et al., 2014). At the same time, they allocate fewer cognitive resources to processing syntactic information than young adults (Beese et al., 2019a; Stine-Morrow et al., 2000), and their comprehension of complex syntax deteriorates (Malyutina et al., 2018; Obler et al., 1991). In light of these findings, older adults can be predicted to rely more on semantic heuristics than algorithmic syntactically-based processing. This prediction was confirmed for "garden-path" sentences with temporary syntactic ambiguities (Christianson

DOI: 10.1177/17470218211053263

et al., 2006; Malyutina & Den Ouden, 2016) and for syntactically complex sentences with a mismatch between syntax and semantics (Amichetti et al., 2016, hearing-impaired older adults in Experiment 2; Obler et al., 1991). However, in some studies, older adults and young adults' reliance on semantic heuristics did not differ (Amichetti et al., 2016, Experiment 1; Yoon et al., 2015). Further evidence is needed to support the claim that reliance on semantic heuristics increases with age. Besides, most previous studies only compared young to older adults (although see Obler et al., 1991). To better elucidate lifespan trajectories of reliance on semantic heuristics, it would be interesting to compare multiple age groups within a single experimental design.

Evidence on the trajectories of structural heuristics across the adult lifespan is scarcer. Structural heuristics do vary across individuals (for example, depending on the working memory span, Fedorova & Yanovich, 2006; Payne et al., 2014) but only few studies have investigated whether they depend on age. For example, Kjelgaard and colleagues (1999) and Titone and colleagues (2006) found comparable structural preferences for late closure in both older and young adults. However, the findings by Payne and colleagues (2014) suggested a more complex pattern, whereby structural heuristics were affected by age in interaction with working memory and reading experience. Namely, they reported a general low-attachment preference in English that was further increased in older adults with greater reading experience but reversed in older adults with lower working memory span. Further experimental evidence is needed to conclude whether reliance on structural heuristics varies across the lifespan.

The current study is the first to investigate the reliance on top-down semantic and structural heuristics during sentence reading in adolescents, young adults, and older adults. We tested three hypotheses. First, we expected that reliance on "fast and frugal" top-down semantic heuristics would increase with age: namely, adolescents might rely on semantic heuristics less than young adults, and young adults less than older adults. Second, we hypothesized that

DOI: 10.1177/17470218211053263

reliance on top-down structural heuristics might also increase with age. To assess the reliance on structural heuristics, we employed structures that are known to bias comprehension: in Russian sentences with a genitive noun phrase consisting of two nouns, and a participial clause attached to one of these nouns, high attachment is strongly preferred (Chernova, 2016; Fedorova & Yanovich, 2006; Sekerina, 1997; although see alternative interpretations of attachment effects based on implicit prosody or discourse prominence: Fodor, 2002; Chernova & Chernigovskaya, 2015; Chernova, 2016). We expected that this bias should be weaker in Russian adolescents and stronger in older adults compared to young adults. Finally, although one of the central claims of the online equilibrium hypothesis (Karimi & Ferreira, 2016) is that reliance on heuristic processing provides a faster route for comprehension, empirical studies have never evaluated whether good-enough representations are indeed formed faster (Ferreira & Yang, 2019). We aimed to investigate whether faster reading speed would be associated with lower response accuracy in implausible sentences, indicating that semantic top-down heuristic processing is faster than bottom-up algorithmic processing.

Method[1]

Participants

Participants belonged to three age groups: 137 adolescents (87 female; age range 13-17 years, M = 15, SD = 1.1), 135 young adults (99 female; age range 20–40 years, M = 25, SD = 5.8), and 77 older adults (57 female; age range 55–91 years, M = 64, SD = 7.8). Adolescents were recruited from the public schools, young adults from universities and via social networks, older adults from social service centers and by word of mouth. All participants were native

[1] The complete set of stimuli, data, and code for the analysis are available online: DOI 10.17605/OSF.IO/4F2PX This study is part of a project investigating good-enough language processing across the lifespan under no-noise and linguistic-noise conditions (preregistration https://osf.io/k3ebs). In this paper, we focus on the effect of participants' age on sentence processing. We analyze the data from the no-noise part of experiments and test only a subset of preregistered hypotheses.

DOI: 10.1177/17470218211053263

speakers of Russian, with normal or corrected-to-normal vision and no history of neurological, psychiatric or speech/language disorders. Young adults were more educated (M = 15 years) than adolescents (M = 9 years) who were in Grades 7-11 ($t = 24$, $p < 0.001$), the latter were on track to achieving the same level of education. Young adults were also more educated than older adults (M = 14 years), $t = 3.4$, $p < 0.001$. Adolescents and young adults received no compensation, while older adults received 400 RUB for participating. All adult participants and the parents of the adolescents gave written informed consent before the experiment. The study was carried out in accordance with the Declaration of Helsinki and the code of conduct of the American Psychological Association and was approved by the HSE Committee on Interuniversity Surveys and Ethical Assessment of Empirical Research.

Materials

The experimental stimuli were Russian sentences with a genitive noun phrase and a participial clause attached to one of the two nouns in the noun phrase. Every sentence was followed by a two-choice comprehension question targeting the attachment site of the participle, see Table 1. We manipulated two factors: (i) the syntactic attachment site of the participle to either the head noun (high attachment) or the dependent noun (low attachment) in the genitive noun phrase (Russian has a preference for high attachment, Chernova, 2016; Sekerina, 1997 reports a preference for high-attachment interpretation of about 80%), and (ii) the semantic match between the participial clause and the head noun (plausible / implausible), resulting in four experimental conditions. All sentences regardless of attachment had a comma before the participial clause that is mandatory in Russian and thus all sentences implied the same prosody.

Importantly, all stimuli sentences were completely unambiguous because syntactic attachment was unambiguously marked by case inflection. In Russian, both nouns and

DOI: 10.1177/17470218211053263

participles are inflected by case, which is marked via case endings. The case of the first noun in the noun phrase is governed by the verb: it can be dative, accusative, instrumental or prepositional depending on the specific verb (for example, in Table 1, the verb '*to dress*' requires the accusative case). The case of the second noun in the noun phrase is governed by the first noun and is always genitive, with the semantics of possession or affiliation (for example, in Table 1, '*the child*' is assigned the genitive case, meaning that the child belongs to the writer). Finally, the critical participle is assigned the case of the noun it agrees with: either the first noun (thus either dative, accusative, instrumental or prepositional case; see conditions A and C in Table 1) or the second noun (thus genitive case; conditions B and D). The case ending of the participle indicates agreement unambiguously.

Notably, Russian uses different sets of case endings for different grammatical genders, for nouns versus participles, and depending on the final consonant of the word stem. Even though noun and participle case endings have a phonological overlap for some cases and genders (for example, -*у [-u]* and -*ую [-uju]* in the accusative case for feminine nouns and participles respectively), more often they do not (for example, *-ы [-y]* and *-ей [-ei]* in the genitive case for feminine nouns and participles respectively) and they are never completely identical. Thus, grammatical parsing can hardly rely on the heuristics of phonological overlap, which could be possible in languages with phonologically identical case endings or prefixes across word classes, genders, etc., like Kiswahili (Alcock & Ngorosho, 2004).

Even though all stimuli sentences were syntactically unambiguous, sentences of such structure were previously demonstrated to lead to as much as 63% errors in comprehension questions in the low attachment condition (Chernova, 2016). While this surprisingly strong bias for high attachment in Russian is intriguing in itself, in this study, we merely relied on the established effect and did not investigate the reasons that might drive it.

DOI: 10.1177/17470218211053263

Semantic plausibility was assessed in a separate norming study.[2] For each potential stimulus sentence, participants (N=188) were presented with a verb phrase, corresponding to the participial clause in the sentence, and two nouns, corresponding to the head and dependent nouns in the genitive noun phrase in the sentence. Participants rated which of the two nouns was more plausible as a subject of the action. On a 5-point scale, one noun was placed at 1 and the other at 5, so that, for example, a rating of 3 would indicate that the two nouns are equally plausible subjects of the action. For the final stimuli list, we selected participles that were rated as highly plausible with one noun from the genitive noun phrase (mean rating 4.52, range 3.82-5.00, or mean rating 1.44, range 1.03-2.00) and implausible with the other noun. A similar norming strategy to obtain plausibility ratings was used in (Stoops et al., 2014) where the authors compared animacy cues and morphological cues in simple unambiguous Russian sentences.

--- Insert Table 1 about here ---

Fifty-six experimental blocks of sentences on each of the four conditions were divided into two sets of 28 blocks, and each was assigned to half of the participants. The two sets were created because each participant took part in the current experiment and in the experiment with linguistic noise interfering with comprehension (not reported here, see also the pre-registration in Footnote 1). The two stimuli sets were matched for sentence length in words and syllables, length in syllables of the critical nouns, gender of the critical nouns, and case of the first critical noun. Each set included an equal number of feminine and masculine critical nouns. To ensure syntactic unambiguity, the head nouns in masculine were in dative, instrumental, and prepositional cases; the head nouns in feminine were always in accusative case. For each set, four experimental lists were used.

[2] The results of the online norming study are available online: DOI 10.17605/OSF.IO/4F2PX

DOI: 10.1177/17470218211053263

Each list contained three types of fillers: plausible sentences structurally similar to the stimuli with either high (N=9) or low (N=9) attachment, followed by comprehension questions addressing parts of the sentences other than the attachment site of the participle; and sentences that were structurally different from the stimuli (N=38). Overall, each participant read 28 experimental items interspersed with 56 fillers.

We assumed that semantic heuristic processing would manifest in lower accuracy rates in implausible compared to plausible sentences. If there would be no difference in accuracy rates between the plausible and implausible condition, this would indicate reliance on algorithmic syntactically-based processing or solely structural heuristics. Similarly, structural heuristic processing would manifest in lower accuracy rates in low-attachment compared to high-attachment sentences. The absence of this difference would indicate reliance on algorithmic processing or solely on semantic heuristics. Importantly, we expected that longer reading times in the implausible condition should be associated with higher accuracy. Based on the online equilibrium theory, we hypothesized that accurate responses in the implausible condition, presumably indicating algorithmic processing, should take more time than incorrect responses reflecting semantic heuristic processing.

Procedure

Each participant was tested in two sessions: one session with linguistic auditory or visual noise and the other session without noise. Session order was pseudorandomized, with a 15-minute break between sessions. In this study, we report only data from the no-noise session.

A non-cumulative self-paced reading paradigm was presented using E-Prime 2.0 (Schneider et al., 2002). The non-cumulative variant of the paradigm was chosen (similarly to Caplan et al., 2011, Chernova, 2016) so that participants could not return to word length information and use it to infer the case after proceeding to read further parts of the sentence.

DOI: 10.1177/17470218211053263

Stimuli were presented on a laptop, with each word appearing in the center of the screen. After each sentence, participants read a two-alternative comprehension question and responded by pressing a key. Before the experiment, participants completed a training of five sentences similar to fillers. One session lasted approximately 20 minutes and included three short breaks.

Analysis

The main dependent variable was response accuracy. In addition, we ran a post-hoc exploratory analyses of reading times (RTs) at four regions: the word preceding the participle (n-1, identical across all conditions), the participle (n, had different endings in all conditions, indicating different case forms: *лепечущ**ую*** 'babble-PP-ACC' / *лепечущ**ей*** 'babble-PP-GEN' / *опубликовавш**ую*** 'publish-PP-ACC' / *опубликовавш**ей*** 'publish-PP-GEN', see Table 1), the first and the second spillover regions (n+1 and n+2, which were identical in pairs of conditions, see Table 1). The results of RT analysis should be treated with caution since we analyzed RTs on different word forms and different lexemes.

Analyses were conducted with the R system for statistical computing (R Core Team, 2016). Response accuracy was analyzed using hierarchical logistic regression, and RTs using hierarchical linear regression (assuming lognormal distribution of the data), fit in the Bayesian framework using the 'brms' package (Bürkner, 2017). Plots were produced with the 'ggplot2' and 'tidybayes' packages (Wickham, 2016; Kay, 2019). Posterior distributions of the parameters are reported in terms of the posterior mode and 95% percentile intervals.

Sum contrasts were used to code attachment and plausibility (with 1 denoting the dispreferred option, i.e., low attachment and implausible conditions, and -1 denoting the preferred option). Age was coded using treatment contrasts with young adults as the reference level, so that all simple effects in the model are estimates for the young adults. The models

DOI: 10.1177/17470218211053263

included interactions between age and plausibility, as well as between age and attachment (this interaction was not preregistered).

To assess whether heuristic-based processing was faster, in the model of response accuracy we also included sentence mean log-reading time (centered and scaled) and its interaction with age, with plausibility, and with attachment (this analysis was not preregistered). To account for potential processing strategy transfer from the noise to no-noise session, we included the interaction between the type of noise in the noise session and session number. This interaction had no effect on the dependent variable and will not be reported.

The models included random intercepts for participants and items as well as by-item random slopes for all main effects and interactions, and by-participant random slopes for the main effects of attachment and plausibility. Note that we could not include by-participant random slopes for age and interactions with age, as the age of each participant is constant. The full model structure for the analysis of response accuracy was as follows:

accuracy ~ age(plausibility + attachment) + log.centered.RT*(age + plausibility + attachment) + plausibility:log.centered.RT:age + noise:session +* (*1 + age**(*plausibility + attachment*) *+ log.centered.RT*(age + plausibility + attachment) + plausibility:log.centered.RT:age || ItemID*) + (*1 + plausibility + attachment || ParticipantID*).

We used regularizing priors that allow for any plausible effect sizes but downweigh extreme values: for the intercept, representing the grand mean for young adults, the prior on the logit scale was *Normal(0, 2)*, which allows for any accuracy in the range from 2% to 98%. The prior for the coefficients for the predictors was *Normal(0, 1)*, allowing for both positive and negative effects spanning up to 76%, depending on the intercept.

In the analysis of word reading times, we additionally included trial accuracy as well as centered word length and centered frequency (only for the participle region) as predictors, rendering the following model structure:

DOI: 10.1177/17470218211053263

RT ~ age(plausibility + attachment)*accuracy + session:noise + word.length.centered + word.freq.centered + (1 + age*(plausibility + attachment)*accuracy || ItemID) + (1 + (plausibility + attachment)*accuracy || ParticipantID).*

For the modeling of RTs, we also used regularizing priors: for the intercept, representing the grand mean for young adults, the prior on the log-normal scale was *Normal(5, 1),* which allows for any RTs in the range from 20 to 1100 ms. The prior for the coefficients for the predictors was *Normal(0, 0.5)*, allowing for both positive and negative effects spanning up to 2600 ms, depending on the intercept.

Results

Reading times

The general pattern of reading times across four analyzed sentence regions can be seen on Figure 1. Detailed results of statistical analysis are presented in Table 2. In the region preceding the participle, we found only slower processing in older adults. In the critical participle region, we again found a slowdown in older adults and in correct trials, as well as a speedup in adolescents, and a speedup for more frequent word forms. Importantly, we found an interaction between plausibility and accuracy: implausible sentences were processed more slowly, but only on trials with correct responses. In the two spillover regions, we again found slower processing in older adults (both n+1 and n+2) and a slowdown in implausible conditions in correct trials (in region n+1, only in adolescents; in n+2, in all groups). In addition, in region n+2, adolescents read faster than young adults. Across all regions, we also found a trivial slowdown in reading longer words.

--- Insert Figure 1 about here ---

DOI: 10.1177/17470218211053263

Response accuracies

Response accuracy in fillers was on average 96% (range 75-100%), indicating that the participants were highly engaged in the task. No participants were excluded from the analysis. For experimental stimuli, mean accuracy and standard deviations are summarized in Table 3. The estimated proportions of correct responses for each participant group across conditions are shown in Figures 2-3, and the posterior distributions of the parameters are shown in Table 4.

--- Insert Table 3 about here ---

As expected, we found that young adults made more errors in the implausible than plausible condition, indicating reliance on semantic heuristics, and in the low-attachment than high-attachment condition, indicating reliance on the structural heuristic of high attachment. Slower reading times in young adults were associated with higher accuracy. There was an interaction between reading speed and plausibility in young adults: for implausible sentences, slower reading times were associated with an additional increase in accuracy. We did not find the interaction between reading speed and attachment in young adults.

Older adults had lower accuracy than young adults across the board and showed a greater decrease in accuracy than young adults in implausible sentences, which indicates greater reliance on semantic heuristics, and the same accuracy decrease in the low-attachment condition, which indicates similar reliance on the structural heuristic of high attachment. As in young adults, slower reading times were associated with greater accuracy, and there was the same interaction between reading speed and plausibility: for implausible sentences, slower reading times were associated with an additional increase in accuracy.

Adolescents did not differ from young adults in the overall accuracy, and had a similar decrease in accuracy in the implausible conditions, indicating reliance on semantic heuristics.

DOI: 10.1177/17470218211053263

However, in the low-attachment conditions, adolescents made fewer errors than young adults. In fact, there was no effect of attachment in adolescents, Prob($\beta < 0$) = 0.86, indicating no reliance on structural heuristics. Similarly to the results of young adults, in adolescents, higher reading times were associated with higher accuracy. The interaction between plausibility and reading speed was also present in adolescents: for implausible sentences, slower reading times were associated with an additional increase in accuracy.

--- Insert Figure 2 about here ---

--- Insert Figure 3 about here ---

--- Insert Table 4 about here ---

Discussion

This study investigated whether and to what extent adolescents, young adults, and older adults relied on top-down semantic and structural heuristics during sentence reading. We confirmed the prediction of the good-enough theory (Christianson et al., 2001; Ferreira et al., 2002) that comprehenders would rely on semantic heuristics and misinterpret semantically implausible sentences. We also confirmed that Russian adults apply the structural heuristic of high attachment: they misinterpret sentences with low attachment more often than sentences with high attachment (Chernova, 2016; Fedorova & Yanovich, 2006). Our novel contribution is that the use of heuristics in comprehension increases across the lifespan. Reliance on semantic heuristics increased in older adults as compared to young adults and adolescents. Reliance on structural heuristics was not detected at all in adolescents, in contrast to young and older adults. In addition, in line with the predictions of the online equilibrium hypothesis, the use of semantic heuristics (as reflected by incorrect responses to comprehension questions) was associated with faster reading in every age group.

DOI: 10.1177/17470218211053263

Semantic heuristics

Our first goal was to ascertain whether reliance on semantic heuristics would increase with age. Although we expected to find lower reliance on semantic heuristics in adolescents than young adults, we found that 13-17-year-old adolescents applied plausibility cues similarly to young adults. Combined with the previous evidence on pre-adolescents and adolescents (Engelhardt, 2014; Joseph et al., 2008; Traxler, 2002), our results suggest that the ability to apply semantic heuristics matures by adolescence, by around the age of 15. Importantly, reliance on semantic heuristics increases as comprehenders become older. We found that older adults misinterpreted implausible sentences more often than young adults. This result was obtained for syntactically unambiguous sentences and is in line with previous experiments with locally ambiguous sentences (Christianson et al., 2006; Malyutina & Den Ouden, 2016). Taken together, these findings may signal a reorganisation in sentence processing strategies in older adults: "fast and frugal" heuristic processing may take over grammar-driven algorithmic processing (cf. "syntactic-to-semantic-processing strategy shift" suggested by Beese et al., 2019a).

The mechanisms driving the sentence processing strategy shift in older adults may be multifold. One possibility is that accumulation of world knowledge and language exposure in older adults (Ramscar et al., 2014) leads to stronger expectations about the message content. That is, older adults may impose more top-down predictions on a sentence representation. This account is in line with increased effects of predictability and context shown in older adults in eye-tracking-while-reading and self-paced-reading studies (Rayner et al., 2006; Zhao et al., 2020). Another possible mechanism is that reliance on heuristic processing is an unconscious strategy allowing to spare older adults' limited cognitive resources for other cognitive needs (Hess, 2000; Peeters et al., 2011), such as verbal working memory (Beese et al., 2019a, 2019b).

DOI: 10.1177/17470218211053263

Structural heuristics

Our second goal was to test whether reliance on structural heuristics would also increase with age. We found that adolescents did not use structural heuristics and interpreted sentences regardless of the attachment site of the participle, whereas young and older adults used the high-attachment heuristic and systematically misinterpreted unambiguous sentences with low attachment. Although reliance on structural heuristics was found in English-speaking pre-adolescents and adolescents (the heuristic of minimal attachment, Traxler, 2002; Engelhardt, 2014), our data showed no evidence that Russian-speaking adolescents applied the structural heuristics of high attachment.

One possible explanation for lack of the structural heuristic of high attachment in adolescents is that in different languages, structural heuristics differ in strength or mature at different stages of language development. Young Russian speakers might need more language experience than they usually have by the age of 15 to develop a preference for high attachment structures. Possibly, in real-life communication, semantic heuristics are applied more often than structural heuristics: the sentence meaning can often be inferred from context, whereas sentences in which one can apply structural heuristics (such as the heuristic of high attachment) are less common in everyday language input. Another possible explanation for why adolescents relied on algorithmic syntactic processing more than adults is the grammatical training that adolescents receive at school during Russian classes. As part of a typical curriculum, adolescents are trained to pay attention to morphological markers and agreement errors in printed texts (during reading and writing), whereas adults do not undergo such training on a regular basis. The design of our study does not allow us to distinguish between heuristic maturation and school training explanations. An experiment with auditory sentence

DOI: 10.1177/17470218211053263

presentation could help to differentiate between these possibilities, as adolescents receive no explicit training in detecting agreement errors in speech.

Relationship between processing speed and accuracy

Our final goal was to evaluate the key prediction of the good-enough processing account: namely, that semantic heuristic processing is faster than algorithmic processing. We found that, indeed, when participants read sentences faster, their accuracy decreased. However, this association alone is not sufficient to make conclusions about the speed of good-enough processing: application of heuristics should affect plausible and implausible conditions differently. This differential effect was also confirmed: implausible participle was read more slowly in the correct than in the incorrect trials. Therefore, our findings are fully compatible with one of the key assumptions behind the online equilibrium hypothesis of the good-enough approach to comprehension: top-down semantic heuristic processing should be faster than the more time-consuming bottom-up algorithmic processing (Karimi & Ferreira, 2016).

Interestingly, structural heuristic processing does not seem to be faster than algorithmic processing — we observed no slowdown in the low-attachment conditions in word-by-word reading times, and no interaction between average reading times and attachment in their effect on comprehension accuracy. On the one hand, this interaction is not necessarily predicted by the online equilibrium hypothesis of the good-enough theory, which is mostly concerned with semantic heuristics. On the other hand, if all heuristics are based on the same underlying principle, such as, for example, *apply the most plausible parse/interpretation and sustain it unless you encounter enough conflicting evidence to warrant reanalysis*, we would expect that in the dispreferred low-attachment conditions, faster reading times would be associated with an increase in incorrect responses. Similar associations between faster response times and decrease in response accuracy were found in processing of ungrammatical sentences with

DOI: 10.1177/17470218211053263

agreement attraction errors (among others, Avetisyan et al., 2020; Schlueter et al., 2019; Staub, 2009) and in the so-called "depth-charge" sentences ("No head injury it too trivial to be ignored"; Paape et al., 2020). One possible explanation for the absence of the expected interaction in our data is that the structural heuristic of high attachment is much weaker than the plausibility heuristic, and we may not have had enough data to detect a modest effect in the application of this heuristic.

Alternatives to good-enough processing

Until now, we considered the decrease in accuracy in implausible sentences and the speed-up in reading times associated with it from the perspective of the good-enough processing account. While it is indeed consistent with the predictions of the online equilibrium hypothesis of the good-enough processing theory, alternative explanations are also possible: one might consider the effect to be a slowdown for correct responses rather than a speed-up for incorrect responses. Correct responses in implausible sentences might take longer because the parser detects the semantic mismatch, and needs additional time to build a representation of an implausible event. For the incorrect responses, mismatch is not detected, and no additional time is used for recovery, therefore incorrect responses are given faster. Word-by-word reading times are consistent with this interpretation: the critical participle was read slower in implausible sentences, but only on the trials that had correct responses.

Whether such reanalysis interpretation is consistent with the online equilibrium hypothesis is an open question. On the one hand, the hypothesis states that the heuristic route finishes processing faster than the algorithmic route even when the algorithmic route encounters no conflicts (faster than the default). On the other hand, the hypothesis seems to also be consistent with the view that algorithmic processing is slowed down with respect to a hypothetical "no conflict" situation (slower than the default): "if there is enough evidence from

DOI: 10.1177/17470218211053263

the output of the algorithms to revise the heuristics-based interim output, the system will fall back to disequilibrium and will therefore allocate more processing resources to restore equilibrium" (p. 7, Karimi and Ferreira, 2016). Unfortunately, our data set cannot distinguish between these two possibilities because we have no neutral condition that is not affected by the plausibility heuristics. Comparing conditions where plausibility is manipulated to a baseline for which plausibility heuristics do not apply (for example, "Rimma helped Jane$_i$ who$_i$ was babbling / published ...") might turn out to be a fruitful avenue for future research.

Another alternative explanation for the emergence of misinterpretation errors suggests that misinterpretations arise at the question response stage after the sentence had been processed, rather than due to heuristics influencing the parsing of the sentence per se. In the implausible conditions in our stimuli, the syntactically correct sentence representation was not semantically sensible. Thus, incorrect question responses could be driven by the comprehenders' post-hoc strategy to choose a response that would be more sensible in real-world communication, rather than reflect their parsing. Some of the previous evidence that drove the good-enough processing theory can also be questioned on whether the comprehenders' behavior reflects their actual sentence processing or strategic priorities in responses. For example, similarly to our stimuli, the correct interpretation of the classic stimulus sentence '*The dog was bitten by the man*' (Ferreira, 2003) is also semantically implausible (cf. garden-path sentences like '*While Anna dressed the baby spit up on the bed*', where the syntactically correct interpretation is also semantically plausible).

Addressing the issue above, Bader and Meng (2018) showed that misinterpretations of implausible passives such as the ones used in Ferreira (2003) are observed only when participants have to answer comprehension questions requiring them to identify the agent or the patient of the action, but not when they assess sentence plausibility. Furthermore, Meng and Bader (2021) showed that participants successfully identify sentences as implausible on

DOI: 10.1177/17470218211053263

the same trials where they misidentify the agent/patient of the action. Is it possible that our participants do not misinterpret the sentences but give priority to sensible meanings at the question answering stage, as suggested by Meng and Bader (2021)? Although we lack direct evidence from plausibility ratings, we still think that it is unlikely: if incorrect responses were solely due to difficulties arising at the question response stage, and participants always built accurate representations during parsing, we would not expect reading times patterns to differ between correct and incorrect trials. In our study, implausible participles were read slower on correct trials, which speaks in favor of correct representation being built during parsing and later reflected at the question response stage.

Limitations and conclusions

Among the limitations of our study, we acknowledge that the preference for high attachment in Russian, utilised in our study as a proxy of top-down structural heuristic, may not always be driven by purely structural mechanisms. Alternatively, attachment preferences may be driven by discourse prominence, specific semantic relations, or implicit prosody of the language (Başer & Hohenberger, 2020, Fodor, 2002, Gilboy et al., 1995). But even though these non-structural factors may play a role in the preference for high attachment, we believe that these heuristics still pertains to top-down preferences in structure-building and thus demonstrates a different kind of top-down heuristic processing than what was measured by semantic plausibility heuristics in our study.

Further, reliance on semantic and structural heuristics may depend on the specific sentence construction we used. Furthermore, our accuracy measures were obtained in the experimental design with two-alternative comprehension questions, meaning that participants were forced to make a choice between only two responses. We cannot be sure that the mental representations of the sentences could be fully mapped onto one of the two responses suggested

DOI: 10.1177/17470218211053263

for the participants. One possible exception is that some participants might have blended representations (e.g. Malyutina & Den Ouden, 2016; Slattery et al., 2013) that cannot be captured in our experimental design. Oral responses would be more ecologically valid and would allow the researchers to investigate heuristic processing in more detail. All these limitations provide exciting opportunities for future research.

To conclude, top-down heuristic processing helps comprehenders to support rapid communication with as little effort as possible, but can sometimes result in misunderstanding. Our study shows that these effective, albeit risky, processing mechanisms appear already in adolescence and then keep maturing across the adult lifespan, via emerging reliance on structural heuristics and increasing reliance on semantic heuristics.

Acknowledgements

We thank Olga Tuzhik, Nikita Beklemishev, Alina Lobanova, Zoya Evdaeva, Anna Mokhova, and Aleksandra Savcenko for their help with data collection.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

Funding

This work was funded by the Russian Foundation for Basic Research, Project No 18-012-00640.

DOI: 10.1177/17470218211053263

Data accessibility statement

The data and materials for the experiment are available at https://osf.io/4f2px/. The experiment was preregistered as part of a project investigating good-enough language processing across the lifespan under no-noise and linguistic-noise conditions (https://osf.io/k3ebs).

References

Alcock, K. J., & Ngorosho, D. (2004). Interaction between phonological and grammatical processing in single word production in Kiswahili. *Language and speech, 47*(1), 1-30. doi.org/10.1177/00238309040470010101

Amichetti, N. M., White, A. G., & Wingfield, A. (2016). Multiple solutions to the same problem: Utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Frontiers in Psychology, 7*, 789. doi:10.3389/fpsyg.2016.00789

Avetisyan, S., Lago, S., & Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension?. *Journal of Memory and Language*, *112*, 104087. doi:10.1016/j.jml.2020.104087

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(8), 1286–1311. doi.org/10.1037/xlm0000519

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition, 21(4)*, 477–487. doi:10.3758/BF03197179

Başer, Z., & Hohenberger, A. (2020). Is there a particular RC attachment preference in Turkish? Negotiating the effects of semantic factors. *Journal of Psycholinguistic Research*. doi:10.1007/s10936-020-09698-4

DOI: 10.1177/17470218211053263

Beese, C., Werkle-Bergner, M., Lindenberger, U., Friederici, A.D., & Meyer, L. (2019a). Adult age differences in the benefit of syntactic and semantic constraints for sentence processing. *Psychology and Aging*, 34(1), 43-55. doi:10.1037/pag0000300

Beese, C., Vassileiou, B., Friederici, A. D., & Meyer, L. (2019b). Age differences in encoding-related alpha power reflect sentence comprehension difficulties. *Frontiers in Aging Neuroscience, 11*, 183. https://doi.org/10.3389/fnagi.2019.00183

Bürkner, P. C. (2017). Advanced Bayesian multilevel modeling with the R package brms. *arXiv preprint arXiv:1705.11123* / Accessed 14 December 2020.

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

Chernova, D., & Chernigovskaya, T. V. (2015). Syntactic Ambiguity Resolution in Sentence Processing: New Evidence from a Morphologically Rich Language. In *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science*, 129-133.

Chernova, D.A. (2016). *The processing syntactically ambiguous sentences: Psycholinguistic research*. Doctoral dissertation, Saint Petersburg State University. (In Russian). https://disser.spbu.ru/dis-list/details/14/949.html / Accessed 27 July 2020.

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

Christianson, K., Williams, C., Zacks, R., & Ferreira, F. (2006). Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes, 42*, 205–238. doi:10.1207/s15326950dp4202_6

DOI: 10.1177/17470218211053263

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*, 538-544. doi: 10.1037/a0018027

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology, 69(5)*, 817–828. doi:10.1080/17470218.2015.1134603

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27 (1)*, 3–42. doi:10.1017/S0142716406060024

Engelhardt, P. E. (2014). Children's and adolescents' processing of temporary syntactic ambiguity: An eye movement study. *Child Development Research, 2014, 475315*, 1–13. doi:10.1155/2014/475315

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540–551. doi:10.1016/S0022-5371(81)90165-1

Fedorova, O.V., & Yanovich I.S. (2006). Early preferences in relative clause attachment in Russian: The effect of working memory differences. In *Proceedings of the 14th annual workshop on formal approaches to Slavic linguistics. The Princeton meeting*. Ann Arbor.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: a study of relative clause attachment. *Language Acquisition, 11*, 127–163. doi:10.1207/s15327817la1103_1

Ferreira, F., Bailey, K.G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11(1)*, 11–15. doi:10.1111/1467-8721.00158

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

DOI: 10.1177/17470218211053263

Ferreira, F., & Yang, Z. (2019). The Problem of Comprehension in Psycholinguistics. *Discourse Processes, 56(7)*, 485-495. doi.org/10.1080/0163853X.2019.1591885

Fillenbaum, S. (1971). Processing and recall of compatible and incompatible question and answer pairs. *Language and Speech, 14(3)*, 256–265. doi:10.1177/002383097101400306

Fillenbaum, S. (1974). Pragmatic normalization: Further results for some conjunctive and disjunctive sentences. *Journal of Experimental Psychology, 102(4)*, 574–578. doi:10.1037/h0036092

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani, (Ed.), Proceedings of NELS 32 (pp. 113-132). Amherst: GLSA Publications.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.

Gilboy, E., Sopena, J.-Mm., Cliftrn, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition, 54*(2), 131–167. doi:10.1016/0010-0277(94)00636-y

Joseph, H. S. S.L., Liversedge, S.P., Blythe, H.I., White, S.J., Gathercole, S.E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 61(5)*, 708–723. doi:10.1080/17470210701400657

Hess, T. M. (2000). Aging-related constraints and adaptations in social information processing. In U. Von Hecker, S. Dutke, & G. Sedek (Eds.), *Generative mental processes and cognitive resources: Integrative research on adaptation and control* (pp. 129–155). Dordrecht: Kluwer.

Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: response bias not sensitivity. *Memory and Cognition, 24*, 687–699.

DOI: 10.1177/17470218211053263

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology, 69(5)*, 1013–1040. doi:10.1080/17470218.2015.1053951

Kay, M. (2018). *tidybayes: Tidy data and geoms for Bayesian models*. R package version, 1(3). https://mjskay.github.io/tidybayes/ Accessed 14 December 2020.

Kjelgaard, M. M., Titone, D. A., & Wingfield, A. (1999). The influence of prosodic structure on the interpretation of temporary syntactic ambiguity by young and elderly listeners. *Experimental Aging Research, 25*(3), 187-207. doi:10.1080/036107399243986

Koornneef, A., & Reuland, E. (2016). On the shallow processing (dis)advantage: Grammar and economy. *Frontiers in Psychology, 7(82)*, 1–17. doi:10.3389/fpsyg.2016.00082

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research, 1146*, 23–49. doi:10.1016/j.brainres.2006.12.063

Malyutina, S., & Den Ouden, D. (2016). What is it that lingers? Garden-path (mis)interpretations in younger and older adults. *The Quarterly Journal of Experimental Psychology*, *69*(5), 880-906. doi:10.1080/17470218.2015.1045530

Malyutina, S., Laurinavichyute, A., Terekhina, M., & Lapin, Y. (2018). No evidence for strategic nature of age-related slowing in sentence processing. *Psychology and Aging, 33(*7), 1045-1059. doi:10.1037/pag0000302

Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences?. *Quarterly Journal of Experimental Psychology, 74*(1), 1-28. doi.org/10.1177/1747021820947940

Obler, L. K., Fein, D., Nicholas, M., & Albert, M. L. (1991). Auditory comprehension and aging: Decline in syntactic processing. *Applied Psycholinguistics, 12*(4), 433-452. doi:10.1017/S0142716400005865

DOI: 10.1177/17470218211053263

Paape, D., Vasishth, S., & von der Malsburg, T. (2020). Quadruplex negatio invertit? The on-line processing of depth charge sentences. *Journal of Semantics, 37(4)*, 509-555.

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., Stine-Morrow, E. A. (2014). Aging and individual differences in binding during sentence understanding: evidence from temporary and global syntactic attachment ambiguities. *Cognition, 130*(2), 157-173. doi:10.1016/j.cognition.2013.10.005

Peters, E., Dieckmann, N.F., & Weller, J.A. (2011). Age differences in complex decision making. In K. W. Schaie & S. L. Willis (Eds.), *Handbook of the psychology of aging* (7th ed., pp. 133–151). Burlington, MA: Academic Press. doi:10.1016/B978-0-12-380882-0.00009-7

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www. R-project.org / Accessed 14 December 2020.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6*(1), 5-42. doi:10.1111/tops.12078

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging, 21*, 448–465. doi:10.1037/0882-7974.21.3.448

Schlueter, Z., Parker, D., & Lau, E. (2019). Error-driven Retrieval in Agreement Attraction rarely leads to Misinterpretation. *Frontiers in psychology*, *10*, 1002. doi:10.3389/fpsyg.2019.01002

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

DOI: 10.1177/17470218211053263

Sekerina, I. (1997). The late closure principle in processing of ambiguous Russian sentences. In *Current Approaches to Formal Slavic Linguistics. Contributions of the Second European Conference on Formal Description of Slavic Languages FDSL II* (pp. 205-217). Potsdam.

Sekerina, I., Petrova, K., & Fernández, E. (2003). Relative clause attachment in Bulgarian. In *Proceedings of the 12th annual workshop on formal approaches to Slavic linguistics. The Ottawa meeting* (pp. 375-394). Ottawa.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*(2), 104-120. doi: 10.1016/j.jml.2013.04.001

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of memory and language*, *60*(2), 308-327. doi:10.1016/j.jml.2008.11.002

Stine-Morrow, E. A., Ryan, S., & Leonard, J. S. (2000). Age differences in on-line syntactic processing. *Experimental Aging Research, 26*(4), 315–322. doi:10.1080/036107300750015714

Stoops, A., Luke, S. G., & Christianson, K. (2014). Animacy information outweighs morphological cues in Russian. *Language, Cognition and Neuroscience, 29*(5), 584-604, doi: 10.1080/01690965.2013.813560

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50*(4), 355-370. doi: 10.1016/j.jml.2004.01.001

Titone, D. A., Koh, C. K., Kjelgaard, M. M., Bruce, S., Speer, S. R., & Wingfield, A. (2006). Age-related impairments in the revision of syntactic misanalyses: Effects of prosody. *Language and Speech, 49*(1), 75–99. doi:10.1177/00238309060490010501

DOI: 10.1177/17470218211053263

Traxler, M.J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology: Section A, 55(1)*, 75–96. doi:10.1080/02724980143000172

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73 (2)*, 89–134. doi:10.1016/s0010-0277(99)00032-3

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging, 18*(2), 332-9. doi:10.1037/0882-7974.18.2.332

Wellwood, A., Pancheva, R., Hacquard, V., & Phillips, C. (2018). The anatomy of a comparative illusion. *Journal of Semantics, 35(3)*, 543–583. doi:10.1093/jos/ffy014

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Cham: Springer International Publishing. doi:10.1007/978-3-319-24277-4

Yoon, J., Campanelli, L., Goral, M., Marton, K., Eichorn N., & Obler, L. K. (2015). The effect of plausibility on sentence comprehension among older adults and its relation to cognitive functions. *Experimental Aging Research, 41*(3), 272-302. doi:10.1080/0361073X.2015.1021646

Zhao, S., Li, L., Chang, M., Wang, J., & Paterson, K.B. (2020). A further look at ageing and word predictability effects in Chinese reading: Evidence from one-character words. *Quarterly Journal of Experimental Psychology*. doi:10.1177/1747021820951131

DOI: 10.1177/17470218211053263

Figure Captions

Figure 1. Geometric mean reading times (back-transformed) across experimental conditions and sentence regions, additionally grouped by trial accuracy.

Figure 2. Estimated accuracy (means and 95% credible intervals) in the four experimental conditions for adolescents, young and older adults.

Figure 3. Estimated interaction between age, plausibility, and reading time (plotted as a factor with three levels corresponding to mean ±1SD). Slower reading times are associated with an additional increase in estimated accuracy.

DOI: 10.1177/17470218211053263

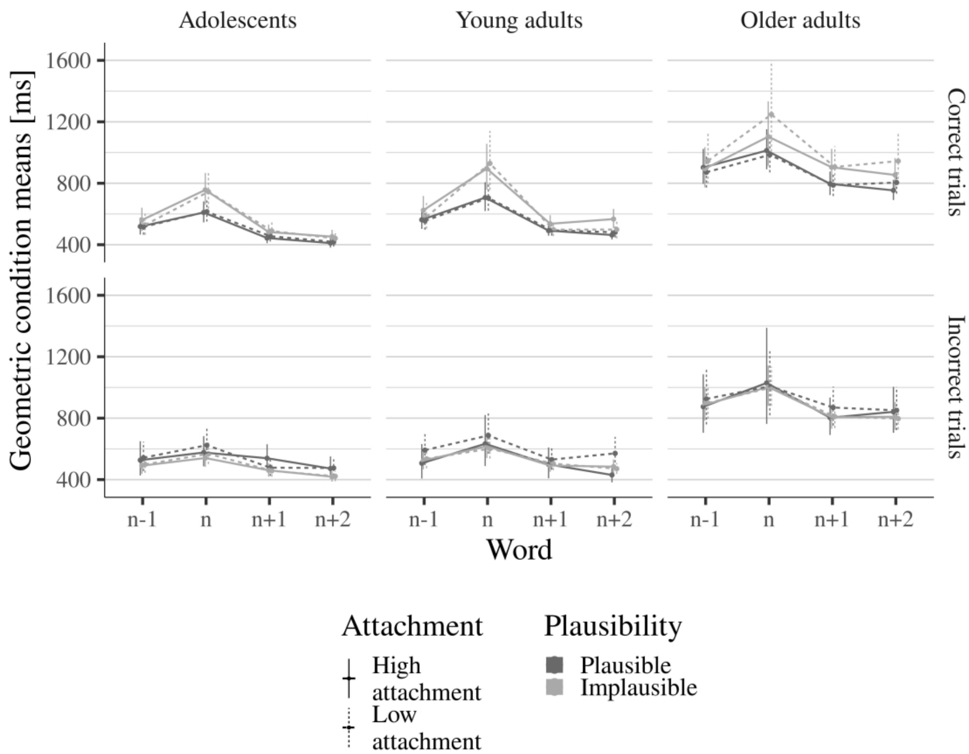

Figure 1. Geometric mean reading times (back-transformed) across experimental conditions and sentence regions, additionally grouped by trial accuracy.

381x304mm (300 x 300 DPI)

DOI: 10.1177/17470218211053263

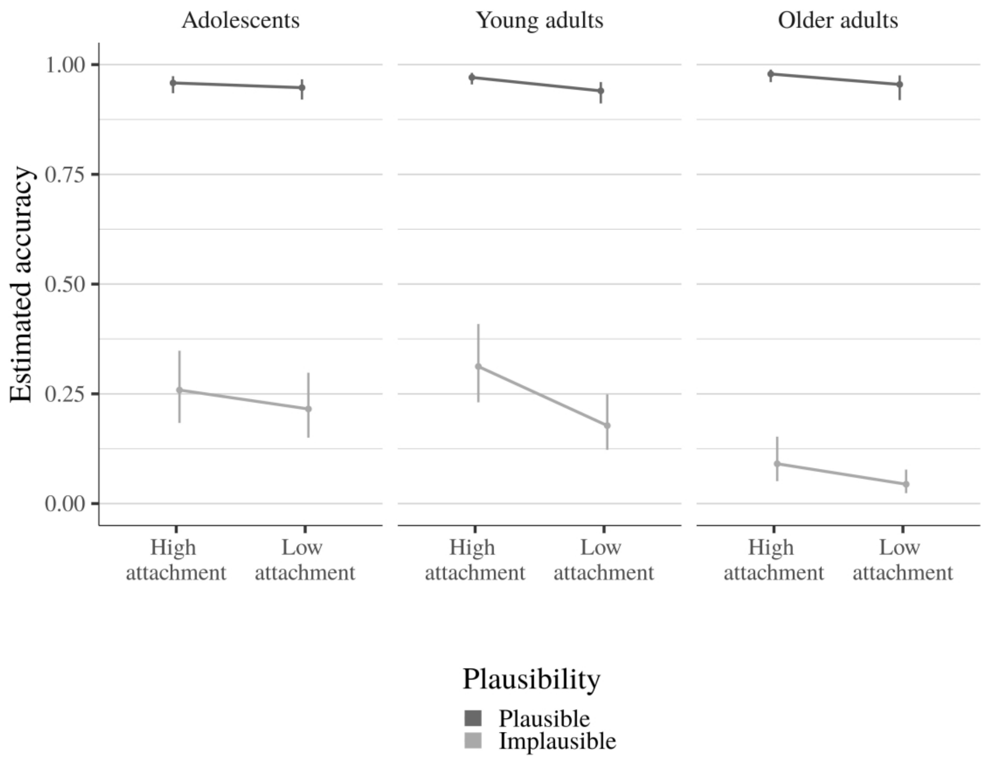

Figure 2. Estimated accuracy (means and 95% credible intervals) in the four experimental conditions for adolescents, young and older adults.

381x304mm (300 x 300 DPI)

DOI: 10.1177/17470218211053263

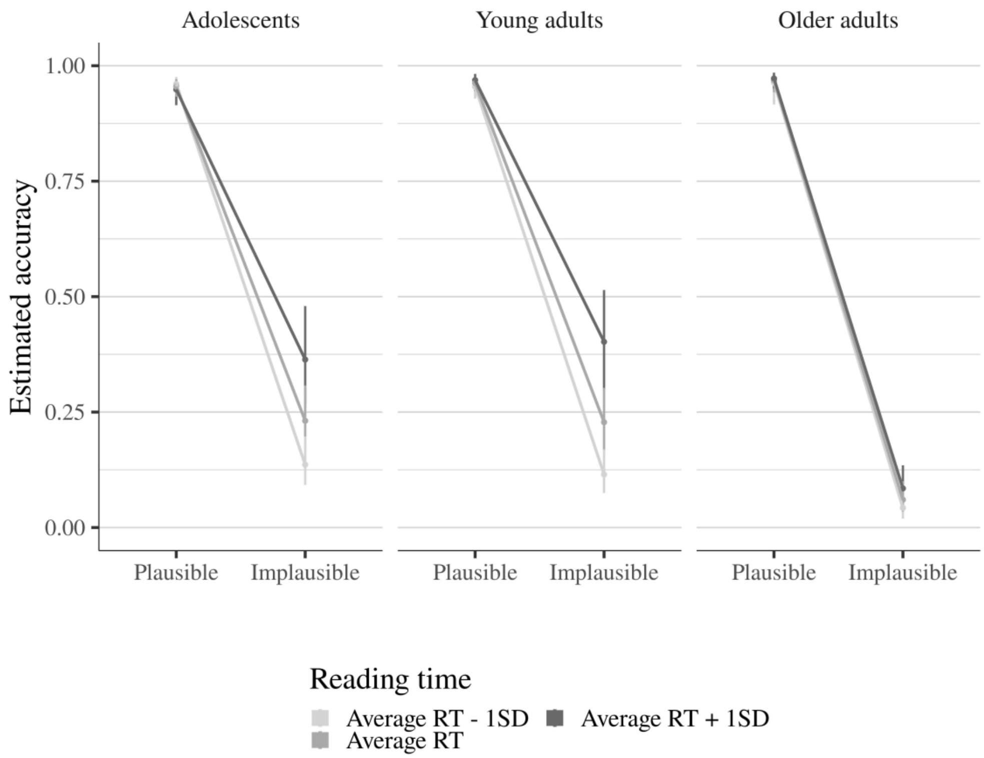

Figure 3. Estimated interaction between age, plausibility, and reading time (plotted as a factor with three levels corresponding to mean ±1SD). Slower reading times are associated with an additional increase in estimated accuracy.

381x304mm (300 x 300 DPI)

DOI: 10.1177/17470218211053263

Table 1. Experimental set of stimuli and fillers

(A) High attachment, plausible	*Римма одевала* ***малышк-у писательниц-ы, лепеч-ущ-ую*** *непонятные слова.* **child-ACC writer-GEN babble-PP-ACC** Rimma dressed **the child$_i$ of the writer who$_i$ was babbling** incomprehensible words.
Comprehension question	*Кто лепетал непонятные слова?* Who was babbling incomprehensible words? ***Малышка*** */ Писательница* **Child** / Writer
(B) Low attachment, plausible	*Римма одевала* ***малышк-у писательниц-ы, опубликова-вш-ей*** *интересный роман.* **child-ACC writer-GEN publish-PP-GEN** Rimma dressed **the child of the writer$_i$ who$_i$ published** an interesting novel.
Comprehension question	*Кто опубликовал интересный роман?* Who published an interesting novel? *Малышка /* ***Писательница*** Child / **Writer**
(C) High attachment, implausible	*Римма одевала* ***малышк-у писательниц-ы, опубликова-вш-ую*** *интересный роман.* **child-ACC writer-GEN publish-PP-ACC** Rimma dressed **the child$_i$ of the writer who$_i$ published** an interesting novel.
Comprehension question	*Кто опубликовал интересный роман?* Who published an interesting novel? ***Малышка*** */ Писательница* **Child** / Writer
(D) Low attachment, implausible	*Римма одевала* ***малышк-у писательниц-ы, лепеч-ущ-ей*** *непонятные слова.* **child-ACC writer-GEN babble-PP-GEN** Rimma dressed **the child of the writer$_i$ who$_i$ was babbling** incomprehensible words.
Comprehension question	*Кто лепетал непонятные слова?* Who was babbling incomprehensible words? *Малышка /* ***Писательница*** Child / **Writer**

DOI: 10.1177/17470218211053263

Fillers	Dima deeply respected **the cook$_i$ of the old woman who$_i$ prepared** a three-course dinner in just an hour. Who respected the cook? Misha / **Dima** The countess admired **the guest's brother$_i$ who$_i$ played** the flute excellently. What did the guest's brother play? **The flute** / The violin The footballer scored two goals in a very important game, and his son was very proud of him. How many goals did the footballer score? **Two** / Three

DOI: 10.1177/17470218211053263

Table 2. Results of statistical analysis of reading times. Estimates are presented on the logarithmic scale. Estimates for which the probability that the true parameter value is greater than 0 exceeds 97.5% or is lower than 2.5% are in bold.

	Precritical region			**Critical region**			**Critical region + 1**			**Critical region + 2**		
Predictors	*Estimate*	*95% Credible interval*	*Prob (β>0)*	*Estimate*	*95% Credible interval*	*Prob (β>0)*	*Estimate*	*95% Credible interval*	*Prob (β>0)*	*Estimate*	*95% Credible interval*	*Prob (β>0)*
Intercept (young adults)	**6.31**	**6.20 – 6.42**	**1**	**6.54**	**6.42 – 6.66**	**1**	**6.20**	**6.13 – 6.28**	**1**	**6.20**	**6.12 – 6.27**	**1**
Adolescents	-0.08	-0.23 – 0.07	0.14	**-0.18**	**-0.34 – -0.01**	**0.02**	-0.06	-0.16 – 0.05	0.15	**-0.13**	**-0.23 – -0.02**	**0.01**
Older adults	**0.46**	**0.29 – 0.63**	**1**	**0.33**	**0.13 – 0.52**	**1.00**	**0.46**	**0.34 – 0.59**	**1**	**0.48**	**0.35 – 0.60**	**1**
Plausibility	0.02	-0.01 – 0.05	0.88	0.01	-0.03 – 0.05	0.64	0.01	-0.02 – 0.04	0.82	-0.02	-0.05 – 0.01	0.08
Attachment	0.00	-0.02 – 0.03	0.62	0.00	-0.02 – 0.03	0.58	0.01	-0.01 – 0.03	0.73	-0.01	-0.03 – 0.01	0.27
Accuracy	0.03	-0.01 – 0.08	0.95	**0.10**	**0.05 – 0.15**	**1**	-0.01	-0.04 – 0.03	0.39	-0.01	-0.04 – 0.03	0.38
Word length				**0.03**	**0.03 – 0.04**	**1**	**0.01**	**0.01 – 0.02**	**1**			
Word frequency				**-0.02**	**-0.02 – -0.01**	**0.001**						
Adolescents x plausibility	-0.02	-0.06 – 0.03	0.25	0.00	-0.05 – 0.06	0.55	-0.03	-0.07 – 0.01	0.07	0.00	-0.04 – 0.04	0.43
Older adults x plausibility	-0.02	-0.07 – 0.04	0.30	0.03	-0.04 – 0.09	0.81	0.00	-0.05 – 0.05	0.54	0.02	-0.03 – 0.07	0.73
Adolescents x attachment	-0.01	-0.04 – 0.02	0.33	0.03	-0.01 – 0.07	0.95	0.00	-0.03 – 0.03	0.51	0.01	-0.01 – 0.04	0.85
Older adults x attachment	0.01	-0.03 – 0.04	0.61	0.00	-0.04 – 0.04	0.54	0.00	-0.03 – 0.03	0.57	0.01	-0.03 – 0.04	0.65
Adolescents x accuracy	0.01	-0.05 – 0.06	0.59	0.04	-0.03 – 0.10	0.85	0.00	-0.05 – 0.05	0.54	-0.01	-0.06 – 0.04	0.31
Older adults x accuracy	-0.06	-0.13 – 0.02	0.06	-0.03	-0.11 – 0.06	0.28	0.03	-0.03 – 0.10	0.86	0.03	-0.03 – 0.10	0.85
Plausibility x accuracy	-0.01	-0.05 – 0.03	0.28	**0.07**	**0.03 – 0.12**	**0.99**	-0.01	-0.05 – 0.03	0.28	**0.06**	**0.02 – 0.10**	**1**
Attachment x accuracy	0.00	-0.03 – 0.03	0.56	0.00	-0.03 – 0.04	0.55	-0.01	-0.03 – 0.02	0.32	0.02	-0.01 – 0.04	0.91
Adolescents x plausibility x accuracy	0.02	-0.04 – 0.08	0.76	0.00	-0.07 – 0.07	0.51	**0.06**	**0.01 – 0.11**	**0.98**	-0.01	-0.06 – 0.05	0.41
Older adults x plausibility x accuracy	-0.02	-0.09 – 0.05	0.31	-0.07	-0.15 – 0.02	0.07	0.02	-0.04 – 0.09	0.74	0.00	-0.07 – 0.07	0.48
Adolescents x attachment x accuracy	-0.01	-0.04 – 0.03	0.39	-0.04	-0.09 – 0.00	0.03	0.01	-0.03 – 0.04	0.69	-0.02	-0.06 – 0.02	0.13

DOI: 10.1177/17470218211053263

Older adults x attachment x accuracy	-0.02	-0.07 – 0.02	0.16	-0.02	-0.07 – 0.04	0.26	-0.01	-0.05 – 0.03	0.29	0.02	-0.03 – 0.06	0.75

DOI: 10.1177/17470218211053263

Table 3. Response accuracy for adolescents, young and older adults in the four experimental conditions.

	Adolescents, mean (SD)	Young adults, mean (SD)	Older adults, mean (SD)
High attachment, plausible	0.89 (0.18)	0.94 (0.10)	0.93 (0.14)
Low attachment, plausible	0.87 (0.18)	0.86 (0.20)	0.87 (0.18)
High attachment, implausible	0.32 (0.31)	0.37 (0.36)	0.21 (0.24)
Low attachment, implausible	0.30 (0.34)	0.29 (0.36)	0.13 (0.21)

DOI: 10.1177/17470218211053263

Table 4. Results of statistical analysis of accuracy. Estimates are presented on the log-odds scale. Estimates for which the probability that the true parameter value is greater than 0 exceeds 97.5% or is lower than 2.5% are in bold.

Coefficient	Estimate (on the log-odds scale)	95% Credible interval	Prob (β > 0)
Intercept (young adults, plausible + high attachment)	**1.01**	**0.76 – 1.27**	**~1**
Adolescents	-0.09	-0.40 – 0.23	0.30
Older adults	**-0.68**	**-1.07 – -0.27**	**0.0007**
Implausible	**-2.23**	**-2.55 – -1.92**	**0**
Low attachment	**-0.38**	**-0.60 – -0.17**	**0.0005**
Mean RT	**0.51**	**0.33 – 0.69**	**1**
Adolescents × implausible	0.10	-0.30 – 0.49	0.69
Older adults × implausible	**-0.86**	**-1.37 – -0.33**	**0.001**
Adolescents × low attachment	**0.26**	**0.02 – 0.50**	**0.98**
Older adults × low attachment	0.00	-0.32 – 0.32	0.51
Adolescents × mean RT	-0.26	-0.51 – 0.00	0.03
Older adults × mean RT	-0.26	-0.59 – 0.08	0.07

DOI: 10.1177/17470218211053263

Implausible × mean RT	**0.31**	**0.11 – 0.53**	**0.99**
Attachment × mean RT	-0.01	-0.12 – 0.10	0.43
Adolescents × implausible × mean RT	0.08	-0.20 – 0.37	0.70
Older adults × implausible × mean RT	-0.20	-0.57 – 0.17	0.14

DOI: 10.1177/17470218211053263