



## OPEN ACCESS

## EDITED BY

Matthew W. Crocker,  
Saarland University, Germany

## REVIEWED BY

Katja Haeuser,  
Saarland University, Germany

## \*CORRESPONDENCE

Julia Edeleva  
✉ [uk107425@uni-kassel.de](mailto:uk107425@uni-kassel.de)

RECEIVED 10 October 2025

REVISED 10 December 2025

ACCEPTED 22 December 2025

PUBLISHED 26 January 2026

## CITATION

Edeleva J (2026) Beyond error-driven adaptation: proactive validation as a goal-directed mechanism of language processing and learning.

*Front. Lang. Sci.* 4:1722280.

doi: 10.3389/flang.2025.1722280

## COPYRIGHT

© 2026 Edeleva. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Beyond error-driven adaptation: proactive validation as a goal-directed mechanism of language processing and learning

Julia Edeleva\*

Department of Research on Foreign Language Learning and Teaching, University of Kassel, Kassel, Germany

Recent advances in psycholinguistics increasingly frame language processing as a predictive process: listeners and readers continuously anticipate upcoming linguistic input. Deviations from those expectations—prediction errors—are assumed to stimulate both moment-by-moment processing and long-term learning. While highly influential, this view implicitly assumes that the adaptation is driven by discrepancies. Such an approach overlooks a crucial aspect of rational human behavior: Rational agents generally act to avoid failure, not to repeatedly learn from it. In the current perspective paper, I review the evidence for prediction error minimization as a proactive, preemptive (rather than reactive repair) mechanism of language processing. Rather than reacting after a mismatch, language users will accumulate evidence by maintaining and validating less probable parses to reduce a risk of failure. By proposing proactive validation as a proactive, goal-directed mechanism, the paper seeks to complement rational models of predictive processing by shifting the temporal and mechanistic focus from *post-hoc* long-term statistical adjustment to anticipatory optimization. This framework provides a unified explanation for cross-linguistic and developmental variability in processing difficulty, such as reduced processing cost for non-canonical structures in morphologically rich languages and the gradual shift from reactive to proactive strategies in language learners. On a broader scale, anticipatory validation can explain why the comprehension system tolerates ambiguity and maintains suboptimal parses—not to correct errors retrospectively, but to pre-empt them prospectively.

## KEYWORDS

language processing, parsing, prediction, prediction error, proactive validation, relative clauses

## Introduction

One of the well-established tenets in psycholinguistics is that language processing is incremental. Language users interpret the input as it arrives rather than wait until they can commit themselves to a fully-fledged structure. The incremental nature of language comprehension creates persistent uncertainty regarding the upcoming linguistic input. While the input unfolds, multiple structural continuations remain viable. This uncertainty emerges from the constant interaction of top-down expectations and bottom-up evidence. Top-down expectations arise from prior linguistic experience (e.g., frequency-based preferences, canonical word order patterns), domain-general knowledge (e.g., event schemas, typical agent-patient relations), and a broader discourse context. Bottom-up information, by contrast, is extracted directly from the incoming signal (e.g., morphological cues, such as case-marking and agreement; lexical constraints,

such as subcategorization frames; and positional constraints, such as word order or prosodic boundaries). This information incrementally confirms, refines or disconfirms top-down priors.

The interaction of top-down and bottom-up processes becomes evident in temporarily ambiguous sentences such as *the boy that the girl kissed*. At the relative pronoun *that*, top-down structural expectations and canonical word order biases may support an initial subject-relative analysis (i.e. the boy is the agent of kissing, cf. the syntactic alternative *the boy that kissed the girl*). However, bottom-up cues introduced later (i.e., *the girl* before the verb *kissed*) validate the target object-relative analysis (i.e., the boy is the patient of kissing).

Thus, predictive language processing—an idea that parallels predictive coding models in neuroscience (Friston, 2010; Rao and Ballard, 1999; Kuperberg and Jaeger, 2016)—suggests that language users actively predict the upcoming linguistic structure rather than purely align their ongoing analysis with the input. Predictive mechanisms using probabilistic cues such as lexical preferences, case marking, and word order make it possible to navigate uncertainty and anticipate likely continuations. Object relative clauses (ORC) generally incur a greater processing cost compared to their subject-relative counterparts (SRC; Lau and Tanaka, 2021 for an overview). Critically, however, empirical evidence shows that this cost is graded and cue-dependent: ORCs become less cost-intensive if the head NP is inanimate (cf. *the ball that the boy kicked*, Gennari and MacDonald, 2008, 2009) or when the embedded NP has been introduced in the previous context (Gordon et al., 2001; Mak et al., 2008).

These and related findings (Federmeier and Kutas, 1999; Häuser et al., 2022; Häuser and Borovsky, 2025) indicate that comprehenders do not commit to one default analysis upfront; rather they pre-activate multiple candidate representations proportionate to their probability. These empirical patterns directly motivate computational approaches that represent sentence continuations as probabilistically weighted alternatives (MacDonald et al., 1994; Spivey-Knowlton et al., 1993; Traxler et al., 2002). Standard probabilistic models assume that the alternatives are activated in parallel and continuously re-ranked regarding their weights as input unfolds (though see Fujita, 2023 for attempts to implement prediction into a non-probabilistic reanalysis model). A less expected parse receives a lower activation weight and will need to be promoted as soon as a more expected parse fails.

Within such probabilistic frameworks,<sup>1</sup> prediction error naturally arises as a complementary mechanism: whenever

predicted input diverges from the actual signal, the system updates its expectations in an effort to statistically fine-tune and minimize errors in the future. Under this view, the system strategically reaches long-term adaptation *after* encountering unexpected input. It thus would imply that the processing system must repeatedly engage in inefficient, “error-prone” behavior to adjust top-down expectations. Thus, an important conceptual question arises: *Does sentence processing primarily operate as a reactive error-correction system, where structural interpretations are adjusted in response to prediction violations, or as a bounded-rational, anticipatory system that accumulates available support for plausible parses to avoid the risk of error?*

While computational models (Dell and Chang, 2014) have largely treated prediction errors as reactive signals, I will argue that it is not prediction error *per se*, but moment-by-moment proactive minimization of risk of failure that provides a more comprehensive account of sentence processing. Rather than reacting to prediction errors as a mechanism behind long-term statistical updates over repeated encounters, the system would allocate available resources to minimize the risk of error before it occurs. This is achieved as the system maintains and validates alternative (less probable) parses as a strategy to ensure stable interpretation.

## Prediction error in language processing

In language processing, prediction errors represent discrepancies between the expected and the observed input. Language users are hypothesized to be initially guided by “fast-and-frugal” heuristics (e.g., the agent-first preference) or more available parses that later clash against subsequent input and engender misinterpretation. The moment prediction error is detected, initial analysis will be repaired, and the top-level probabilities updated, incurring additional processing cost. Some computational approaches (Chang et al., 2006; Dell and Chang, 2014) treat prediction error as an essential and ubiquitous mechanism for learning. Over repeated encounters, multiple mismatches and moment-by-moment adjustments may allow language learners to update their mental representations, refine their linguistic expectations and gradually converge on the target language system.

In L2 learners, predictive processing is subject to multiple sources of variation (Kaan and Grüter, 2021) which sparks ongoing empirical investigations into the role of prediction error as an implicit driver of language learning (cf. Kaan and Grüter, 2021). Eye-tracking and ERP studies in adult L2 learners show that when incoming input violates more available L1-based structural expectations, processing delays and neural signatures such as the N400 are observed as correlates of prediction failure and the ensuing adjustment (Safak and Hopp, 2023, 2025; Bovolenta and Marsden, 2022). Similar effects have been documented in children acquiring their first language. Discrepancies between the expected and the observed input guide attention redirection, reanalysis, and representational updating (Reuter et al., 2019; Gambi and Messenger, 2023).

While these findings support the implicit corrective function of prediction errors, they raise a theoretical tension: *If a rational system is designed to minimize discrepancies due to the risk of failure*

<sup>1</sup> Other computational approaches, such as cue-based retrieval models (Lewis and Vasishth, 2005), abstract away from the assumption that an entire representation becomes fully activated. They conceptualize sentence processing as a series of activation states for potential referents and share the concept of weak priors which is also critical for the current paper. Each referent is encoded with specific features and become cue-addressable in memory. When a target needs to be selected, potential referents are retrieved based on these features. The associated processing cost corresponds to the activation effort for a specific referent. While these models successfully simulate observed processing costs, they don't make any assumption about predictive mechanisms. They therefore fall beyond the scope of the current contribution, focusing on predictive and goal-directed frameworks.

([Crocker, 2005](#)), why would it rely on them as a learning signal that stimulates processing? In domains outside language, particularly in decision-making, agents are assumed to act rationally under uncertainty. They are not driven by failure, but by the minimization of expected loss ([Borovcnik, 2015](#)). They continuously weigh up possible actions (also less likely ones) against their costs and benefits. In doing so, they aim to actively reduce the risk of failure rather than induce repairs.

Applying this logic to language comprehension, each “error”, even when it happens only once, carries a cost: the processing effort required to revise the ongoing linguistic analysis and the risk of misinterpretation. Permanent “error-prone” behavior would contradict rational efficiency in a system like language which is evolutionarily optimized for real-time communication where miscomprehension is penalized. Instead, rational agents could use all the available resources to anticipate the most likely outcome and gatekeep potential errors.

## Proactive validation

As previously discussed, language processing operates under conditions of ambiguity where early commitment to a single, highly probable parse can increase the risk of future repair (cf. [Häuser et al., 2022](#); [Häuser and Borovsky, 2025](#) for recent empirical evidence in favor of multiple representations to be activated and maintained in parallel). Some approaches ([Crocker, 2005](#)) offer a rational solution by proposing that the parser occasionally favors a globally less probable, but a more “specific” analysis. Error minimization is achieved indirectly as the parser locally reduces the cost of misanalysis: it can easily drop a less probable parse and shift to a more probable one. While current approaches capture the efficiency of adaptive parsing, they still frame rationality as a reactive process as errors are inevitable. The system’s goal is, thus, to engage in developing efficient error recovery routines.

A different question could be asked: *Is there a mechanism that the system uses to gatekeep errors in the first place?*<sup>2</sup> As soon as it encounters some cues in a conflicting environment (e.g., sentence-initial conflict of case-marking and word order in morphologically-rich languages), the parser activates and maintains less probable parses in a low-weighted state. These initially less probable alternatives are temporarily kept and subsequently validated once supporting information becomes available, rather than generated through ongoing reranking (or reanalysis).

In a way, the two complementary processes can be illustrated by the following analogy: Imagine a familiar road which you take home that is under construction. On a typical day, you would go straight, and you have been told that the construction is finished, while a road sign indicates a bend ahead. A **reactive strategy** would rely on previous experience (though keeping in mind the road sign) and go straight which will result in a sudden “slam on the brakes” when the bend appears. One could also strongly commit to the bend as a less probable, but highly specific route. If the construction is actually finished and the road is clear, but the workers just forgot to remove

the sign, a “slam on the brakes” would still take place though going straight is less taxing than bending. A **proactive validation strategy** would instead slow down and pay attention before the bend and maintain alternative (less preferred) paths in mind, weighing the conflicting information from the announcement and the sign. As the bend becomes visible, the weak prediction is confirmed, and steering can be adjusted gradually while resolving residual uncertainty. If the road turns out to be clear, the weak prediction is simply suppressed, and the driver will continue along the preferred route smoothly with minimal effort.

If the weak prediction (the bend) is consistently confirmed **over repeated trips**, the new route becomes retained and strengthened, and anticipatory navigation becomes more efficient. However, if the weak prediction is not confirmed (the road is actually straight and the workers were absent-minded), it is dropped, and its weight is reduced in future predictions.

## Why proactive validation?

Proactive validation finds empirical support in cross-linguistic data. Processing cost for non-canonical structures has been shown to vary depending on the availability of predictive cues in each language. The German version of the above relative clause *der Junge, den das Mädchen küsst* [Eng. the boy that the girl kisses] contains early information about the syntactic function of the head NP (*der Junge*.the boy) via the accusative marking on the relative pronoun (*den.that.ACC.SG.*)—except for case syncretism—allowing for early disambiguation. In Russian, case and agreement are even more diagnostic: the syntactic function can almost always be signaled by case-marking on the relative pronoun (*kotor-yi.that.NOM.SG* vs. *kotor-ogo.that.ACC.SG*). Speaker-hearers can, thus, generate an expectation of an object relative clause while simultaneously maintaining the subject-first parse. The expectation of an SRC aligns with word order preferences in German where canonical subject-first structures are processed more efficiently ([Adani and Fritzsche, 2015](#); [Bornkessel et al., 2002](#); [Edeleva and Slioussar, accepted](#)) though empirical evidence toward word order canonicity effects in Russian is less robust (cf. [Edeleva and Slioussar, accepted](#); [Edeleva et al., 2020](#)).

In English—a language with a very rigid word order and late disambiguation—garden-path structures produce robust signatures of reanalysis such as elevated reading times in the disambiguating region, left anterior negativity (LAN) and late positivity ([Osterhout and Holcomb, 1992](#); [King and Kutas, 1995](#)). Comparable effects are reduced in languages with a canonical SVO word order but early cues that support anticipatory disambiguation. In German, adults and children exploit case-marking early to fine-tune syntactic predictions ([Özge et al., 2022](#); [Brandt et al., 2016](#); [Bornkessel-Schlesewsky and Schlesewsky, 2009](#)). In Dutch ([Johnson, 2005](#)), French ([Van Heugten and Shi, 2009](#)), Spanish ([Lew-Williams and Fernald, 2007](#)) and Czech ([Smolík and Bláhová, 2019](#)), children can use gender information on articles and adjectives to anticipate upcoming nouns though predictive ability may be constrained by children’s grammatical development.

Cross-linguistic differences are evident in the magnitude of the processing trade-off for canonical and non-canonical structures.

<sup>2</sup> cf. also [Staub \(2025\)](#) who reviews extensive empirical evidence on word frequency and garden-path-processing and concludes that incremental processing difficulty is not reducible to the effect of a word’s predictability.

Edeleva et al. (2020) did not evidence the typical subject-first advantage with Russian native speakers at the point of disambiguation in a self-paced reading task when the initially occurring noun was syncretic between the nominative and the accusative (cf. *uchitelja.teacher/teachers.ACC.SG/NOM.PL*). Critically, the subject preference did not show up even for animate targets, suggesting high fidelity of morphological cues. Combined with considerable word order flexibility, this likely explains why Russian native speakers are more efficient at processing non-canonical structures and exhibit shorter response latencies than German native speakers in tasks like picture-matching (Kempe and MacWhinney, 1999).

If prediction error were the driving force of learning, one would expect users of English—where cues are very salient (e.g., lexical or positional constraints) and errors more frequent—to learn very rapidly. In cue-rich languages, prediction-error learning should be slower as redundant cues make disconfirmation events rare and less salient. However, empirical evidence shows the opposite: language users achieve greater processing efficiency for non-canonical structures in cue-rich environments.

Proactive cue-integration is further supported by psychophysiological studies. Sustained frontal negativity has been interpreted as a maintenance and expectation-checking process when the participants were requested to actively predict prior to disambiguation (Lai et al., 2024). When cues are available early, ERP amplitudes (N400; P600) can be smaller or delayed (Aurnhammer et al., 2021; Bornkessel-Schlesewsky and Schlesewsky, 2019). It indicates that the analysis becomes validated when confirmatory cues emerge.<sup>3</sup> Eye-tracking studies report longer gaze durations, go-past and reaction times in the pre-validation regions when multiple parses remain viable (Staub et al., 2007; Levy et al., 2013). These findings are best captured through the lens of *interpretive stability* which refers to the tendency to maintain a coherent, contextually grounded interpretation of an unfolding sentence rather than overturning ongoing analysis at every small mismatch. The parser does not wait to reconfigure the entire syntactic and semantic representation whenever the input becomes inconsistent. Instead, it maintains a leading interpretation. Simultaneously, it preserves probabilistically weaker alternatives as graded competitors and monitors the incoming signal for confirmation or disconfirmation.

Developmental and second-language research provides additional evidence. Pre-school children show improved comprehension of object relative clauses when the head and the embedded NP mismatch in gender or number (Belletti et al., 2012; Adani et al., 2010; Arosio et al., 2012), suggesting that they consult late-occurring cues to ensure interpretive stability. Young children and L2 speakers who rely more heavily on structural heuristics exhibit strong garden-path effects (Trueswell et al., 1999; Clahsen and Felser, 2006). With increasing experience and growing knowledge, sensitivity to cue validity improves and cues such as case and gender marking are exploited early (Dussias et al., 2013;

Hopp and Lemmerth, 2018). Thus, proficiency growth aligns with a gradual transition from error-driven adjustments to goal-directed, proactive control.

The study by Hopp and Lemmerth (2018) delivers further evidence that challenges prediction error as the primary mechanism for language learning. They found that advanced German L2 learners with Russian as L1 could use gender marking on both adjectives and determiners to predict upcoming nouns in German similarly to native speakers. By contrast, high-intermediate L2 learners consistently used gender marking predictively only on adjectives. Notably, the predictive condition elicited significantly longer reaction times ( $\approx 1,600$ – $1,700$  ms) compared to the non-predictive condition ( $\approx 1,400$ – $1,500$  ms). Since language processing is incremental, learners would be maintaining at least one structural alternative in any case. Why, then, is maintaining a single structural alternative more costly in the predictive than in the non-predictive condition? Further, the researchers manipulated gender congruence between L1 and L2. They included congruent and incongruent nouns which were matched for corpus frequency in German. If learners use gender marking predictively, the general likelihood to encounter a prediction error is comparable across congruent and incongruent nouns. Interestingly, however, high-intermediate L2 learners could use gender marking on articles predictively only when the upcoming noun was lexically congruent. In fact, incongruent nouns should theoretically generate greater prediction error (they are still less expected due to L1–L2 contrasts), but learners failed to use predictive cues in this condition altogether. Proactive validation offers a more plausible account: learners maintain multiple alternatives and validate predictions, a process that succeeds for gender-congruent nouns but not for gender incongruent nouns as their gender is less well-entrenched due to L1–L2 contrasts.

Recent experimental work by Edeleva (2023, 2024) delivers distinct evidence of validation-like behavior during comprehension. The studies examined how children comprehend subject and object relative clauses in Russian and German using a character selection task. The participants could opt for the head NP *the cat* as the relative clause subject or object or the embedded NP *the hedgehog* (cf. Figure 1). The experimental sentences (cf. Table 1) contained two types of cues: (1) a case-marked relative pronoun as an early cue (*der.NOM.SG* vs. *den.ACC.SG//kotoryi.NOM.SG* vs. *kotorogo.ACC.SG*) and (2) a case-marked embedded NP (the nominative vs. the accusative case form of the definite determiner in German and the respective noun inflection in Russian) as a late cue. Two types of non-target responses were quantified: (1) syntactic reversal errors when children misassigned the syntactic function of the head NP and selected the subject-*cat* instead of the object-*cat*; (2) embedded NP errors where children selected the embedded NP character (*the hedgehog*).

Both Russian and German children showed greater commitment to the embedded NP character in ORCs. In the ORC condition (*Where is the cat that the hedgehog is feeding?*), German children were more likely to fixate the hedgehog than in the SRC condition (*Where is the cat that is feeding the hedgehog?*). In Russian, accuracy in the character selection task revealed a similar pattern: children selected the embedded NP character as the intended target more frequently in ORCs than in SRCs.

<sup>3</sup> Note that Aurnhammer et al. (2021) and Lai et al. (2024) contrast a highly probable alternative against a highly improbable one. A more gradual manipulation of probability would be desirable to indeed measure maintenance of less probable alternatives.

Edeleva and Slioussar (accepted) used an identical experimental design with adult native speakers of Russian, German and French. In a classical visual world paradigm, the participants' eye movements were tracked while they were listening to SRCs or ORCs. Russian and German speakers fixated more on the embedded NP character in ORCs and this effect extended into the verb region. By contrast, French native speakers – whose embedded NPs do not provide additional morphological cues either on the determiner or on the noun itself (*le hérisson*) – showed little engagement with the embedded NP character. These results complement classical garden-path profiles: comprehenders use early cues to pre-activate less probable ORCs and confirm them later if additional evidence is available. Late validation effects suggest that the parser maintains flexibility to minimize the likelihood of potential downstream repairs.

Language learning also takes place in a noisy environment. Children and L2 learners need to sample the input across different speakers or contexts, where variability in accent, speech rate, grammatical realization challenges stable cue-outcome mappings, as inconsistencies create transient statistical irregularities. Under a purely prediction-error framework, such variability would induce overfitting, as the system continuously readjusts to every inconsistency. This process would ultimately compromise generalization and consolidation, i.e. learners would internalize regularities specific to particular contexts or individual speakers, rather than robust, generalizable patterns.

By contrast, if learners engage in proactive validation, they can stabilize representations by confirming reliable regularities across noisy instances instead of recalibrating after each deviation. This mechanism provides a built-in buffer against overfitting and explain how learners extract consistent patterns from highly variable input, i.e. why, despite pervasive noise, children and proficient L2 speakers converge on robust grammatical systems.



FIGURE 1  
Sample visual display from Edeleva (2023).

TABLE 1 Sample stimulus sentences from Edeleva (2023, 2024).

GER_SRC	Wo ist	der.NOM.SG Kater,	der.NOM.SG gleich bestimmt	den.ACC.SG Igel füttert?
RUS_SRC	Gde kot.NOM.SG		kotor-yi.NOM.SG vozmozhno	pokormit ezha.ACC.SG?
	Where is	the cat	that will likely	feed the hedgehog?
GER_ORC	Wo ist	der.NOM.SG Kater,	den.ACC.SG gleich bestimmt	der.NOM.SG Igel füttert?
RUS_ORC	Gde kot.NOM.SG		kotor-ogo.ACC.SG vozmozhno	pokormit ezh.NOM.SG?
	Where is	the cat	that the hedgehog will likely feed?	

Word order manipulation in Russian relative clauses is not presented.

Eye-tracking evidence suggests that children and adult native speakers (Edeleva, 2023; Edeleva and Slioussar, accepted) engage in this validation: they anticipate and check predictions when supporting cues are available. The system can operate under residual uncertainty—language users actively maintain multiple structural alternatives and process the input without immediately collapsing when strong predictions fail. Even if comprehension ultimately breaks down, as often occurs with ORCs, the system is still actively processing and validating the input. By contrast, prediction-error models define rational behavior as the gradual reduction of uncertainty. It would force constant and costly recalibration and eliminate functional states in which uncertainty is a resource for interpretive flexibility rather than a risk of not producing a single error-free outcome.

Another key point is differences in learning outcomes for children and L2 speakers. Prediction-error models would suggest that all learners should benefit from exposure in a technically similar way as each error provides a signal to update internal representations. Yet, adult L2 learners often exhibit fossilized errors. They persist to maintain non-target forms or constructions despite repeated exposure to correct input, whereas children acquire the same structures more fully. Prediction-error models fail to explain this discrepancy: if errors alone drove learning, repeated exposure to correct input would eventually eliminate and overwrite fossilized patterns, which rarely occurs.

A likely explanation involves differences in prior knowledge. Children have relatively weak or unstable priors which allow them to explore a broader range of predictions and integrate new patterns efficiently. Adult L2 speakers bring strong, entrenched priors from their L1 and prior experience. Proactive validation initially reinforces strong predictions as they are consistent with entrenched priors. Weak predictions, such as novel forms or cues, can still be maintained and evaluated if the input provides highly informative cues and ample exposure. Crucially, entrenched priors can stabilize both target-like and non-target-like forms. Strong priors bias interpretation such that even partially inconsistent input may be perceived as compatible with the prediction, causing the system to confirm it.

In contrast to prediction error, proactive validation provides opportunities for incremental improvement. It predicts that adult L2 learners continue processing even if communication succeeds (= no error noticed by the language user) and no overt modifications to entrenched non-target patterns occur. The system actively checks predictions against the input but updating takes place only once enough evidence has accumulated to override entrenched expectations. This mechanism captures why fossilized forms can persist silently and re-emerge under stress, anxiety, or

divided attention, when limited resources reduce active validation and increase reliance on strong priors.

In summary, rational and predictive models of sentence processing (cf. a recent special issue edited by Crocker et al., 2021) have convincingly demonstrated that processing operates under probabilistic principles. Yet they typically frame adaptation as a reactive process. The parser adjusts dynamically after prediction failures. Efficiency is achieved *post-hoc* – by minimizing surprisal or optimizing recovery from error. Prediction error remains a necessary by-product of learning.

The current perspective introduces a complementary, proactive dimension. Rather than reacting to errors, the processor acts to minimize the likelihood of such errors in advance. It maintains low-probability parses as weak predictions and verifies them when confirmatory cues arrive. This validation layer enables weak predictions to persist in long-term during learning without the necessity to invoke errors. Therefore, the system doesn't primarily track what goes wrong, but samples and consolidates structures that consistently work.

## Mathematical formalization of proactive validation

A formalization of proactive validation within a probabilistic parsing framework should represent candidate parses, cue uptake and validation. Let  $P = \{p_1, p_2, \dots, p_n\}$  be a set of possible parses, a hypothesis space at time point  $t$ . Then the weight of parse  $p_i$  evolves as:

$$\omega(p_i; t) = \lambda \cdot \omega_i(p_i; t-1) + \alpha \cdot \text{support}(p_i; c_t)$$

where

- $\omega(p_i; t)$  = the weight of parse  $p_i$  at any time point  $t$ ;
- $\omega_i(p_i; t-1)$  = the weight of parse  $p_i$  at previous time point  $(t-1)$ ;
- $\text{support}(p_i; c_t)$  = cue confirming parse  $p_i$  (0 for non-informative input).

To model within-sentence maintenance,  $0 < \lambda < 1$  controls the retention or decay of possible parses. If  $\lambda < 1$ , weak parses will decay slowly if not confirmed for a long time. For  $\lambda = 1$ , the parse will persist, and weight will grow via validation. This way, weak alternatives will eventually fade if not confirmed.

A further parameter  $\alpha$  represents the learning rate. It controls how strongly the confirmatory cue modifies the weight of a given parse.  $\alpha$  quantifies how quickly the system stabilizes a parse once it has been validated. A higher  $\alpha$  represents rapid consolidation of confirmed structural configurations. In long term, this would result in faster convergence on stable expectations in cue-rich environments. A lower  $\alpha$  produces more conservative updating and allows the parser to remain flexible under variability or cross-linguistic interference.

Thus,  $\alpha$  has a dual function: locally, it controls the strength with which an ongoing weak parse is validated; globally, it regulates how quickly these structures are internalized into long-term probabilistic representations. In this sense,  $\alpha$  embodies the speed-stability trade-off in proactive validation. Too high  $\alpha$  overfits to

transient regularities; while too low  $\alpha$  slows adaptation to reliable cues. An optimal  $\alpha$  will balance stability and flexibility, which aligns proactive validation with principles of rational adaptation rather than reactive correction.

While proactive validation primarily relies on confirmatory evidence, late non-informative cues may still induce residual mismatches. Thus, it is reasonable to include residual prediction error into the equation as a fine-tuning signal to adjust the strength of validated parses:

$$rPE_i = (\omega_i^{\text{expected}} - \omega_i^{\text{observed}})^2$$

The squared prediction error always remains positive; it ensures symmetry in that it treats over- and under-predictions equally. It provides a smooth correction function that downweights small discrepancies but preserves sensitivity to larger ones. Thus, the parser maintains representational stability while fine-tuning expectations across exposures. In a way, it learns not from failure itself, but from residual uncertainty that remains when a prediction is largely accurate.

It may also be reasonable to treat residual prediction error as a variance term. When a late cue (e.g., a syncretic case marking on the determiner) does not deliver confirmatory evidence, it still provides valuable information about the general uncertainty of ongoing parses. By quantifying the spread of observed vs. expected weights, the parser tracks how uncertain each parse remains without overreacting to minor deviations.

Mathematically:

$$rPE_i = \sum_j P(p_j|c_t) (\omega_i^{\text{expected}} - \omega_i^{\text{observed}})^2$$

where

$\sum_j P(p_j|c_t)$  is the sum of conditional probabilities of alternative parses  $p_j$  given the cue  $c_t$ .

Thus, residual prediction error corresponds to total residual uncertainty that remains after the less expected parse has been validated.

Integrating all components, the evolution of parse weights under proactive validation can be expressed as:

$$\omega(p_i; t) = \lambda \cdot \omega_i(p_i; t-1) + \alpha \cdot \text{support}(p_i; c_t) + rPE_i$$

An additional anticipatory component can be incorporated to account for expected future evidence. The system might also pre-weight parses as to the likelihood that the upcoming input (e.g., number/gender information on the verb in a verb-final relative clause) will confirm them. Anticipatory validation allows the parser to maintain weak alternatives that are expected to be relevant shortly rather than rely on present and past evidence.

$$\omega(p_i; t) = \lambda \cdot \omega_i(p_i; t-1) + \alpha \cdot \text{support}(p_i; c_t) + rPE_i + \beta \cdot \mathbb{E}[\text{support}(p_i; c_{t+1:t+T})]$$

where

$\beta$  is the weighting of the anticipatory component;

$\mathbb{E}[\text{support}(p_i; c_{t+1:t+T})]$  is the expected confirmatory support from upcoming cues over horizon  $T$ .

The full equation formalizes proactive validation as a stability-optimizing mechanism: the parser distributes computational

resources across both current evidence and likely upcoming validation, which allows to support interpretive stability without over-relying on reactive error correction.

## Future research directions

More extensive empirical work is needed to determine how reactive adjustment through prediction error correction and proactive validation complement each other during language processing. A key challenge is to disentangle their processing signatures. A tentative hypothesis is that prediction error reflects an early signal of repair. By contrast, proactive validation can have an extended spell-out, supporting interpretive stability. The two processes may also interact differently with cognitive effort measures. Prediction error produces localized spikes in processing effort at points of conflict and adjustment. Validation distributes cognitive effort earlier—at points where multiple alternatives are maintained. Critically, the trade-off between short-term effort and long-term interpretive stability could explain why late-disambiguating structures, such as object relative clauses, continue to elicit processing load beyond the frequency of exposure. The parser distributes cognitive resources early to maintain multiple alternatives. Because the input unfolds incrementally, some ambiguity persists until supporting cues arrive. Thus, the processing load after early allocation may be reflective of ultimate commitment to the less probable parse. Processing cost, in this sense, may reveal a rational balance between resource expenditure and interpretive stability.

Empirical validation of this framework requires combining fine-grained temporal measures with computational modeling. In ERP studies, prediction error has been linked to an N400 (Bornkessel-Schlesewsky and Schlesewsky, 2019). By contrast, proactive validation can correspond to a sustained frontal negativity (Lai et al., 2024), reflecting anticipatory maintenance of possible alternatives which might mitigate the magnitude of the N400. Eye-tracking measures can further differentiate between the two. Prediction error predicts disruptions stipulated in early eye-tracking measures (e.g., first fixation duration, regression proportions). Validation can become evident in longer inspection times in pre-disambiguation regions—where multiple alternatives are entertained—but smoother transitions post-disambiguation. In other words, validation as a process cannot receive sufficient empirical support unless temporal dynamics are examined in interaction.

The shift from reactive error correction to proactive minimization offers a testable explanation for cross-linguistic variation. Languages with limited morphosyntactic cues (e.g., English) may rely on top-down heuristics, which provokes frequent prediction errors. Conversely, cue-rich languages (e.g., Russian, German) may reduce the need for reactive correction as case marking and agreement support early activation and reliable validation of alternative parses. Additionally, word order flexibility introduces global uncertainty, encouraging the parser to activate and maintain low-probability, cue-driven alternatives. Developmental studies have so far been unsuccessful in replicating the beneficial effect of statistical optimization through prediction error for learning (Gambi et al., 2024). Both

first- and second-language learning research could test how increasing morphological awareness and cue integration capacities gradually result in more rational behavior and facilitate transitions from error-driven to proactively controlled processing.

Häuser et al. (2022) raise an important concern regarding the working memory demands associated with multiple graded predictions. Their findings show that younger comprehenders routinely used sentence context to keep several potential continuations active while older adults generated narrower and more selective predictions. This suggests that alternative analyses might overload working memory.

Within the proactive validation framework, however, these alternatives are not full structural representations but skeletal hypotheses—abstract representations which contain just enough information for rapid confirmation or dismissal. These alternatives function as latent candidates: they remain below the full retrieval threshold (and thus do not tax working memory) but can be reactivated on demand when relevant cues appear. In this way, the system minimizes the risk of a costly reanalysis and at the same time avoids the burden of keeping richly specified parses in working memory.

Computational implementations could formalize how probabilistic weights are assigned to these latent alternative parses across different cue environments and under varying cognitive load. Adding a validation layer—a mechanism that maintains low-weight alternatives and verifies them when confirmatory cues arrive—offers a principled way to model cross-linguistic differences in non-canonical structure processing. Extending existing cue-based retrieval models in this direction may simulate how the parser uses latent alternatives to uphold interpretive stability, reduce reanalysis cost, and flexibly adapt its expectations when the unfolding input supports an alternative parse.

## Conclusion

The current perspective examined empirical evidence to reconsider the role of prediction error as the principal driver of probabilistic language processing and learning. While prediction error provides valuable feedback to the processing system, it is paradoxically grounded in discrepancies that a rational processor would strive to avoid. Rather than being a system tuned to detect prediction errors and adjust error correction strategies, language processing may instead operate to prevent them through proactive validation of alternative parses. We can assume that the system activates and maintains less probable parses until they can be verified or rejected against the input. In other words, the system does not determine what went wrong, but what consistently works. Weak predictions will only be kept if they are verified.

Being a complementary mechanism, **proactive validation** reframes processing cost as an adaptive trade-off between flexibility and interpretation stability as it moves beyond the assumption of error-prone behavior. As this view is integrated into existing models of predictive processing, it can bridge current divides between top-down, probabilistic and cue-based approaches, and align language processing with more general accounts of rational cognition. If learning is viewed as maintaining structural representations beyond their immediate processing, maintaining less probable parses is

beneficial only insofar as these parses are periodically validated and, consequently, reinforced against the input. In doing so, the system minimizes the risk of failure before it escalates into a costly revision. Simultaneously, the processor samples information about the global statistical structure of a given language. Over repeated exposures, this process yields more calibrated expectations, more efficient cue integration, and the gradual consolidation of less probable parses.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

JE: Conceptualization, Writing – review & editing, Writing – original draft.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This publication was funded by the University Library of the University of Kassel.

## References

Adani, F., and Fritzsche, T. (2015). "On the relation between implicit and explicit measures of child language development: evidence from relative clause processing in 4-year-olds and adults," in *Proceedings of the 39th Boston University Conference on Language Development*, Vol. 1, eds. E. Grillo, and K. Jepson (Somerville, MA: Cascadilla Press), 14–26.

Adani, F., van der Lely, H. K. J., Forgiarini, M., and Guasti, M.T. (2010). Grammatical feature dissimilarities make relative clauses easier: a comprehension study with Italian children. *Lingua* 120, 2148–2166. doi: 10.1016/j.lingua.2010.03.018

Arosio, F., Yatsumihi, K., Forgiarini, M., and Guasti, M. T. (2012). Morphological information and memory resources in children's processing of relative clauses in German. *Lang. Learn. Dev.* 8, 340–364. doi: 10.1080/15475441.2011.634691

Aurnhammer, C., Delogu, F., Schulz, M., Brouwer, H., and Crocker, M. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PLoS ONE* 16:e0257430. doi: 10.1371/journal.pone.0257430

Belletti, A., Friedmann, N., Brunato, D., and Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122, 1053–1069. doi: 10.1016/j.lingua.2012.02.007

Bornkessel, I., Schlesewsky, M., and Friederici, A.D. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition* 85, B21–B30. doi: 10.1016/S0010-0277(02)00076-8

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Lang. Linguist. Compass* 3, 19–58. doi: 10.1111/j.1749-818X.2008.00099.x

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10:298. doi: 10.3389/fpsyg.2019.00298

Borovcnik, M. (2015). Risk and decision making: the "logic" of probability. *Math. Enthusiast* 12:14. doi: 10.54870/1551-3440.1339

Bovolenta, G., and Marsden, E. (2022). Prediction and error-based learning in L2 processing and acquisition: a conceptual review. *Stud. Second Lang. Acq.* 44, 1384–1409. doi: 10.1017/S0272263121000723

Brandt, S., Lieven, E., and Tomasello, M. (2016). German children's use of word order and case marking to interpret simple and complex sentences: testing differences between constructions and lexical items. *Lang. Learn. Dev.* 12, 156–182. doi: 10.1080/15475441.2015.1052448

Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Rev.* 113, 234–272. doi: 10.1037/0033-295X.113.2.234

Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/S0142716406060024

Crocker, M. (2005). "Rational models of comprehension: addressing the performance paradox," in *Twenty-First Century Psycholinguistics*, ed. A. Cutler (Routledge: Four Cornerstones).

Crocker, M., Jäger, G., Kuperberg, G., Rohde, H., Teich, E., Turnbull, R. (eds.). (2021). Research topic: rational approaches in language science. *Front. Psychol.* 12. doi: 10.3389/978-2-88974-765-8

Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. Biol. Sci.* 369:20120394. doi: 10.1098/rstb.2012.0394

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., and Gerfen, C. (2013). When gender and looking go hand in hand: grammatical gender processing in L2 Spanish. *Stud. Second Lang. Acquis* 35, 353–387. doi: 10.1017/S0272263112000915

Edeleva, J. (2023). Embedded NP error in German object relative clause comprehension: a case for a universal developmental pathway. *Q. J. Exp. Psychol.* 76, 1220–1232. doi: 10.1177/17470218221114747

Edeleva, J. (2024). Russian children and their relatives: what can a free word order language reveal about the subject/object asymmetry? *Lang. Learn. Dev.* 21, 118–141. doi: 10.1080/15475441.2024.2372247

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Edeleva, J., Chrabaszcz, A., and Demareva, V. (2020). Resolving conflicting cues in processing of ambiguous words: the role of case, word order and animacy. *Q. J. Exp. Psychol.* 73, 1173–1188. doi: 10.1177/1747021820902429

Edeleva, J., and Slioussar, N. (accepted). Non-target processing and response patterns attest to cross-linguistic differences in structural analysis. *Q. J. Exp. Psychol.*

Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Fujita, H. (2023). Predictive structure building in language comprehension: a large-sample study on incremental licensing and parallelism. *Cogn. Process.* 24, 301–311. doi: 10.1007/s10339-023-01130-8

Gambi, C., Lelonkiewicz, J. R., and Crepaldi, D. (2024). Do children (and adults) benefit from a prediction error boost in one-shot word learning? *J. Cogn.* 7, 1–16. doi: 10.5334/joc.342

Gambi, C., and Messenger, K. (2023). The role of prediction error in 4-year-olds' learning of English direct object datives. *Languages* 8:276. doi: 10.3390/languages8040276

Gennari, S. P., and MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *J. Mem. Lang.* 58, 161–187. doi: 10.1016/j.jml.2007.07.004

Gennari, S. P., and MacDonald, M. C. (2009). Linking production and comprehension processes: the case of relative clauses. *Cognition* 111, 1–23. doi: 10.1016/j.cognition.2008.12.006

Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1411–1423. doi: 10.1037/0278-7393.27.6.1411

Häuser, K., and Borovsky, A. (2025). Got it right up front? Further evidence for parallel graded prediction during prenominal article processing in a self-paced reading study. *Glossa Psycholinguist.* 4, 1–40. doi: 10.5070/G6011.1636

Häuser, K., Kray, J., and Borovsky, A. (2022). Hedging bets in linguistic prediction: younger and older adults vary in the breadth of predictive processing. *Collabra Psychol.* 8:36945. doi: 10.1525/collabra.36945

Hopp, H., and Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Stud. Second Lang. Acquis.* 40, 171–199. doi: 10.1017/S0272263116000437

Johnson, E. K. (2005). "Grammatical gender and early word recognition in Dutch," in *Proceedings of the 29th Boston University Conference on Language Development*, eds. A. Brugos, M. R. Clark-Cotton, and S. Ha (Somerville, MA: Cascadilla Press), 320–330.

Kaan, E., and Grüter, T. (eds.). (2021). *Prediction in Second Language Processing and Learning*. Amsterdam; Philadelphia, PA: John Benjamins Publishing. doi: 10.1075/bpa.12

Kempe, V., and MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Lang. Cogn. Process.* 14, 129–171. doi: 10.1080/016909699386329

King, J. W., and Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *J. Cogn. Neurosci.* 7, 376–395. doi: 10.1162/jocn.1995.7.3.376

Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299

Lai, M. K., Payne, B. R., and Federmeier, K. D. (2024). Graded and ungraded expectation patterns: prediction dynamics during active comprehension. *Psychophysiology* 61:e14424. doi: 10.1111/pyps.14424

Lau, E., and Tanaka, N. (2021). The subject advantage in relative clauses: a review. *Glossa J. Gen. Linguist.* 6:34. doi: 10.5334/gjgl.1343

Levy, R., Fedorenko, E., and Gibson, T. (2013). The syntactic complexity of Russian relative clauses. *J. Mem. Lang.* 69, 461–495. doi: 10.1016/j.jml.2012.10.005

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 275–419. doi: 10.1207/s15516709cog0000\_25

Lew-Williams, C., and Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychol. Sci.* 18, 193–198. doi: 10.1111/j.1467-9280.2007.01871.x

MacDonald, M. C., Pearlmuter, N. J., and Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703. doi: 10.1037/0033-295X.101.4.676

Mak, W. M., Vonk, W., and Schriefers, H. (2008). Discourse structure and relative clause processing. *Mem. Cogn.* 36, 170–181. doi: 10.3758/MC.36.1.170

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Özge, D., Kornfilt, J., Maquate, K., Küntay, A. C., and Snedeker, J. (2022). German-speaking children use sentence-initial case marking for predictive language processing at age four. *Cognition* 221:104988. doi: 10.1016/j.cognition.2021.104988

Rao, R., and Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Reuter, T., Borovsky, A., and Lew-Williams, C. (2019). Predict and redirect: prediction errors support children's word learning. *Dev. Psychol.* 55, 1656–1665. doi: 10.1037/dev0000754

Safak, D. F., and Hopp, H. (2023). Cross-linguistic differences in predicting L2 sentence structure: the use of categorical and gradient verb constraints. *Stud. Second Lang. Acquis.* 45, 1234–1260. doi: 10.1017/S0272263123000256

Safak, D. F., and Hopp, H. (2025). Learning L2 grammar from prediction errors? Verb biases in structural priming in comprehension and production. *Biling. Lang. Cogn.* 1–17. doi: 10.1017/S1366728925000033

Smolik, F., and Bláhová, V. (2019). Czech 23-month-olds use gender agreement to anticipate upcoming nouns. *J. Exp. Child Psychol.* 178, 251–265. doi: 10.1016/j.jecp.2018.10.004

Spivey-Knowlton, M. J., Trueswell, J., and Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Can. J. Exp. Psychol./Revue Canadienne de psychologie expérimentale* 47, 276–309. doi: 10.1037/h0078826

Staub, A. (2025). Predictability in language comprehension: prospects and problems for surprisal. *Annu. Rev. Linguist.* 11, 17–34. doi: 10.1146/annurev-lingistics-011724-121517

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., and Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 1162–1169. doi: 10.1037/0278-7393.33.6.1162

Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten path effect: studying on-line sentence processing in young children. *Cognition* 73, 89–134. doi: 10.1016/S0010-0277(99)00032-3

Van Heugten, M., and Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Dev. Sci.* 12, 419–425. doi: 10.1111/j.1467-7687.2008.00788.x