


ORIGINAL ARTICLE

The predictive processing of number information in subregular verb morphology in a first and second language

Eva Marie Koch^{1,*} , Bram Bulté¹, Alex Housen¹ and Aline Godfroid²

¹Department of Linguistics and Literary Studies, Vrije Universiteit Brussel, Brussels, Belgium and

²Department of Linguistics and Languages, Michigan State University, East Lansing, MI, USA

*Corresponding author. Email: koch.eva.marie@gmail.com

(Received 17 November 2021; revised 19 January 2023; accepted 20 March 2023; first published online 08 May 2023)

Abstract

We investigated the predictive processing of grammatical number information through stem-vowel alternations in German strong verbs by adult first language (L1) speakers and Dutch-speaking advanced second language (L2) learners of German, and the influence of working memory and awareness (i.e., whether participants consciously registered the predictive cue) thereon. While changed stem vowels indicate a singular referent (e.g., /ɛ/ in *fällt*_{3SG}, “falls”), unchanged vowels indicate plural (e.g., /a/ in *fallt*_{2PL}, “fall”). This target structure presents a challenge for L2 learners of German due to its subregularity and low salience. With their eye movements being tracked, participants matched German auditory sentences (VSO order) with one of two pictures, displaying identical action scenes but varying in agent number. The number cue provided by the strong verbs allowed participants to predict whether the upcoming subject would be singular or plural. The analyses revealed significant prediction, measured as predictive eye movements toward the target picture and faster button-press responses. Prediction in the L2 group was weaker than in the L1 group and present in the eye movement data only. Higher working memory scores were linked to faster predictive presses. Approximately half of the participants had become aware of the predictive cue, and being aware facilitated prediction to a limited extent.

Keywords: eye tracking; L1/L2 predictive processing; verb morphology; working memory; awareness

Introduction

In first language (L1) listening comprehension, we continuously integrate various types of linguistic and contextual information not only to interpret what is being said but also to generate predictions about what will be said next (for reviews, see Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). Predictive (or anticipatory) processing allows for dialogic interaction, and

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

communication in general, to proceed rapidly, smoothly, and efficiently (Kutas et al., 2011; Pickering & Garrod, 2004). However, recent research suggests that L1 listeners do not always engage in predictive processing (Huettig & Guerra, 2019; Mishra et al., 2012), and that prediction may not be as automatic and unconscious as it was often assumed to be (see Brothers et al., 2017). Moreover, many questions remain regarding the extent to which prediction also plays a role in second language (L2) listening (for a review, see Godfroid, 2020), and regarding the internal and external factors that influence prediction.

In this study, we further explored the nature of grammatical prediction in L1 and L2 sentence processing. So far, most studies on morphosyntax-based prediction have used gender markings in the noun phrase (NP) as predictive cue (e.g., Hopp & Lemmerth, 2018; Huettig & Guerra, 2019), and there is a need to complement this research area with new target structures. Therefore, we focused on verb morphology, which has rarely been the object of earlier prediction research. We investigated the predictive processing of stem-vowel alternations that occur in German strong verbs in the simple present – a subtle, subregular morphosyntactic feature that is often incompletely acquired in second language acquisition (SLA; see Godfroid, 2016). The vowel alternation (e.g., /ɛ/ in *fällt*_{3SG}, “falls” vs. /a/ in *fallt*_{2PL}, “fall”) encodes grammatical number information and represents a potential cue to predict whether the upcoming subject in VSO sentences will be singular or plural. Using visual world eye tracking, we compared prediction in adult L1 speakers of German and advanced L2 learners of German with Dutch as their L1; furthermore, we explored whether prediction was influenced by working memory capacity (see Huettig & Janse, 2016), and by the language users’ (un)awareness that the strong verbs’ morphology provided a predictive number cue (see Curcic et al., 2019). The participants in our study had previously participated in a related experiment (see Koch et al., 2021), revealing robust prediction effects based on number markings encoded in regular German verb affixation. It is thus interesting to investigate whether these same participants are also able to use strong verb morphology – in comparison a much more complex and challenging structure, especially for learners – for predictive number processing.

L1 and L2 predictive sentence processing

In its strict sense, predictive language processing means that expected upcoming linguistic information is preactivated in the brain before it is actually encountered in the input (see Huettig, 2015). A widely implemented method to experimentally measure such predictive processing is the visual world paradigm, exposing participants simultaneously to pictures on a computer screen and auditory stimuli, while their eye movements are being recorded (for reviews, see Huettig et al., 2011; Magnuson, 2019). Most commonly, prediction is operationalized as more looks toward a target picture before onset of the target in the auditory input in trials containing a predictive cue as compared to uninformative trials. In addition to this strict definition, prediction is sometimes operationalized in a less strict sense as “facilitation,” meaning that certain information allows for subsequent words to be processed faster and more easily, an effect which may become visible only after target onset. Following Karaca et al. (2021), we see such facilitation as a part of the prediction

process and therefore do not differentiate between facilitation and prediction in its strict sense in this study.

L1 listeners exploit a wide range of linguistic cues predictively, including lexical-semantic information (e.g., Kamide *et al.*, 2003), grammar (e.g., Altmann & Kamide, 2007), prosody (e.g., Henry *et al.*, 2017), or the wider discourse-context (e.g., Van Berkum *et al.*, 2005). Based on such cues, they can for instance predict semantic aspects of upcoming words (e.g., Altmann & Kamide, 1999), syntactic structure (e.g., Arai & Keller, 2013), or turn ends (e.g., Ruiter *et al.*, 2006). As for L1 morphosyntax-based predictive processing, several studies have provided positive evidence (e.g., Dahan *et al.*, 2000; Lew-Williams & Fernald, 2007), but note that most studies on this subject used gender markings on determiners as their target structure.

Recently, there has been an increase in studies exploring L2 predictive processing (Godfroid, 2020). Research has provided ample evidence that L2 listeners can make predictions based on lexical-semantic cues (e.g., Dijkgraaf *et al.*, 2017, 2019; Peters *et al.*, 2018). In contrast to L1 prediction, the outcomes regarding L2 prediction based on morphosyntactic cues have been more mixed. For instance, no L2 predictive processing could be found in some studies using case (e.g., Hopp, 2015; Mitsugi, 2017) or gender marking (Lew-Williams & Fernald, 2010). Positive evidence comes mainly from studies on gender-based prediction in L2 learners (e.g., Hopp & Lemmerth, 2018; Morales *et al.*, 2016). Some of these studies only found prediction in learners with an advanced proficiency (e.g., Dussias *et al.*, 2013), or in participants who exhibited solid knowledge of the target structure during production measures (e.g., Hopp, 2013). Moreover, prediction was facilitated when gender was congruent between the L2 and the learner's L1 (e.g., Dussias *et al.*, 2013; Hopp & Lemmerth, 2018), suggesting a dependency of prediction on L1–L2 similarity.

Studies comparing L1 and L2 predictive processing generally found prediction effects in the L2 to be smaller (e.g., Chun & Kaan, 2019; Grüter *et al.*, 2012), but this does not necessarily mean that the underlying mechanisms are different. Several factors may in fact account for the different prediction outcomes in L1 and L2 processing, suggesting that both rely on the same mechanisms (see Kaan, 2014): L2 processing in general may be affected by a diminished accuracy of lexical representations (Kaan, 2014), enhanced competing interlingual information (Weber & Broersma, 2012), and slower lexical access (Shook *et al.*, 2015). Combined, this may contribute to an increased cognitive load in the L2, potentially reducing the cognitive resources necessary for prediction (Kaan, 2014; also see Grüter *et al.*, 2014, 2017, who express this idea in their Reduced Ability to Generate Expectations – or RAGE – hypothesis).

The predictive use of verb number markings

To complement existing, mainly gender-marking-based research on morphosyntax-based prediction, we introduced a new target structure by focusing on subregular German verb morphology providing predictive cues for upcoming subject number (see below, for details and examples). Relatively few studies have investigated the predictive processing of number markings in verb morphology, and existing studies exclusively investigated the L1 English processing of the copula *be*, comparing its third-person singular (3SG) and third-person plural (3PL) forms *is* and *are* of

the simple present tense (PRES). Generally, they used visual world designs, contrasting single-object and multiple-objects pictures, accompanied by auditory stimuli in which the copula provided predictive number information. Kouider et al. (2006) used auditory stimuli such as *Look, there is a/are some loopite(s)*, and reported a tendency in 2-year-old toddlers to cast anticipatory looks after hearing the verb; however, their design did not allow to assess with certainty whether predictive eye movements were based on the verb, the quantifier that immediately followed it, or a combination of both. Lukyanenko and Fisher (2016) found predictive processing of the copula in 3-year-olds and adults when exposed to sentences such as *Where is/are the good apple(s)?*. Riordan et al. (2015), however, could not find predictive eye movements in a study with adult participants. A possible explanation for this null result might be that the time slot during which prediction could have occurred was too brief to measure any anticipatory eye movements (which may also apply to Kouider et al., 2006). Brown et al. (2021) found positive evidence for prediction; moreover, they found that the number cue provided by the copula reduced phonological competition effects when competitor words did not match the target number. Reuter et al. (2021) found predictive processing of the copula in adults and children when two simple pictures were contrasted, but not when complex visual stimuli (photographs including four potential referents) were used.

In contrast to English, German has a relatively rich morphosyntactic agreement system and its native speakers may thus be more inclined to rely on morphosyntactic cues than speakers of English. Moreover, research should go beyond using only the copula as carrier of predictive information, as its conjugation is highly irregular and achieved through suppletion, resulting in a very salient singular–plural distinction, while conjugation through affixation or allomorphy is less salient and harder to learn and process (Krause et al., 2015). Prediction based on regular (“weak”) affixation in German verbs has been investigated in Koch et al. (2021), a visual world study that is part of the same project as the present study and that involves the same sample (see Procedure). The German weak conjugation paradigm is still productive and marks morphosyntactic information (person, number, mood, and tense) uniquely through affixation. In the auditory stimulus sentences, the verb’s suffix (e.g., *folg-t*_{3SG}, “follows” vs. *folg-en*_{3PL}, “follow”) was the only morpheme allowing to distinguish between singular and plural and thus to anticipate the upcoming subject’s number. Both adult L1 and L2 speakers of German showed strong suffix-based prediction effects. All had become aware that the suffixes provided a number cue, suggesting prediction was controlled and strategic.

Here, we focus on German strong conjugation, which marks morphosyntactic information through a combination of affixation and stem-vowel alternations. In PRES, the most common alternations are $a \rightarrow \ddot{a}$ ($/a/ \rightarrow /ɛ/$) and $e \rightarrow i(e)$ changes (realized as $/e/ \rightarrow /i/$ or $/ɛ/ \rightarrow /ɪ/$), marking second- and third-person SG. These alternations entail the coexistence of two variants of the same verb stem, each of which marks different functional properties. Table 1 provides examples. Opposing the 3SG and 2PL PRES forms of strong verbs (e.g., *fäll-t*_{3SG} vs. *fall-t*_{2PL}) creates a minimal pair in which only the stem vowel provides a disambiguating number cue. This is the target structure of the present study. The question is whether L1 German speakers and advanced L2 learners of German (L1 Dutch), who were already found to

Table 1. Examples of German conjugation in the present tense

	Strong verbs		Weak verbs	Dutch cognate
	<i>a</i> → <i>ä</i> type	<i>e</i> → <i>i/ie</i> type		
	“fall”	“see”	“follow”	“fall”
Infinitive	fall- <i>en</i>	seh- <i>en</i>	folg- <i>en</i>	vall- <i>en</i>
1SG	fall- <i>e</i>	seh- <i>e</i>	folg- <i>e</i>	val
2SG	fäll- <i>st</i>	sieh- <i>st</i>	folg- <i>st</i>	val- <i>t</i>
3SG	fäll- <i>t</i>	sieh- <i>t</i>	folg- <i>t</i>	val- <i>t</i>
1PL	fall- <i>en</i>	seh- <i>en</i>	folg- <i>en</i>	vall- <i>en</i>
2PL	fall- <i>t</i>	seh- <i>t</i>	folg- <i>t</i>	vall- <i>en</i>
3PL	fall- <i>en</i>	seh- <i>en</i>	folg- <i>en</i>	vall- <i>en</i>

Note: Phonemic realization: *a* = /a/, *ä* = /ɛ/, *e* = /e/ (or /ɛ/ in other phonological contexts), *i* = /i/, *ie* = /i/. Changed vowels are marked in bold. A Dutch cognate of *fallen* (*vallen*, strong verb) is provided for comparison: Dutch also has a strong conjugation paradigm, but stem-vowel alternations are limited to past tense forms.

use the relatively “easy” suffixes of the weak paradigm predictively, also exploit such subtle cues predictively.

The strong paradigm represents a challenge for learners for a number of reasons. The paradigm is no longer productive, and strong verbs are thus a closed verb class; nonetheless, they have a high frequency in everyday German (Köpcke, 1998). When learners encounter a German verb infinitive that is novel to them, it is impossible to know what class the verb belongs to, making it hard to know when the alternation needs to be applied (Bybee & Newman, 1995). Yet, strong conjugation represents a *subregular* paradigm because – despite its unpredictable, irregular appearance – strong verbs are organized following specific functional and morphophonemic regularities (Godfroid, 2016; Köpcke, 1998).

Several studies (Godfroid, 2016; Godfroid & Uggen, 2013; Koch et al., 2023; Krause et al., 2015) that have used the vowel change as their target structure provide evidence for its persistent learning difficulty and found the vowel-change-free weak paradigm to be the default conjugation system in learners’ interlanguages. It is a challenge for learners to create and store solid, correct morphological representations of the stem variants with the changed vowels and to map them onto the correct morphosyntactic properties; the coexisting, incorrect default representations with unchanged vowels compete with and easily overrule these target representations. Factors that contribute to this learning difficulty are the unpredictable nature of the vowel change, its limited perceptual salience (see Simoens et al., 2017) and information redundancy (DeKeyser, 2005), because the morphosyntactic information encoded in the stem variants is also provided through the suffixes. Against this backdrop, one may expect prediction based on the vowel change in strong conjugation to be much more taxing – if at all possible – for L2 learners of German than prediction based on the weak suffixes.

Prediction and awareness

Besides broadening the range of morphosyntactic target structures, research also needs to explore potential moderating factors to better understand the nature of morphosyntactic prediction. In this study, we focus on awareness and working memory capacity (see below). Investigating awareness (or consciousness) – which we specify here as the participants’ conscious realization that certain linguistic structures provide cues to predict upcoming input – is interesting not only from a theoretical but also methodological perspective. Although prediction – especially in the L1 – is often assumed to be automatic and unaware, recent studies have questioned this by showing that prediction can also happen in a top-down, controlled and strategic way. Brothers et al. (2017) conducted an electroencephalography (EEG) study on the L1 semantic prediction of final words in written English sentences. Every participant performed one task that was presented as a mere comprehension task, and another that explicitly encouraged them to predict the sentence-final word. Prediction effects were stronger when participants were made aware of the task’s prediction goal. Huettig and Guerra (2019) investigated the effect of several task-related parameters (slow vs. normal speech rate, short vs. long preview time, presence vs. absence of an explicit prediction instruction) on the predictive use of gender-marked determiners in L1 Dutch in a visual world eye tracking study. When speech rate was normal and preview short, a small prediction effect was found but only when participants were instructed to predict; no prediction was found when there was no prediction instruction.

Awareness also represents a much-debated topic within SLA research (e.g., Leow & Hama, 2013; Williams, 2009). Curcic et al. (2019) conducted a learning experiment in which Dutch native speakers learned a miniature language based on Fijian. They explored the effect of several cognitive aptitude variables on the anticipatory usage of gender-marked determiners and assessed the (spontaneous) emergence of what they call *prediction awareness* (i.e., participants realized “that determiners sometimes helped them during the test to arrive at the correct answer faster”, Curcic et al., 2019, p. 57) by means of post-experiment interviews. The authors measured prediction as a higher proportion of looks toward the target and faster button-press reaction times for picture selection when determiners provided a predictive cue as compared to when they did not. The aptitude variables influenced prediction indirectly: they predicted the emergence of awareness, and only aware learners were found to engage in anticipatory processing. Similarly, in Andringa (2020) – a study that not only measured *if* but also *when* learners of an artificial L2 became aware – anticipatory processing of determiners marked for animacy and distance occurred only after participants had developed awareness. Both studies underscore the importance of assessing participants’ awareness through careful debriefing.

Prediction and working memory

Morphosyntax-based prediction might also depend on working memory. In a visual world study, Huettig and Janse (2016) explored the influence of individual differences on gender-marking-based prediction in L1 Dutch speakers and found a strong link between working memory capacity and prediction. Since then, several

prediction studies investigated the role of working memory for predictive processing, though none of them used morphosyntactic target structures. In a semantic prediction experiment, Ito *et al.* (2018) manipulated cognitive load by having half of the participants (either L1 or advanced L2 speakers of English) perform a memory task in addition to the prediction task. Increased cognitive load delayed prediction, and this was similar in the L1 and L2 groups. These findings indicate that cognitive resources play a role in prediction. Chun and Kaan (2019) investigated the relation between working memory and L2 semantic predictive processing more directly. In their visual world experiment, complex stimulus sentences were used to augment cognitive load. Prediction effects were found in both the L1 and advanced L2 English groups, starting somewhat earlier in the L1 group. Against the authors' expectations, L2 prediction was not modulated by working memory. More recently, in a study using ambiguous relative clauses, Chun (2020) did not find successful semantic prediction to depend on the L2 speakers' working memory. Both Chun and Kaan (2019) and Chun (2020) suggest small group size or a lack of variance among the participants as possible explanations for the absence of an effect of working memory on prediction. Favier *et al.* (2021) investigated whether L1 speakers of Dutch use passive constructions to predictively attribute agent roles. Working memory was a significant predictor of prediction, but the size of the effect was small. Clearly, there is a need for more experimental research to establish the precise nature of the relationship between working memory and predictive processing.

The current study

The present study further explores predictive L1 and L2 language processing in a visual world eye tracking experiment by using a new morphosyntactic target structure and by assessing the roles of awareness (i.e., the presence/absence of the spontaneously emerging awareness that one can focus on the verb morphology to anticipate subject number) and working memory for prediction. The L2 learners of German in our study were native speakers of Dutch, who had extensive prior knowledge of the weak and strong German conjugation paradigms. Dutch and German are typologically closely related Germanic languages that share many similarities. Crucially, Dutch also has a strong conjugation paradigm involving stem alternations in past tense forms, and many verbs are cognates in both languages. Yet, Dutch strong verbs never have alternating stems in PRES (see Table 1), potentially causing negative L1–L2 transfer involving blocking effects (Ellis, 2006), which makes it particularly difficult for Dutch speakers to fully master the strong conjugation paradigm (see Koch *et al.*, 2023). That is, learners first learn the association between 3SG PRES and the regular *-t* suffix, which matches their L1 and therefore allows for direct positive L1–L2 transfer; subsequently, it may be more taxing for them to learn at a later stage that stem vowels of strong verbs also constitute a cue for the same morphosyntactic information, because the firstly learned cue “overshadows” the new cue.

Our investigation addressed the following research questions:

- RQ1: To what extent do L1 and advanced L2 speakers of German exploit the morphosyntactic number information encoded in the strong verbs' stem vowels (e.g., *fällt*_{3SG} vs. *fallt*_{2PL}) to predict the number of the upcoming subject?
- RQ2: Is the prediction effect influenced by language status (L1 vs. L2)?
- RQ3: Is the prediction effect influenced by prediction awareness?
- RQ4: Is the prediction effect influenced by working memory capacity?

We measured prediction in a twofold manner, using the participants' button-press reaction times and their eye gaze data. Note that while RQs 1–3 were analyzed using both measures, the analysis of RQ4 was based uniquely on the reaction-time data (for full details, see Analysis).

We expected the L1 group to engage in predictive processing, given the sensitivity of native speakers to number mismatches in subject–verb agreement (e.g., Wagers et al., 2009). Because of the reduced saliency of the target structure, we expected these prediction effects to be small. In the L2 group, we only expected very limited prediction effects despite their advanced proficiency levels, given the difficulty of the vowel change for this population. This would also be in line with theories that claim form-function mappings in the L2 to be incomplete and/or weaker than in the L1, causing slower processing (e.g., weaker links hypothesis, Gollan et al., 2008; Missing Surface Inflection Hypothesis, Prévost & White, 2000). Note that in the analyses, we only included strong verbs that the participants were able to produce correctly during a conjugation task. Overall, we hypothesized a facilitatory role of awareness (e.g., Curcic et al., 2019), and a positive correlation between prediction and working memory (see Huettig & Janse, 2016).

Methods

Participants

Thirty-one L1 speakers of German (age range 18–45 years; mean age 30 years; 23 identified as female, eight as male) and 30 L2 learners of German with L1 Dutch (age range 18–50 years; mean age 27 years; 22 identified as female, eight as male), residing in Brussels, Belgium, participated in our study. All signed an informed participation consent form and were given monetary compensation. All participants were students or adults with academic degrees. None of them had hearing problems, dyslexia, or color vision deficiency. All participants spoke several other foreign languages besides German, in particular English and French, and Dutch in the L1 group. Most L2 speakers had started to learn German in high school or at university. On average, their first intensive encounter with the German language was at the age of 16.47 years ($SD = 6.88$); however, there was quite some variation across participants (range 1–47 years, or 6–20 years when disregarding two outliers). Their mean score on the German version of the LexTALE (www.lextale.com; see Lemhöfer & Broersma, 2012), an untimed lexical decision task that can be used to approximate general German proficiency, was 73.67% ($SD = 10.31\%$). This score corresponds to the CEFR level B2 upper-intermediate and differed significantly from the L1 group ($M = 90.40\%$, $SD = 4.18\%$; $W = 867$, $p < .001$, $r = -.74$).

During the post-experiment interviews, all L2 participants reported being familiar with the phenomenon of the vowel-changing strong verbs. Almost all reported having learned about this through formal language instruction in the past. Most of them had the impression that they knew the strong paradigm well, though they also found it effortful to apply the vowel change correctly and reported this to be item-dependent, with frequent verbs being known best. Two participants had learned the paradigm exclusively via immersion and reported mastering it rather well.

Stimuli

The main task of the experimental session was a picture-matching task with visual world eye tracking, combining visual and auditory stimuli (available at <https://www.iris-database.org/details/QCXMA-3yDNr>). The target structure was the stem-vowel alternation that occurs in German strong verbs in the simple present tense (see above; see Table 1): some of the auditory stimulus sentences contained strong verbs that were either inflected in 3SG (e.g., *fäll-t*) or 2PL (e.g., *fall-t*), that preceded the subject, and whose stem vowels (here: /ɛ/ vs. /a/) provided a potential predictive cue for subject number. A brief story introduced the task: a group of blue and green aliens has landed on Earth and started doing all kinds of comical things; a girl named Anna observes them and describes with surprise what she sees. Each trial consisted of two pictures that depicted the aliens carrying out actions, combined with a sentence in German – Anna’s statements – presented auditorily and matching only one of the pictures. The participants were instructed to listen to Anna and to indicate the target picture as fast and accurately as they could by pushing a button on a game controller, without turning their gaze away from the screen. Both the eye movement recordings and the reaction-time latencies of the button presses would later be used to measure prediction.

All items of the task consisted of German verbs, embedded in grammatical sentences describing the aliens’ actions. An items spreadsheet can be found in the supplementary materials of this paper, accessible through <https://osf.io/vjcmz/>. There were three item classes: strong items (32 German strong verbs), weak items (32 German weak verbs), and filler items (a mixture of 32 weak, strong, and irregular verbs). We selected verbs that were preferably highly frequent or cognate with Dutch to facilitate comprehension in the L2 group (90% of the 96 verbs were cognates). All verbs were inflected either in 3SG or 2PL PRES and thus always received the *-t* suffix (see Table 1). Only the strong verbs in 3SG would undergo a stem-vowel change in addition to the suffix, allowing the strong verbs’ stem vowel to become a predictive number cue. The weak and filler items could not be used predictively (weak items, because their 3SG and 2PL forms are fully identical, and filler items, because number was never contrasted in the picture pairs of filler trials); these items would not be analyzed and mainly served the purpose of making our prediction-related research aim less obvious.

The onsets of the auditory and visual stimuli were always synchronous. Figure 1 shows how the trials were presented (videos illustrating the course of trials in real time are available in the supplementary materials: <https://osf.io/vjcmz/>). All pictures were developed for the purpose of the experiment and always showed one or two

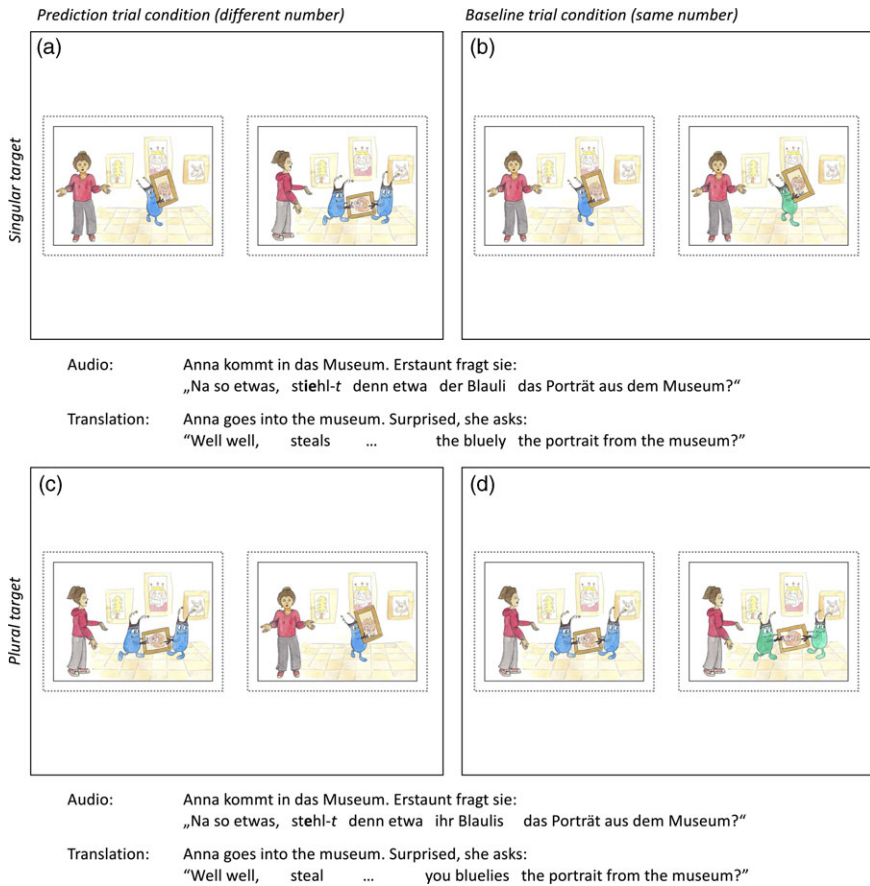


Figure 1. Screenshots illustrating the presentation of trials.

Note: In all screenshots, the left picture is the target. The verb's stem vowel represents a predictive number cue in prediction trials (a, c) but not in baseline trials (b, d; in these examples, alien color is blue on the left picture and green on the right picture). The dotted rectangles indicate interest areas.

aliens carrying out the actions referred to by the verb. The pictures had dimensions of 300×400 pixels and were contoured by black rectangles. The target and distractor pictures were aligned horizontally on a $1,024 \times 768$ pixels screen, with a 112 pixels distance between them. The target and distractor interest areas had dimensions of 350×450 pixels and were thus slightly larger than the corresponding pictures. All fixations within these interest areas would be registered as looks toward the target or distractor picture.

All audio stimuli were built up in the same way: they had a duration of 9,000 ms, and all subcomponents had the same duration across trials (see Figure 2). They started with an introduction that served as 4,500 ms of preview time, and part of which was Anna's exclamation of surprise, enabling her next utterance to be formulated as a question with VSO word order (instead of the default SVO order in German declarative sentences). The verb thus always preceded the subject, enabling

Time frame		Audio sentence	Translation	Onset	Offset	Duration
PREVIEW	Introduction	Anna kommt in das Restaurant. Erstaunt fragt sie:	Anna enters the restaurant. Surprised, she asks:	0 ms	3420 ms	4500 ms
	Expression of surprise	„Na so etwas,	“Well well,	3420 ms	4500 ms	
VERB	Verb	isst	eats _{3SG PRES}	4500 ms	5150 ms	1500 ms
	Modal particle	denn etwa		5150 ms	6000 ms	
SUBJECT	Subject NP	der Blauli	the bluely	6000 ms	6900 ms	900 ms
END	Object NP	die Spaghetti mit den Fingern?“	the spaghetti with his hands?“	6900 ms	9000 ms	2100 ms

Figure 2. Structure and timing of the auditory stimulus sentences.

us to test whether participants would exploit the verbs’ morphology to anticipate subject number, or if they would wait for the subject NP to select the target. To facilitate measurement, we broadened the interval between verb and subject NP by inserting the German modal particle *denn etwa*, which emphasizes astonishment.

The subject always referred to the blue and green aliens; solely alien color (blue vs. green) and syntactic number and person (3SG vs. 2PL) varied systematically in order to create different trial conditions (see below). Blue aliens were called *Blaulis* (“bluelies”; 3SG: *der Blauli-Ø*; 2PL: *ihr Blauli-s*), and the green ones were called *Grünlis* (“greenlies”; 3SG: *der Grünli-Ø*; 2PL: *ihr Grünli-s*). *Der* is the 3SG determiner (nominative, masculine); *ihr* is the 2PL personal pronoun nominative, combined with the referent noun in apposition. Both *der* and *ihr* thus carry number information, as does the noun’s suffix ($-\emptyset_{SG}$ or $-s_{PL}$). On pictures involving one alien, Anna would look toward the participant, as if talking *about* this alien (using 3SG); on pictures involving multiple aliens, she would look toward them, as if talking *to* them (justifying the use of 2PL).

All sound recordings were spoken by the experimenter, a female High German native speaker. She applied a rather slow pace (roughly 3.3 syllables per second; see Fernandez *et al.*, 2020) and stopped briefly after each sentence subcomponent (see Figure 2) to avoid coarticulation. All subcomponents were edited to have exactly the same duration across all trials; care was taken to remove all gaps between the different subcomponents within each sentence.

Main task design

All trials involving strong items systematically varied in the number of the target referent (SG vs. PL), and in the number displayed on the two pictures (same vs. different number). There were thus four possible trial types: (a) different-number pictures, SG target; (b) different-number pictures, PL target; (c) same-number pictures, SG target; and (d) same-number pictures, PL target (see Figure 1). In same-number trials, the only parameter differing between both pictures was alien color.

All 32 strong items were only shown once to each participant, in one of the four trial types. This was counterbalanced between participants, using a Latin square rotation: if an item was shown to one participant in trial type *a*, the next participant would get the same item in type *b*, etcetera. Target alien color (blue and green) and target picture location (left and right) were counterbalanced over the four trial types.

To measure predictive processing, we distinguished two main trial conditions. Different-number trials were the *prediction trial condition*, in which the verb's stem vowel represented the first reliable cue for target subject number. Same-number trials were the *baseline trial condition*: as number was not contrasted, the number information encoded in the verb was not informative; instead, participants were obliged to wait for the subject NP, providing the disambiguating color cue. We used a pseudo-random trial order: trials appeared in a new, random order for each participant, but there could never be two strong items in a row.

Familiarization task

An online, untimed familiarization task, taking place one week prior to the experiment and taking about 45 minutes to complete (see Procedure), introduced the participants to the materials of the picture-matching task. Our goal was to prevent any difficulties at the level of comprehension or interpretation during the picture-matching task, thereby maximally reducing the potential impact of external factors on processing and eye movements; furthermore, it also served as a production premeasure of the test verbs. The participants saw 96 pictures, one per verb and presented one by one. The pictures always showed a singular alien carrying out an action and also included the German labels of the verbs and objects (see Figure 3). The participants' task was to write brief descriptions of each picture in German, using the labels (but no further linguistic aids).



Figure 3. Sample visual stimulus of the familiarization task.

Operation span task

During the experimental session in the laboratory, we administered the operation span task by Klaus and Schriefers (2016, www.socsci.ru.nl/memory/) as a measure of working memory capacity. The online test had a duration of about 12 minutes. Instructions were provided in the participants' L1. The participants judged the correctness of mathematical equations while memorizing series of two to six digits, presented to them in 14 blocks of two to six trials. Cronbach's alpha for the 14 test blocks was .77. The mean of the percentage of correctly recalled digits per test block represented the participants' operation span score (see Zheng & Lemhöfer, 2019).

Awareness interview

We administered an extensive, carefully structured retrospective interview immediately after the second picture-matching task, following the guidelines of Rebuschat (2013). The protocol is available in the supplementary materials (<https://osf.io/vjcmz/>). The experimenter started by asking general questions about how the participants had experienced the eye tracking task. She also invited the participants to report what they thought the experiment was about, and whether they had used any strategies. She then asked increasingly specific questions about language, first inquiring about the general role of grammar in the experiment, and eventually asking questions about verb form and its predictive function. The participants of the L2 group were also asked to provide information about their prior knowledge of German strong verb conjugation.

The main purpose of this interview was to assess the status of *prediction awareness* (see Curcic *et al.*, 2019): participants would be categorized as *aware* or *unaware*, depending on whether or not they had consciously realized that they could focus on the strong verbs' morphology to determine the target picture in different-number trials, without having to wait for the subject NP.

Oral conjugation task

During the oral conjugation task at the session end, participants conjugated all 64 strong and weak verbs aloud in 3SG PRES, and we explicitly asked to produce the stem vowels as accurately as possible. After each oral production, participants were prompted to provide confidence ratings and source attributions (adapted from Rebuschat, 2013): they had to indicate how certain they were that they had produced the right stem vowel (not certain, a little certain, very certain) and which knowledge source they had relied on (guess, intuition, explicitly learned knowledge). All strong verbs that were produced with an incorrect stem vowel or for which participants had indicated "guess" and "not certain" were considered *unknown* and excluded from the final dataset for these particular participants. This was necessary because such verbs could not be expected to be used predictively. Note that even if participants were unable to produce the verbs accurately, they still may have some implicit, passive knowledge of the target forms. In that sense, a production measure to exclude unknown items is a rather conservative way of assessing knowledge of the target structure.

Procedure

All participants were first familiarized with the experimental materials through the online familiarization task that they did from home. To hide the experiment's actual focus on morphosyntactic processing, we portrayed our study as a pilot project targeted at assessing the usability of newly developed drawings as language learning materials for children.

One week after familiarization, we tested the participants individually in a quiet lab. The main tasks of the experimental session were two picture-matching tasks with visual world eye tracking, which were always administered in the same order, and each of which took approximately 25 minutes to complete. Both tasks shared the same design, yet they differed in their target structures. The first task focused on prediction based on German weak verb inflection by contrasting 3SG and 3PL PRES and investigating whether the regular suffixes *-t* and *-en* could be used predictively. The second task is the picture-matching task described above, which had the stem-vowel alternation in strong verbs as its target structure. In this paper, we focus uniquely on the second task; the first task (for a full report, see Koch et al., 2021) is only reported on to the extent necessary to understand and interpret the second task. The first task revealed robust, suffix-based prediction effects; moreover, these were found to be of an aware and strategic kind (also see subsection Awareness, under Results).

Between the picture-matching tasks, we administered the operation span task, the LexTALE, and a background survey.

The second picture-matching task started with five practice trials and a 13-point calibration, after which participants completed 96 experimental trials. These were always preceded by drift checks, and there were brief breaks after every 24 trials. The mean accuracy of target picture identification on test trials was 98.99% in the L1 group ($SD = 1.87\%$, range = 93.75%–100%) and 98.44% in the L2 group ($SD = 2.43\%$, range = 90.63%–100%); we can thus be confident that there were no comprehension problems. Given that the second picture-matching task was preceded by the first picture-matching task, the participants' experience with the first task could be assumed to have shaped their expectations regarding the second task. Keep in mind, however, that although verb form still contained a number cue, the target structure from the first task no longer provided predictive information. Care was taken to prevent any learning effects between the two tasks: throughout the experimental session, a participant would never encounter the same verb twice in SG or twice in PL, and never twice in the prediction condition or twice in the baseline condition.

The second picture-matching task was immediately followed by the retrospective awareness interview. The oral conjugation task concluded the session. The full session duration was approximately 100 minutes. We audio-recorded the interviews and the conjugation task.

Analysis

A laptop-mount EyeLink Portable Duo (SR Research Ltd., 2016) registered the participants' eye movements with a sampling rate of 500 Hz. Each sample was coded as

either looks toward the target picture, looks toward the distractor picture, or as falling outside of these regions. The original dataset contained 1,952 trials, but we excluded all incorrect responses (25 trials removed) and, subsequently, any unknown strong items (an additional 156 trials removed), resulting in a data loss of 9.27% and a dataset consisting of 1,771 trials.

We assessed predictive processing by means of two outcome variables: the proportion of looks toward the target picture (computed exclusively across all samples assigned to the target or distractor picture regions, excluding samples during blinks or saccades)¹, and reaction-time latency of the button-press responses (RT; measured in milliseconds from trial onset until button press). As for the eye movement data, prediction of subject number based on the verb's stem vowel was operationalized as a higher proportion of looks toward the target picture in prediction trials than in baseline trials during the prediction time frame. This time frame is the critical interval between verb onset and subject NP onset (see Figure 2), ranging from 4,710 to 6,219 ms (corrected for 10 ms of audio delay caused by the experiment computer, plus 200 ms for the time needed to plan and execute eye movements; Saslow, 1967). Prediction at the level of the RTs was operationalized in a different manner, independently from the prediction time frame, as faster button presses in prediction trials than in baseline trials (comparable to Curcic *et al.*, 2019). While the operationalization of eye-gaze-based prediction matches the strict definition of prediction (i.e., effects measured before target onset, see above), our measurement of RT-based prediction can strictly speaking be seen as facilitation, since our measurement does not depend on target onset (i.e., effects occurring before or after target onset would be taken into account). We analyzed the eye movement data and the RT data separately by means of two different but complementary analyses in R (R Core Team, 2018; data and scripts are available in the supplementary materials: <https://osf.io/vjcmz/>).

Analysis of the eye movement data

For the eye movement analyses, we excluded an additional 29 trials for which more than 50% of the gaze data were looks outside the interest areas or missing (see Curcic *et al.*, 2019). For the remaining trials, the average of samples contributing per trial was 85.97% ($SD = 4.75\%$). We analyzed these data by means of cluster-based permutation analysis (Maris & Oostenveld, 2007; for examples using eye tracking, see Andringa, 2020; Barr *et al.*, 2014; Guerra *et al.*, 2021) with the *EyetrackingR* package (Dink & Ferguson, 2015). This method allows to identify time intervals (so-called *time clusters*) during which participants are significantly more likely to look at the target picture in prediction trials than in baseline trials (or vice versa). If a significant time cluster signaling more target looks in prediction trials overlaps with the prediction time frame, this represents a prediction effect, and the cluster's range informs us about its timing. In step one of this analysis, the participant-averaged gaze data are subdivided into time bins of 50 ms, and for each bin, paired sample *t* tests are run to test the difference in the proportion of target looks in prediction versus baseline trials. This results in the identification of potential clusters of adjacent bins for which a significant difference was found. In step two, 2,500 iterations of randomly shuffling the labels of the conditions being tested,

repeating step one, and comparing the resulting clusters among each other are performed, ultimately resulting in probabilities indicating whether or not the clusters identified in step one are significant or due to chance.

We performed seven cluster-based permutation analyses. For analysis 1, we used the full dataset and the procedure just described to establish whether there was a prediction effect in the sample as a whole (RQ1). To test if prediction depended on language status (RQ2), we directly compared the L1 and L2 groups by computing for each 50 ms time bin the proportion of target looks in prediction trials minus the proportion of target looks in baseline trials, and using the resulting difference score as dependent variable for the *t* tests (analysis 2). In addition, we repeated the procedure from analysis 1 in the L1 and L2 groups separately (analyses 3 and 4). To establish whether prediction was influenced by awareness (RQ3), we directly compared the unaware and aware groups by again using the difference between the proportion of target looks in prediction and baseline trials as dependent variable (analysis 5), and we also repeated analysis 1 in the unaware and aware groups separately (analyses 6 and 7).

Analysis of the reaction-time data

We analyzed the RTs of the button presses by means of linear mixed-effects modeling, using the *lme4* package (Bates et al., 2015; version 1.1-23). We used treatment-coded contrasts and restricted maximum likelihood for model estimation; effects were reported as significant at $p < .05$. The model took the following form: (reaction time) \sim 1 + Trial condition*Target + Trial condition*Group + Trial condition*Working memory + Trial condition*Awareness*Trial number + (Trial condition|Participant) + (1|Item). The term left of the tilde (\sim) is the outcome variable; the terms at its right are the model terms. The number 1 represents an intercept. As fixed effects, the model included the within-subject factors Trial condition (baseline, prediction) and Target (singular, plural), the between-subjects factors Group (L1, L2) and Awareness (unaware, aware), and the continuous variables Working memory and Trial number. The asterisks mark interactions. The terms including the bar character (|) are the random effects. In modeling the variance components, we tried to balance type I and type II errors by fitting the maximal random structure (as recommended by Barr et al., 2013), but only if it is supported by the data (see e.g., Matuschek et al., 2017). We started with a model containing crossed random intercepts for participants and items and included theoretically motivated additional variance components that both yielded a stable (i.e., converging) model and significantly improved model fit (as indicated by the Akaike information criterion and the likelihood ratio test). The resulting model included random intercepts over participants and over items (thus, the model took into account variation based on individual differences among subjects and item-level differences), as well as random slopes for participants depending on Trial condition.

We included all fixed effects for theoretical reasons, regardless of their contributions to model fit. Trial condition was the key variable for assessing prediction (RQ1); its respective two-way interactions with Group, Awareness, and Working memory allowed us to establish whether the prediction effect differed

in the L1 and L2 groups (RQ2), whether it depended on awareness (RQ3), and whether it was modulated by working memory (RQ4). Trial number was included to assess potential changes in the course of the experiment. As for Target, this variable reflects whether the target referent was singular (picture with one alien + verb in 3SG) or plural (picture with two aliens + verb in 2PL). This factor needed to be included to account for potential effects of verb form (e.g., 3SG might be processed more easily than 2PL for having a higher frequency in general language use) and of picture complexity (i.e., more complex pictures have been found to attract more fixations than less complex pictures; e.g., Schlerer, 2019).

Results

Awareness

The interviews confirmed that the participants' expectations regarding the second picture-matching task were shaped by the experience they had had with the first task during which all of them had engaged in strategic processing: at trial onset in task one, they generally scanned the two pictures for their differences (number, color, objects, and setting or action) and subsequently waited for the element in the sentence that would provide the disambiguating cue. This also means that all of them had developed prediction awareness of the verbs' suffixes, and they perceived this as relatively easy. For the second task, they expected verb form to again provide a reliable cue in different-number trials; however, almost half of the participants reported that they soon thought that the verb no longer provided a cue "because singular and plural now had identical forms" – not realizing at all that the verb could still be used predictively under specific circumstances. The other half did develop prediction awareness of the stem vowels, though they generally reported that this was more difficult and took more effort to apply than during the first task. We categorized 26 participants as unaware (14 L1, 12 L2) and 35 participants as aware (17 L1, 18 L2).

Eye movements

Table 2 summarizes the results of the cluster-based permutation analyses. Analysis 1 revealed a significant cluster in the sample as a whole, ranging from 5,300 to 6,750 ms (for visualization, see Figure 4). This effect had a positive directionality, indicating there were significantly more looks toward the target picture in prediction trials than in baseline trials. As this cluster overlapped with the prediction time frame (4,710–6,210 ms), we can be confident that it reflects a prediction effect in the full sample (RQ1), starting 800 ms after verb onset and spilling over into the subject time frame (6,000–6,900 ms; see Figure 2). In baseline trials, participants started looking more toward the target around 6,400 ms (Figure 4).

Comparable significant positive clusters representing prediction effects (with spillovers into the subject) were found in the L1 and L2 groups separately (RQ2; Table 2: analyses 3 and 4; see Figure 5). Although the effect started earlier in the L1 (5,300–6,800 ms) than in the L2 group (5,550–6,650 ms), this 250 ms difference

Table 2. Outcomes of the cluster-based permutation analyses

Cluster	Sum statistic	Range (ms)	Probability (p)
Analysis 1 (Full sample)			
1	2.76	400–450	.830
2	172.26	5,300–6,750	.000
3	–9.29	4,750–4,950	.462
Analysis 2 (L1 vs. L2)			
1	2.43	150–200	.864
2	2.10	2,200–2,250	.927
3	2.04	8,300–8,350	.951
4	–9.15	3,500–3,700	.494
Analysis 3 (L1 only)			
1	2.23	400–450	.897
2	154.63	5,300–6,800	.000
3	2.15	6,900–6,950	.922
4	6.75	8,200–8,350	.638
Analysis 4 (L2 only)			
1	8.72	2,800–3,000	.471
2	4.20	3,550–3,650	.804
3	2.46	5,400–5,450	.831
4	84.31	5,550–6,650	.000
5	–6.38	4,650–4,800	.634
Analysis 5 (Unaware vs. Aware)			
1	15.31	2,500–2,800	.228
2	–6.46	8,900–8,950	.689
Analysis 6 (Unaware only)			
1	74.75	5,800–6,650	.000
2	–2.09	2,550–2,600	.957
3	–13.01	4,800–5,000	.299
Analysis 7 (Aware only)			
1	2.03	400–450	.942
2	4.56	1,250–1,350	.744
3	19.33	2,550–2,950	.163
4	146.89	5,300–6,750	.000
5	–4.49	7,350–7,450	.755
6	–4.13	8,900–9,000	.818

Note: Significant time clusters are marked in bold. Dependent variable of analyses 1, 3, 4, 6, and 7: proportion of looks toward the target picture, with factor Trial condition (prediction, baseline). Dependent variable of analyses 2 and 4: difference between the proportion of looks toward the target picture in prediction and baseline trials, with factors Group (L1, L2) and Awareness (unaware, aware), respectively.

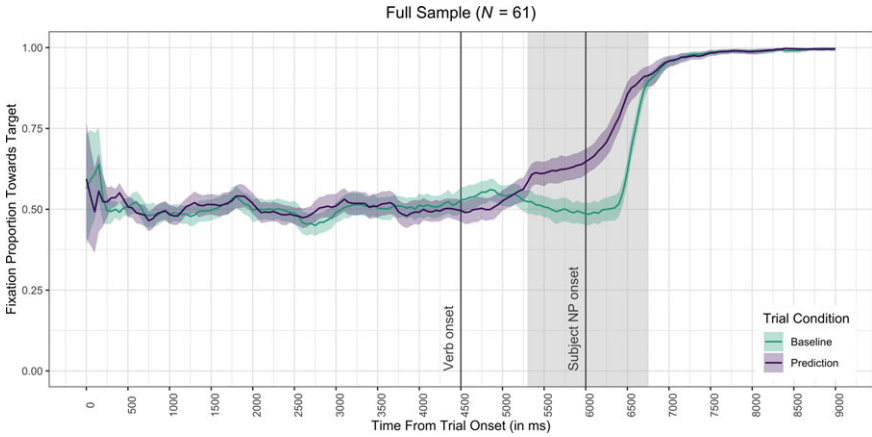


Figure 4. Proportion of looks toward the target picture unfolding over time in the full sample. *Note:* Error bands show 95% confidence intervals. Shaded areas indicate time clusters during which there was a significant difference between prediction and baseline trials.

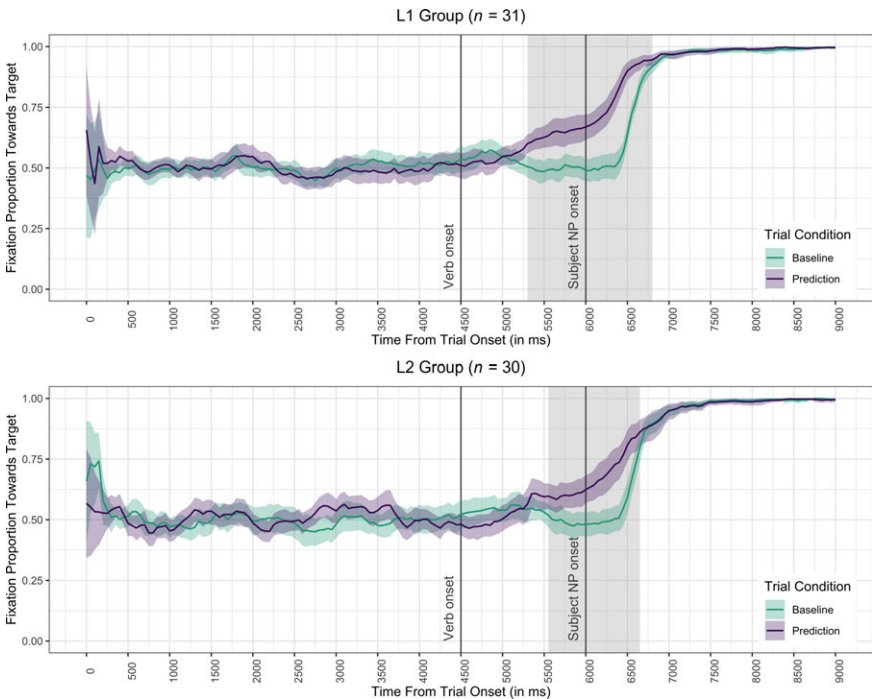


Figure 5. Proportion of looks toward the target picture unfolding over time in the L1 and L2 groups. *Note:* Error bands show 95% confidence intervals. Shaded areas indicate time clusters during which there was a significant difference between prediction and baseline trials.

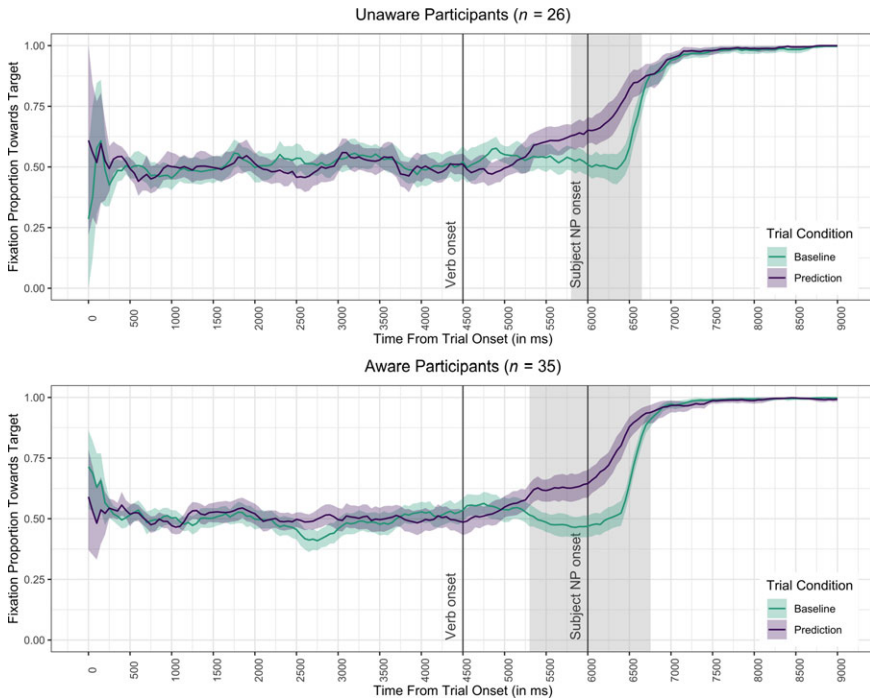


Figure 6. Proportion of looks toward the target picture unfolding over time in unaware and aware participants.

Note: Error bands show 95% confidence intervals. Shaded areas indicate time clusters during which there was a significant difference between prediction and baseline trials.

was not significant according to analysis 2 that directly compared both groups (for visualization, see Supplementary Figure S1 in the supplementary materials).

Similarly, significant clusters reflecting prediction were found in the unaware and aware groups separately (RQ3; Table 2: analyses 6 and 7; see Figure 6). While prediction started 500 ms earlier in the aware (5,300–6,750 ms) than in the unaware group (5,800–6,650 ms), this difference was not significant according to analysis 5 which directly compared the unaware and aware participants (also see Supplementary Figure S2).

Reaction times

The RT model output, which can be found in Supplementary Table S1, revealed three significant parameters: the two-way interactions between Trial condition and Group ($p = .009$) and Trial condition and Working memory ($p = .019$), and the three-way interaction between Trial condition, Awareness, and Trial number ($p = .004$). To interpret effects involving categorical variables, we used estimated marginal means (see Table 3) and pairwise comparisons (least square means method). We examined 10 planned contrasts (see Table 4) and applied the Benjamini–Hochberg correction for multiple testing (Benjamini & Hochberg,

Table 3. Estimated means based on the reaction-time mixed-effects model

	RTs (ms)	
	Baseline trials	Prediction trials
Across Target, Group, and Awareness	6,950 [6,862, 7,038]	6,735 [6,614, 6,855]
Target: Singular	6,935 [6,841, 7,029]	6,730 [6,606, 6,855]
Target: Plural	6,966 [6,872, 7,059]	6,738 [6,613, 6,864]
Group: L1	6,869 [6,750, 6,988]	6,551 [6,384, 6,717]
Group: L2	7,031 [6,908, 7,155]	6,919 [6,748, 7,089]
Awareness: Unaware	6,973 [6,843, 7,103]	6,823 [6,642, 7,003]
Awareness: Aware	6,928 [6,817, 7,039]	6,647 [6,492, 6,801]

Note: Estimated means and 95% confidence intervals by Trial condition, by Trial condition and Target, by Trial condition and Group, and by Trial condition and Awareness.

Table 4. Pairwise comparisons among estimated means, based on the reaction-time mixed-effects model

Contrast	MD (ms)	SE	t	p
Overall				
Baseline vs. Prediction	216 [140, 291]	38	5.74	<.001
Target				
Singular: Baseline vs. Prediction	204 [116, 293]	45	4.57	<.001
Plural: Baseline vs. Prediction	227 [138, 316]	45	5.06	<.001
Baseline-prediction difference Singular vs. baseline-prediction difference Plural	23 [-73, 118]	49	0.47	.640
Group				
L1: Baseline vs. Prediction	319 [215, 422]	52	6.19	<.001
L2: Baseline vs. Prediction	113 [3, 223]	55	2.05	.045
Baseline-prediction difference L1 vs. baseline-prediction difference L2	206 [55, 358]	76	2.72	.009
Awareness				
Unaware: Baseline vs. Prediction	150 [34, 266]	58	2.60	.012
Aware: Baseline vs. Prediction	281 [185, 378]	48	5.85	<.001
Baseline-prediction difference Unaware vs. baseline-prediction difference Aware	131 [-20, 282]	75	1.74	.087

Note: Estimated mean differences (MD) and their 95% confidence intervals (squared brackets); p-values in bold print are significant after a correction for multiple testing. Contrasts were specified for Trial condition (baseline vs. prediction) overall, and at each level of Target, Group, and Awareness. We also compared the mean difference between baseline and prediction trials between the two levels of Target, Group, and Awareness, respectively.

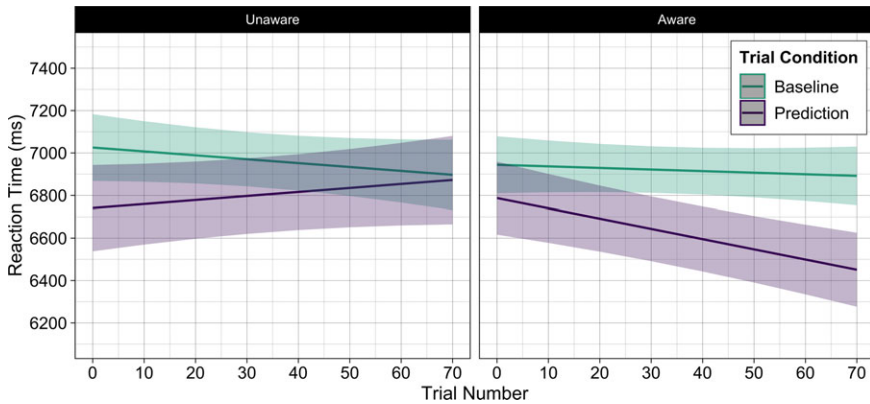


Figure 7. Reaction time as a function of Trial condition, Awareness, and Trial number.

1995), which involves using an adjusted significance threshold for p -values based on a 5% false discovery rate and the number of contrasts.

The pairwise comparisons (Table 4) revealed a significant prediction effect in the full sample (RQ1): the model estimated that overall, participants made button-press responses that were on average 216 ms faster on prediction trials than on baseline trials (see Table 3, for the estimated average RTs). The prediction effect was not further modulated by Target: Table 4 shows that the prediction effect was present and similar within singular-target and plural-target trials.

The significant interaction between Trial condition and Group shows that prediction was different in the L1 and L2 groups (RQ2). This is confirmed by the significance of the contrast between the prediction effects (i.e., the estimated RT difference between baseline and prediction trials) in both groups (Table 4). Moreover, whereas the estimated difference between prediction and baseline trials in the L1 group (319 ms) was significant, this was not the case for the L2 group (113 ms; note that p was .045, but this was not significant after the correction for multiple testing).

There was no significant interaction between Trial condition and Awareness (RQ3). Both the aware and unaware groups showed a significant prediction effect (see Table 4). Although the aware participants' estimated prediction effect was on average 131 ms larger than that of the unaware participants, this difference was not significant. The significant interaction between Trial condition, Awareness, and Trial number indicates that Trial number significantly modulated the RTs, but only on prediction trials and only in the aware group (see Figure 7): under those conditions, the model estimated the RTs to become 5 ms faster with every trial. Figure 7 illustrates this: while RTs on baseline trials remain steady in the aware group, those on prediction trials become faster. The plot for the unaware group suggests that participants engaged in predictive processing only at task outset, but that this effect vanished over time because responses became slower on prediction trials but faster on baseline trials.

The significant interaction between Trial condition and Working memory indicates that prediction was modulated by working memory capacity (RQ4). While

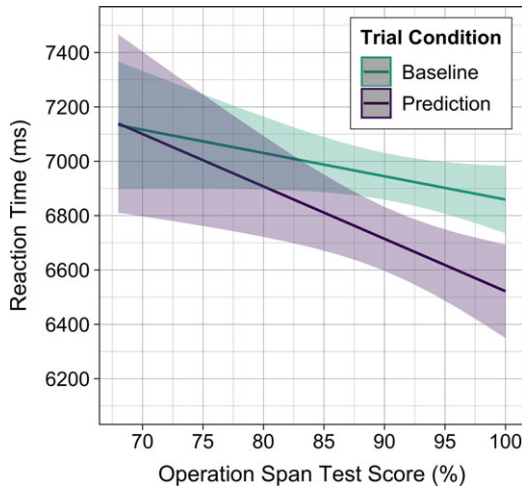


Figure 8. Reaction time as a function of Trial condition and Working memory.

Working memory did not significantly influence RT on baseline trials, the model estimated that on prediction trials, every unit increase of the operation span test score was accompanied by a 19-ms reduction in RTs. Thus, the higher the participants working memory capacity, the faster they were at responding on prediction trials (see Figure 8).

Discussion

The present study investigated the extent to which L1 and advanced L2 speakers of German exploit the stem allomorphy in strong verbs to anticipate subject number. The analyses revealed a significant prediction effect in the full sample (RQ1): in the eye gaze data, this effect started 800 ms after verb onset; in the RT data, the effect manifested itself as button-press responses that were 216 ms faster in prediction trials than in baseline trials. Comparable RT-based prediction effects were present in both singular-target trials and plural-target trials, indicating that both the changed and unchanged stems were used predictively. In the gaze data, there was significant prediction in both language groups, though this effect started 250 ms earlier in the eye movements of the L1 group than in the L2 group. In the RT data, the estimated difference between baseline and prediction trials was significantly larger (by 206 ms) in the L1 group than in the L2 group. In fact, the difference in RTs between baseline and prediction trials was only significant in the L1 group but not in the L2 group. Prediction awareness modulated prediction, but only to a certain extent (RQ3). Significant prediction effects in the gaze and RT data were present in both the unaware and aware groups. In the aware group, prediction in the gaze data started 500 ms earlier than in the unaware group. The prediction effect in the RTs was 131 ms larger in the aware group than in the unaware group, but this difference was not significant. Throughout the experiment, the aware but not the unaware group became faster at responding

on prediction trials. Working memory positively influenced prediction (RQ4): the higher the participants' working memory capacity, the faster they were at responding on prediction trials.

L1 and L2 morphosyntax-based prediction

Our study extends the evidence of morphosyntax-based L1 and L2 predictive processing to a new target structure. So far, the most common target structure in this domain has been gender marking (e.g., for L1 processing: Dahan et al., 2000; Huettig & Guerra, 2019; for L2 processing: Hopp & Lemmerth, 2018; Morales et al., 2016); thus, it is interesting to see that prediction also extends to number information processing, and to a target structure that is of similar or even greater difficulty than gender markings because of its low saliency and informative redundancy. Our finding of successful predictive L1 processing is in accordance with the fact that German has a relatively rich morphological system and that its users even rely on subtle morphosyntactic cues for smooth, anticipatory processing.

Although no (or, at best, only a marginal) prediction effect was found in the RT data of the L2 group, the eye gaze data did reveal a significant prediction effect. To date, only a limited number of studies has found positive evidence for morphosyntax-based prediction in adult L2 sentence processing (e.g., Curcic et al., 2019; Trenkic et al., 2014), and some studies found negative evidence (e.g., gender: Lew-Williams & Fernald, 2010; case: Hopp, 2015; Mitsugi, 2017). That we found prediction in L2 speakers is therefore noteworthy, given the learning and processing difficulties that strong conjugation represents in L2 learners, particularly when their L1 is Dutch (Koch et al., 2023). The learners' advanced proficiency level potentially played an important role in this respect (see Grüter et al., 2017; Hopp, 2013); perhaps the results would have looked different in learners with lower proficiency. Furthermore, despite the hypothesized blocking effects in this population (see above), the typological proximity between German and Dutch may have played in the learner group's favor. It may be interesting to replicate this study with learners whose L1 is typologically less related to the target language, as for instance Swedish, which does not have subject–verb agreement and whose speakers may thus be less inclined to rely on the grammatical number cues.

A comparison between the L1 and L2 groups revealed that prediction started slightly earlier in the gaze data of the L1 group; in the RT data, a clear prediction effect could only be found in the L1 but not the L2 data. This L1–L2 difference is in line with prior research (e.g., Chun & Kaan, 2019; Grüter et al., 2012) and likely is a consequence of overall slower lexical access, increased cognitive load (see Kaan, 2014) and less stable mental representations of the different stem variants in the L2 group. Recall that all participants had been familiarized extensively with the experimental materials and that we had only kept those items in the dataset for which the participants demonstrated accurate conjugation knowledge, ruling out a lack of comprehension of the stimulus materials in the learner group as a potential cause for the L1–L2 differences.

It remains important to keep in mind that our findings are based on two different outcome measures. Eye movements are a covert, automatic response type, reflecting online interpretive processes, while button presses are an overt, controlled response

that occur after the interpretive processes and the additional operations following them (e.g., action or response planning; see Ayasse *et al.*, 2017). Furthermore, as the eye-gaze-based prediction effects in our data were observable before the target onset, they correspond to a strict definition of prediction; our RT-based prediction effects, however, were observable only after the target onset (which can be assumed to be a direct consequence of the additional steps of controlled processing that are involved in the button-press responses), and some may prefer to refer to such effects as facilitation.

Prediction and awareness

Aware and unaware predictive processing

Our study provides evidence that morphosyntax-based predictive processing can happen spontaneously and unconsciously. Nonetheless, while prediction is often tacitly assumed to be an unconscious, automatic process and while time-sensitive measures such as EEG or eye tracking are often associated with natural, unaware processing, our findings also demonstrate that L1 and L2 prediction can happen consciously and strategically (see Andringa, 2020; Curcic *et al.*, 2019). The significant interaction between trial condition, awareness, and trial number underscores this strategic behavior: responses became faster on prediction trials throughout the task, but only for aware participants. Thus, future prediction research should not take unaware processing for granted and integrate careful debriefing to assess awareness.

Keep in mind that we coded awareness in a binary way, thus applying strict cut-offs to categorize data that might in reality be of a much more complex nature. That is, perhaps there was variation with respect to the moment in time in which the participants became aware, or perhaps some aware participants found it more difficult to use the stem vowels for prediction than others. We did not quantify such nuances.

As for prediction in the unaware group, the trends in Figure 7 might provide further insight: the model estimated prediction in the unaware group to only occur early in the task; after that, RTs become faster on baseline trials and slower on prediction trials. Possibly, the unaware participants first tried to strategically focus on the verbs to identify the target in different-number trials, just like they had done in the first task. This actually led to successful predictive processing, but without reaching these participants' consciousness. After that, participants dropped this strategy that they believed to no longer work in the second task; they thus started to pay less attention to the verb and to only focus on the subject instead, and processing was no longer predictive. This may also explain why RTs on baseline trials became faster. What is interesting is that prediction, in this context, seems to be linked to attention to form, without implying awareness. Though such an interpretation is based on the visual inspection of trends and not on significant results, it would be in line with what the unaware participants reported during the interviews.

Limited effect of awareness on prediction

Our study suggests that awareness had a limited, facilitatory effect on prediction, which is in line with prior research (Brothers *et al.*, 2017; Huettig & Guerra, 2019). It would be

interesting to investigate how awareness and language status interact; however, our subgroups were too small to integrate such an interaction in our analyses. Nonetheless, visual inspection of the eye gaze data plotted over time (see Appendix S1 in the supplementary materials) shows predictive trends within both the aware and unaware subgroups of the L1 and L2 groups. This would suggest that our findings deviate from those of Curcic et al. (2019) and Andringa (2020), who found predictive L2 processing – which they took as evidence for L2 learning – to fully depend on the emergence of awareness. Note, however, that their studies were learning experiments involving an artificial miniature language and thus no prior knowledge of the target structure, while we investigated L2 prediction based on existing linguistic knowledge. The divergence between our findings and theirs might indicate that the role of awareness for prediction also depends on the stage of knowledge development: possibly, awareness only plays a decisive role for prediction at the beginning stages of learning; in L1 processing or at more advanced stages of L2 learning, knowledge is (more) consolidated and allows for automatized predictive processing, leaving less room for awareness to play a role.

As we had expected a more important effect of awareness, we studied the unaware group in more detail. Note that this group also included 11 participants (6 L1, 5 L2) who recalled having noticed the presence of vowel-changing strong verbs in the task, without realizing they could be used predictively. To rule out the possibility that prediction in the unaware group was contingent on whether or not participants noticed the target structure, we compared the *noticing* and *non-noticing* subgroups. Visual inspection of the eye movement data (see Appendix S2 in the supplementary materials) revealed predictive looking behavior in the unaware non-noticing subgroup, too, both within L1 and L2 speakers; in fact, prediction seemed slightly stronger in the non-noticing than in the noticing subgroup. In the unaware L2 subgroup, noticing the strong verbs appeared to be linked to a RT slowdown (average RTs were even slightly slower on prediction trials than on baseline trials). In the unaware L1 subgroup, noticing seemed to have a small beneficial effect. On the basis of these explorative observations of descriptive statistics, we tentatively conclude that prediction in the unaware group was not dependent on noticing.

What predicted awareness?

When discussing why some participants were aware while others were not, it is important to keep in mind that the picture-matching task reported on here was always administered 15 minutes after the first picture-matching task (see Koch et al., 2021). The participants' experience of the second task was thus affected by their experience of the first task. Note that this task-order effect affected *all* participants in a similar way, as confirmed by the interviews; thus, it does not jeopardize the validity of the results. Also note that counterbalancing task order was not a suitable alternative. The 3SG–2PL contrast in the second picture-matching task was more unusual than the 3SG–3PL contrast in the first task: when we describe pictures, we typically use a third and not a second person to refer to the agents visible in the pictures. If the experimental session had started with the 3SG–2PL contrast, participants could have thought during the remainder of the session that plural-agent pictures corresponded not only to 3PL but also to 2PL in the audio. This might

Table 5. Means and standard deviations (between parentheses) for potential predictors of awareness

	L1		L2	
	Unaware <i>n</i> = 14	Aware <i>n</i> = 17	Unaware <i>n</i> = 12	Aware <i>n</i> = 18
Operation span score	89.70 (8.62)	88.82 (9.43)	91.29 (9.84)	93.12 (6.93)
LexTALE	90.71 (3.17)	90.15 (4.94)	69.69 (11.52)	76.32 (8.76)
Target structure knowledge	98.89 (1.98)	98.66 (2.31)	68.24 (25.12)	88.93 (8.76)

have compromised cue reliability in the suffix-based prediction task, because weak verbs in 3SG and 2PL share identical forms. Exposing all participants first to picture-matching task 1 and subsequently to task 2 prevented such undesired confounding effects.

Considering that the entire sample had engaged in aware predictive processing in the first picture-matching task, it is noteworthy that only half of our participants developed prediction awareness during the second picture-matching task. The context in which participants performed the second task was thus clearly an awareness-inducing one. Both picture-matching tasks were almost identical and only differed in their target structure, suggesting that (un)awareness was closely linked to the target structure's degree of difficulty. As pointed out earlier, the suffixes of the regular, productive weak paradigm were perceived as easy, while the vowel change of the subregular, unproductive strong paradigm required more effortful processing in both participant groups. In the L2 group, participants had robust knowledge of the regular suffixes and positive transfer facilitated prediction, while their knowledge of the strong paradigm was less robust, and transfer was negative.

A question that arises is why some participants developed awareness, while others did not. We therefore performed a post hoc exploratory analysis, using binomial logistic regression. As potential predictors of awareness, we considered working memory and the LexTALE scores (as an index of German proficiency). To this list, we added *target structure knowledge*: this was the average percentage of correctly produced strong verbs (i.e., with application of the correct stem vowel) during the (written) familiarization task, which should serve as an indicator of how well participants mastered the target structure. Table 5 provides the descriptive statistics. The model outcomes (for details, see Appendix S3) revealed that none of the variables predicted awareness in the L1 group, suggesting the cause for the emergence of awareness in this group lies elsewhere. A potential and tentative explanation is that different levels of metalinguistic awareness in both subgroups may have played a role here.

In the L2 group, target structure knowledge but not the other variables significantly predicted awareness ($p = .044$), with mean accuracy scores of 89% in the aware subgroup against only 68% in the unaware subgroup. This finding suggests that the emergence of awareness was associated with a deep level of knowledge of the target structure. One possibility is to interpret this association in the context of theories of L2 learning, in which awareness represents a central topic. More traditional models of learning assume that learning starts with explicit, conscious forms of knowledge, and that practice subsequently leads to more implicit, unconscious

forms of knowledge at more advanced stages of development (e.g., Skill Acquisition Theory; DeKeyser, 1998). However, that we found an association between more robust forms of knowledge and awareness does not directly fit into such a model of learning; rather, it fits Cleeremans' Radical Plasticity Thesis (Cleeremans, 2007, 2011), which assumes a reversed order of the stages of L2 knowledge development: learners first acquire implicit/unconscious, weak representations; then, based on input and experience, these representations start becoming more robust, of improved quality, and explicit/conscious, involving access to aware inspection of the knowledge; ultimately, knowledge becomes automatic. Thus, awareness is not seen as a starting point for learning, but as a sign of knowledge having reached a robust stage, which would apply to our L2 sample. Alternatively, it may also be the case that participants with better target structure knowledge simply had more opportunities to predict throughout the experiment than participants with less extensive knowledge. Assuming that each prediction trial of an item for which a participant knows the target conjugation increases the odds of participants becoming aware of the predictive cue, this may explain why the emergence of awareness depended on target structure knowledge.

Prediction and working memory

The facilitatory effect of higher working memory capacity on predictive button presses is in line with prior research (Favier et al., 2021; Huettig & Janse, 2016). Recall that some studies (Chun, 2020; Chun & Kaan, 2019) did not find an effect of working memory on semantic prediction. A tentative explanation of why working memory appears to modulate morphosyntactic but not semantic prediction could be that semantic prediction is achieved with much more "ease," leaving no room for working memory to be of influence.

Our finding applies to the sample overall; due to limited (sub)sample size, we could not integrate interactions between working memory and awareness or language status in our model. Nonetheless, our post hoc exploratory analyses suggest that working memory scores were similar in the aware and unaware subgroups of the L1 and L2 groups. This is interesting, given that working memory is often associated with explicit learning aptitude (e.g., Li, 2015, 2016): on such an account, working memory could be expected to determine the emergence of awareness, yet this was not the case for our data. We invite future research to further explore the precise relationship between working memory and prediction.

Conclusions

This study extends the positive evidence for morphosyntax-based predictive processing in L1 speakers and L2 learners to stem-vowel alternations in German strong verbs, a new target structure with high difficulty due to subregularity, redundancy, and low salience. Both native speakers and learners used the strong verbs' stem vowels to predict upcoming subject number, though this effect remained limited in the learner group. About half of the participants spontaneously developed prediction awareness and thus engaged in strategic, controlled processing – in the L2 group, this depended on how robust their knowledge of the target structure was.

The other half remained unaware, meaning prediction happened unconsciously. Awareness facilitated prediction, but only to a limited extent. Working memory capacity had a facilitatory effect on predictive RTs.

Our study suggests a clear reliance on subtle morphosyntactic cues during online processing in L1 speakers and, to some extent, also in advanced learners of German, a language with a rich morphosyntactic system. Future prediction research may further explore how language status, awareness, working memory, and other individual differences (e.g., processing speed; Huettig & Janse, 2016) interact, preferably with an increased sample size that allows for the analysis of complex interactions and/or subgroup analysis. Moreover, our experiment could easily be combined with learning treatments (see Cintrón-Valentín & Ellis, 2015) or instruction manipulations (e.g., Huettig & Guerra, 2019).

Replication package. For supplementary material (appendices, supplementary tables and figures, videos illustrating experimental trials) accompanying this paper, visit <https://osf.io/vjcmz/>. This OSF site also provides research materials allowing replication (data, analysis code, interview protocol). The stimulus materials (audio recordings and pictures) are available via the IRIS database: <https://www.iris-database.org/details/QCXMA-3yDNr>.

Acknowledgments. This study was conducted within the framework of Eva Koch's doctoral research, supported by an FWO (Fonds Wetenschappelijk Onderzoek, Flanders) grant awarded to Alex Housen and Aline Godfried. The authors would like to thank Marije Michel for her excellent idea to use aliens as referents in the experimental stimuli, Florian Hintz for his advice regarding the general study design, and Falk Huettig for his recommendation to increase the length of the critical prediction time frame. Finally, the authors thank all reviewers for their highly valuable feedback and suggestions for improvement.

Conflict of interest. The authors declare none.

Ethical approval. This project was approved by the Ethical Committee in Humane Sciences (Vrije Universiteit Brussel) on the 4 December 2014.

Notes

1 Any eye movements beyond the button press were recoded as looks towards the target picture. This was done because the post-response gaze data no longer reliably indicate sentence processing: the task of target picture identification has already been completed and is no longer relevant. Instead of discarding these data, we recoded them in a meaningful way.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4), 502–518. <https://doi.org/10.1016/j.jml.2006.12.004>
- Andringa, S. (2020). The emergence of awareness in uninstructed L2 learning: A visual world eye tracking study. *Second Language Research*, 36(3), 335–357. <https://doi.org/10.1177/0267658320915502>
- Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, 28(4), 525–560. <https://doi.org/10.1080/01690965.2012.658072>
- Ayasse, N. D., Lash, A., & Wingfield, A. (2017). Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. *Frontiers in Aging Neuroscience*, 8, 1–12. <https://doi.org/10.3389/fnagi.2016.00329>

- Barr, D. J., Jackson, L. & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, *143*(1), 404–413. <https://doi.org/10.1037/a0031813>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://arxiv.org/abs/1406.5823>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. <https://doi.org/10.1016/j.jml.2016.10.002>
- Brown, V. A., Fox, N. P., & Strand, J. F. (2021). “Where are the . . . fixations?”: Grammatical number cues guide anticipatory fixations to upcoming referents and reduce lexical competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0001019>
- Bybee, J., & Newman, J. (1995). Are stem changes as natural as affixes? *Linguistics*, *33*, 633–654. <https://doi.org/10.1515/ling.1995.33.4.633>
- Chun, E. (2020). L2 prediction guided by linguistic experience. *English Teaching*, *75*, 79–103. <https://doi.org/10.15858/engtea.75.s1.202006.79>
- Chun, E., & Kaan, E. (2019). L2 prediction during complex sentence processing. *Journal of Cultural Cognitive Science*, *3*(2), 203–216. <https://doi.org/10.1007/s41809-019-00038-0>
- Cintrón-Valentín, M., & Ellis, N. C. (2015). Exploring the interface: Explicit focus-on-form instruction and learned attentional biases in L2 Latin. *Studies in Second Language Acquisition*, *37*, 197–235. <https://doi.org/10.1017/S0272263115000029>
- Cleeremans, A. (2007). Consciousness: The radical plasticity thesis. In R. Banerjee & B. K. Chakrabarti (Eds.), *Progress in brain research* (Vol. *168*, pp. 19–33). Elsevier. [https://doi.org/10.1016/S0079-6123\(07\)68003-0](https://doi.org/10.1016/S0079-6123(07)68003-0)
- Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology*, *2*, 1–12. <https://doi.org/10.3389/fpsyg.2011.00086>
- Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*, 42–71. <https://doi.org/10.1111/lang.12321>
- Dahan, D., Swingle, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480. <https://doi.org/10.1006/jmla.1999.2688>
- DeKeyser, R. M. (1998). Beyond focus on form. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge University Press.
- DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*, 1–25. <https://doi.org/10.1111/j.0023-8333.2005.00294.x>
- Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism*, *20*(5), 917–930. <https://doi.org/10.1017/S1366728916000547>
- Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881–900. <https://doi.org/10.1080/23273798.2019.1591469>
- Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis*. <http://www.eyetracking-r.com>
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, *35*(2), 353–387. <https://doi.org/10.1017/S0272263112000915>
- Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194. <https://doi.org/10.1093/applin/aml015>

- Favier, S., Meyer, A. S., & Huettig, F. (2021). Literacy can enhance syntactic prediction in spoken language processing. *Journal of Experimental Psychology: General*. Advance online publication. <http://dx.doi.org/10.1037/xge0001042>
- Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G. & Allen, S. E. M. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, 73(12), 2348–2361. <https://doi.org/10.1177/1747021820948019>
- Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, 38, 177–215. <https://doi.org/10.1017/S0272263115000388>
- Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge. <https://www.routledge.com/Eye-Tracking-in-Second-Language-Acquisition-and-Bilingualism-A-Research/Godfroid/p/book/9781138024670>
- Godfroid, A., & Uggem, M. S. (2013). Attention to irregular verbs by beginning learners of German. *Studies in Second Language Acquisition*, 35, 291–322. <https://doi.org/10.1017/S0272263112000897>
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814. <https://doi.org/10.1016/j.jml.2007.07.001>
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28(2), 191–215. <https://doi.org/10.1177/0267658312437990>
- Grüter, T., Rohde, H., & Schafer, A. J. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In W. Orman & M. J. Valteau (Eds.), *Proceedings of the 38th annual Boston university conference on language development* (pp. 179–191). Cascadia Press. <https://par.nsf.gov/servlets/purl/10028988>
- Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, 7(2), 199–229. <https://doi.org/10.1075/lab.15011.gru>
- Guerra, E., Bernotat, J., Carvacho, H., & Bohner, G. (2021). Ladies first: Gender stereotypes drive anticipatory eye-movements during incremental sentence interpretation. *Frontiers in Psychology*, 12, 1–10. <https://doi.org/10.3389/fpsyg.2021.589429>
- Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience*, 32(10), 1229–1249. <https://doi.org/10.1080/23273798.2017.1327080>
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29, 33–56. <https://doi.org/10.1177/0267658312461803>
- Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, 53(3), 277–306. <https://doi.org/10.1515/iral-2015-0014>
- Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171–199. <https://doi.org/10.1017/S0272263116000437>
- Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118–135. <https://doi.org/10.1016/j.brainres.2015.02.014>
- Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, 1706(November 2018), 196–208. <https://doi.org/10.1016/j.brainres.2018.11.013>
- Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, 31(1), 80–93. <https://doi.org/10.1080/23273798.2015.1047459>
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. <https://doi.org/10.1016/j.actpsy.2010.11.003>
- Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, 21(2), 251–264. <https://doi.org/10.1017/S1366728917000050>
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 4(2), 257–282. <https://doi.org/10.1075/lab.4.2.05kaa>

- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8)
- Karaca, F., Brouwer, S., Unsworth, S., & Huettig, F. (2021). Prediction in bilingual children: The missing piece of the puzzle. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 116–137). John Benjamins. <https://doi.org/10.1075/bpa.12.06kar>
- Klaus, J., & Schriefers, H. (2016). Measuring verbal working memory capacity: A reading span task for laboratory and web-based use. <https://doi.org/10.31219/osf.io/nj48x>
- Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, *5*(1), 115–132. <https://doi.org/10.22599/jesla.79>
- Koch, E. M., de Vos, J. F., Housen, A., Godfroid, A., & Lemhöfer, K. (2023). Learning second language morphosyntax in dialogue under explicit and implicit conditions: An experimental study with advanced adult learners of German. *Bilingualism: Language and Cognition*, 1–14. <https://doi.org/10.1017/S1366728922000797>
- Köpcke, K. M. (1998). Prototypisch starke und schwache Verben der deutschen Gegenwartssprache. *Germanistische Linguistik*, *141–142*, 45–60.
- Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*(1), 1–25. https://doi.org/10.1207/s15473341lld0201_1
- Krause, H., Bosch, S., & Clahsen, H. (2015). Morphosyntax in the bilingual mental lexicon. *Studies in Second Language Acquisition*, *37*, 597–621. <https://doi.org/10.1017/S0272263114000564>
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195395518.003.0065>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Leow, R. P., & Hama, M. (2013). Implicit learning in SLA and the issue of internal validity. *Studies in Second Language Acquisition*, *35*, 545–557. <https://doi.org/10.1017/S027226311300003X>
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. <https://doi.org/10.1111/j.1467-9280.2007.01871.x>
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*, 447–464. <https://doi.org/10.1016/j.jml.2010.07.003>
- Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408. <https://doi.org/10.1093/applin/amu054>
- Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*(4), 801–842. <https://doi.org/10.1017/S027226311500042X>
- Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, *146*, 349–370. <https://doi.org/10.1016/j.cognition.2015.10.012>
- Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(2), 113–139. <https://doi.org/10.1007/s41809-019-00035-3>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>

- Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research*, 5(1), 1–10. <https://doi.org/10.16910/jemr.5.1.3>
- Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics*, 38(4), 953–983. <https://doi.org/10.1017/S0142716416000515>
- Morales, L., Paolieri, D., Dussias, P. E., Valdés Kroff, J. R., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism: Language and Cognition*, 19(2), 294–310. <https://doi.org/10.1017/S1366728915000176>
- Peters, R. E., Grüter, T., & Borovsky, A. (2018). Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied Psycholinguistics*, 39(6), 1439–1469. <https://doi.org/10.1017/S0142716418000383>
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002–1044. <https://doi.org/10.1037/bul0000158>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190. <https://doi.org/10.1017/S0140525X04000056>
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–33. <https://doi.org/10.1191/026765800677556046>
- R Core Team. (2018). *R: A language and environment for statistical computing* [Computer software]. Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63(3), 595–626. <https://doi.org/10.1111/lang.12010>
- Reuter, T., Dalawella, K., & Lew-Williams, C. (2021). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience*, 36(4), 474–490. <https://doi.org/10.1080/23273798.2020.1839665>
- Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. *Frontiers in Psychology*, 6, 1–11. <https://doi.org/10.3389/fpsyg.2015.00590>
- De Ruiter, A. J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515–535. <https://www.jstor.org/stable/4490203>
- Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, 57(8), 1030–1033. <https://doi.org/10.1364/JOSA.57.001030>
- Schlechter, J. (2019). *Predictive language processing in late bilinguals: Evidence from visual-world eye-tracking* [Doctoral dissertation, University of Potsdam]. Publication Server of the University of Potsdam. <https://doi.org/10.25932/publishup-43249>
- Shook, A., Goldrick, M., Engstler, C., & Marian, V. (2015). Bilinguals show weaker lexical access during spoken sentence comprehension. *Journal of Psycholinguistic Research*, 44, 789–802. <https://doi.org/10.1007/s10936-014-9322-6>
- Simoens, H., Housen, A., & De Cuypere, L. (2017). The effect of perceptual salience on processing L2 inflectional morphology. In S. M. Gass, A. Mackey, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (1st ed., pp. 107–130). Routledge. <https://doi.org/10.4324/9781315399027>
- SR Research Ltd. (2016). *EyeLink Portable Duo* [Apparatus and software]. <https://www.sr-research.com/eyelink-portable-duo/>
- Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism*, 17(2), 237–257. <https://doi.org/10.1017/S1366728913000321>
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(3), 443–467. <https://doi.org/10.1037/0278-7393.31.3.443>
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237. <https://doi.org/10.1016/j.jml.2009.04.002>

- Weber, A., & Broersma, M.** (2012). Spoken word recognition in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5368–5375). John Wiley & Sons. <https://doi.org/10.1002/9781405198431>
- Williams, J. N.** (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–353). Emerald Press.
- Zheng, X., & Lemhöfer, K.** (2019). The “semantic P600” in second language processing: When syntax conflicts with semantics. *Neuropsychologia*, *127*, 131–147. <https://doi.org/10.1016/j.neuropsychologia.2019.02.010>

Cite this article: Koch, EM., Bulté, B., Housen, A., and Godfroid, A. (2023). The predictive processing of number information in subregular verb morphology in a first and second language. *Applied Psycholinguistics* *44*, 750–783. <https://doi.org/10.1017/S014271642300022X>