Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment

EVA M. KOCH (ID)

BRAM BULTÉ (ID)

ALEX HOUSEN

ALINE GODFROID (ID)

Author affiliations can be found in the back matter of this article

WHITE ROSE
UNIVERSITY PRESS
Universities of Leeds, Sheffield & York

ABSTRACT

We investigated whether adult German native speakers and Dutch-speaking second-language learners of German exploit German regular verb morphology for predictive sentence processing and whether such predictive processing is moderated by working memory capacity and awareness of the predictive cue. In a picture-matching task with visual-world eye-tracking, the participants (first-language group: $n = 31$; second-language group: $n = 30$) saw two pictures of action scenes, varying in the number of referents depicted (singular versus plural). Simultaneously, they heard sentences in German in which the verb suffix represented the first reliable cue for grammatical number of the upcoming subject noun phrase. Successful exploitation of this number cue was measured as anticipatory button-press reaction times and eye-movements toward the correct picture, before subject onset. We found significant prediction effects in both the first-language and the second-language group, with anticipatory processing being somewhat faster in the native speakers than in the learners. Faster reaction times on prediction trials were associated with higher working memory scores. Debriefing interviews revealed that all participants had become aware that they could use the verb form predictively, adding further evidence that prediction can be a conscious process.

CORRESPONDING AUTHOR:
Eva M. Koch

Vrije Universiteit Brussel, Brussels, Belgium

koch.eva.marie@gmail.com

1. INTRODUCTION

As a fundamental element of human cognition, prediction also plays an important role in language comprehension. When listeners hear a sentence in their native language (L1), they continuously predict upcoming input, enabling fast and incremental language comprehension. Although such predictive language processing does not occur all the time (e.g., Huettig & Guerra, 2019), it facilitates efficient communication (Kutas et al., 2011). L1 listeners have been found to make predictions about upcoming information based on a vast range of linguistic cues (for reviews, see Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018), such as word semantics or grammatical information (e.g., case or gender markings, verb tense). Such predictive processing appears to be further modulated by various individual differences, including vocabulary size (Rommers et al., 2015) and working memory capacity (Huettig & Janse, 2016).

Given calls to view L1 and second-language (L2) acquisition and processing not as separate phenomena but as part of a larger continuum (e.g., Ortega, 2009), it is important to establish the commonalities and potential differences between L1 and L2 predictive processing. As comprehension processes in L2 listening are usually slower, more cognitively demanding and less accurate than in L1 listening (e.g., Weber & Broersma, 2012), one might expect a reduced ability to make such predictions (cf. RAGE hypothesis in Grüter et al., 2017). At the same time, a growing body of research has found L2 learners to successfully employ lexical-semantic cues for predictive sentence processing, while studies using morphosyntactic cues have produced mixed outcomes (for a review, see Godfroid, 2020). More research is therefore needed to address and better understand the different factors that influence the extent to which L2 learners engage in predictive language processing, especially in the area of morphosyntax.

Using visual-world eye-tracking (a method during which participants look at pictures while listening to auditory stimuli), we investigated prediction in L1 and L2 speakers for a hitherto understudied grammatical structure, namely grammatical number encoded in German regular verb morphology. As target verb forms, we used the third-person singular (3SG) and plural (3PL) in the simple present tense (PRES). In sentences such as *Lach-$t_{3SG\ PRES}$ der_{SG} Blauli-$\emptyset_{SG}$?/Lach-$en_{3PL\ PRES}$ die_{PL} Blauli-s_{PL}?* ("Laughs the bluely?/Laugh the bluelies?"), the verb's suffixes *-t* and *-en* provide a predictive number cue for the upcoming subject noun phrase (NP). We compared prediction in advanced L2 learners and L1 speakers of German, expecting somewhat reduced prediction effects in the learners, and we also explored the roles of working memory capacity and awareness in prediction.

2. BACKGROUND

2.1. PREDICTIVE L1 AND L2 PROCESSING

An increasing number of studies has investigated whether and how prediction in L2 listeners differs from L1 predictive processing (Godfroid, 2020). Such studies have generally found prediction effects in the L2 to be weaker compared to L1 processing (e.g., Grüter et al., 2012) or even entirely absent (e.g., Hopp, 2015). Nonetheless, prediction can be seen as being fundamentally similar in L1 and L2 processing (Kaan, 2014), as several factors may account for these L1-L2 processing differences: Prediction in the L2 could be slowed down because of slower lexical access (Shook et al., 2015), a reduced quality of lexical representations, or a stronger impact of frequency effects (Kaan, 2014). Moreover, enhanced activation of competing interlingual information in the L2 may increase cognitive load, leaving fewer cognitive resources available for prediction (Kaan, 2014; Weber & Broersma, 2012).

L2 prediction may further depend on the level of language that is being processed. There is now extensive evidence that learners can generate predictions at the semantic level (e.g., Dijkgraaf et al., 2017). At the morphosyntactic level, however, the findings are mixed. While there is ample evidence for the predictive use of case and gender markings in L1 processing (e.g., Dahan et al., 2000; Huettig & Guerra, 2019; Mitsugi & MacWhinney, 2016), several studies failed to find morphosyntax-based prediction in L2 learners (e.g., Hopp, 2015; Lew-Williams & Fernald, 2010). Yet, positive evidence for L2 predictive gender processing has started to accumulate (e.g., Curcic et al., 2019; Hopp & Lemmerth, 2018), also shedding light on important modulating

learner-related factors, such as L1 transfer effects (e.g., Morales et al., 2016) and proficiency (cf. Grüter et al., 2017). In several studies (e.g., Hopp, 2013; Hopp & Lemmerth, 2018), only learners with a near-native L2 mastery were found to exploit gender markings predictively.

2.2. THE ROLE OF WORKING MEMORY AND AWARENESS IN PREDICTION

More generally, our understanding of what influences predictive language processing is still developing. A first potential factor that we focus on is working memory capacity, which was found to be a strong predictor for the predictive processing of gender-marked determiners in an individual-differences study by Huettig and Janse (2016), testing 105 Dutch native speakers by means of visual-world eye-tracking. Such a strong correlation between working memory and L1 prediction has possible implications for L2 processing. As the cognitive resources for prediction in the L2 may be reduced (see above; Kaan, 2014), the impact of working memory may be even stronger in L2 predictive processing. Chun and Kaan (2019) explored this topic in a visual-world prediction experiment in which they increased cognitive load by using complex stimulus sentences. Both English L1 and advanced L2 speakers successfully exploited semantic cues for prediction, though the effect started 180 ms later in the L2 group. The authors assessed the link between L2 prediction and working memory but against their expectations, they did not find a correlation, possibly due to small subsample size ($n = 23$). More replication research is therefore needed to confirm the findings of Huettig and Janse (2016) and to establish the precise relationship between working memory capacity and predictive L1 and L2 processing.

A second variable of interest is awareness, defined here as whether or not participants are conscious of the fact that specific linguistic cues may allow them to anticipate upcoming input. It still seems rather uncommon for studies to report if their participants were aware of the experiment's prediction goal. Such information, however, could potentially elucidate the extent to which prediction involves automatic and controlled processing (cf. Huettig, 2015). Being aware of specific cues in the input enables listeners to make predictions in a conscious and controlled manner; being unaware may mean that predictive processing happens unconsciously and automatically. In a visual-world experiment, Curcic et al. (2019) explored the roles of cognitive aptitude components for the anticipatory processing of gender-marked determiners in an artificial L2. Although the participants' cognitive aptitudes did not influence prediction directly, they were related to the participants' development of prediction awareness, which in turn played an important facilitatory role for prediction. Using a miniature language based on Esperanto, Andringa (2020) found the predictive usage of determiners marking distance and animacy to fully depend on the emergence of prediction awareness. In an EEG experiment using written stimuli, Brothers et al. (2017) manipulated awareness by comparing the semantic prediction of sentence-final words in two tasks that differed in their instructions. While one task was a mere comprehension task, the other task explicitly invited the participants to anticipate the final word. Prediction was stronger in the latter task. Huettig and Guerra (2019) conducted a visual-world experiment in which they investigated the effects of the presence/absence of the explicit instruction to predict, speech rate (slow versus normal), and preview time (4 seconds versus 1 second) on the L1 predictive processing of Dutch gender-marked determiners. Under conditions of a normal speech rate and short preview, prediction occurred only when participants were instructed to predict and therefore made aware of the predictive cues in the input. In the absence of such an instruction, prediction remained absent. The authors' general conclusion was that the conditions under which language processing takes place can be decisive for prediction. In sum, these studies suggest that the presence of prediction awareness can enhance predictive processing; in the case of L2 morphosyntactic processing, prediction even seems to fully depend on awareness, at least during the initial stages of learning.

2.3. VERB NUMBER MARKINGS AS A PREDICTIVE CUE

As pointed out above, morphosyntax-based L2 prediction research has produced mixed findings, and we aim to help resolve this state of affairs by expanding the range of target structures investigated. An interesting, yet scarcely explored morphosyntactic cue for prediction is inflectional verb morphology, and more specifically verb number marking. L1 speakers are expected to use verb number markings predictively, as research found them to be sensitive to subject-verb number agreement mismatches (e.g., Wagers et al., 2009). In the relevant

visual-world studies that investigated the L1 predictive use of verb number markings, test trials generally contrasted one single-object picture and one multiple-objects picture, combined with an auditory sentence in which the morphology of the copula *be*, inflected in 3SG or 3PL PRES, provided the first cue for a singular-plural distinction. In K ouider et al. (2006), for instance, 20-to-36-month-old children heard English sentences containing pseudowords, such as *Look, there **is a** blicket/Look, there **are some** blicket**s***. Toddlers as young as 24 months started to make anticipatory looks to the correct picture as soon as they heard the verb. Riordan et al. (2015), however, could not find any predictive number processing in their eye-tracking study that exposed English-speaking adults to sentences such as *There **is** the lion/There **are** the lion**s***. The authors suggest that the absence of prediction was a consequence of English inflected verbs providing a redundant number cue that is less reliable than those provided in NPs. However, their critical time frame during which prediction should have occurred was possibly too short (approximately 150 ms) to reliably measure prediction. Lukyanenko and Fisher (2016) measured prediction by comparing informative trials (the verb provided a number cue: *Where **are** the good cookie**s**?*) to uninformative trials (the referent was not preceded by an agreeing verb: *Do you see the good cookie**s**?*). Adults as well as 3-year-olds were found to use the agreeing verbs in informative trials predictively.

To our knowledge, Schlenter (2019) is the only study looking into L2 verb number processing. Using visual-world eye-tracking, the author tested 28 adult native speakers of German and 25 learners (upper B2 – C1 Common European Framework of Reference for Languages [CEFR] levels; L1 Russian), using German sentences such as *In the mailbox **is/are** at noon **a/several** newspaper/newspaper**s*** (translation with German word order). Although a visualization of the looking data showed predictive trends in both groups, no significant prediction effects were found. This result needs to be interpreted with caution given that the experimental design only contained informative trials, and no uninformative trials to compare them with.

Taken together, existing research on the predictive use of verb number markings is scarce, and almost exclusively limited to L1 English studies using the copula. Because English has a rather reduced morphosyntactic agreement system, its speakers may be generally less predisposed to rely on morphological cues than speakers of morphologically rich(er) languages. Furthermore, the copula tends to be highly irregular and saliently marks the singular-plural distinction through suppletion (English: *is/are*; German: *ist/sind*). Results therefore may not generalize to other forms of number marking, such as when number is expressed by verb affixes, which are less salient and more difficult to process and learn than suppletion (Krause et al., 2015). This highlights a need to build on this research area with more studies using morphologically richer languages that allow for different ways of encoding number agreement; moreover, research should also place the focus on L2 processing.

2.4. THE PRESENT STUDY

The following research questions (RQs) guided our investigation:

- RQ1: Do L1 and L2 speakers use the morphosyntactic information encoded in the suffixes of German weak verbs to predict grammatical number of the upcoming subject? To what extent is the prediction effect – if present at all – similar in the L1 and L2 groups?
- RQ2: Is the prediction effect influenced by working memory capacity?
- RQ3: Is the prediction effect influenced by participants' awareness status?

As there is ample evidence that native speakers can use morphosyntactic structures predictively, we expected the L1 group to process the verbs' suffixes predictively. As for the L2 group, inflectional morphology in general represents a source of persistent difficulty for learners, both at the level of production and comprehension (DeKeyser, 2005). This might decrease the chances for predictive processing of the number information in the verbs. Yet, the weak verb conjugation can be hypothesized to be relatively easier for learners to acquire due to its regularity and productivity (Godfroid, 2016). Moreover, Dutch-speaking learners of German might benefit from positive transfer, since both languages share the same inflectional suffixes for 3SG and 3PL PRES. In line with prior findings, we expected any prediction effects in the L2 group to start later than in the L1 group. As for working memory, we expected a positive correlation with prediction (cf. Huettig & Janse, 2016). Moreover, we expected stronger prediction in participants who developed prediction awareness than in participants who remained unaware.

3. METHOD
3.1. PARTICIPANTS

The participants were 30 advanced L2 learners of German whose native language was Dutch (mean age = 26.47, *SD* = 9.46; 22 females), and 31 L1 German speakers (mean age = 29.77, *SD* = 6.52; 23 females). All participants had good hearing, normal or corrected-to-normal vision, and no history of dyslexia or color-blindness. All were university students or adults with a higher education degree, living in Brussels or Flanders, Belgium. As a proxy of the participants' general German proficiency, we administered the German version of the LexTALE (*www.lextale.com*; cf. Lemhöfer & Broersma, 2012), a non-speeded lexical decision task. The L2 group's average score was 73.67% (*SD* = 10.31, *range* = 50.00–96.25), which corresponds to a B2 upper intermediate CEFR level (Lemhöfer & Broersma, 2012) and which differed significantly from the L1 group (*M* = 90.40, *SD* = 4.18, *range* = 81.25–97.50; *W* = 866.5, *p* < .001, *r* = –.74).

3.2. TARGET STRUCTURE

The German conjugation system distinguishes between weak, strong, and irregular verbs. In this study we focus on weak conjugation, a regular, productive paradigm in which affixation encodes morphosyntactic information of number, person, tense and mood (**Table 1**). As target forms, we selected 3SG and 3PL PRES. When juxtaposed, these verb forms (e.g., *lach-**t**$_{3SG\ PRES}$* versus *lach-**en**$_{3PL\ PRES}$*) represent a minimal pair, where the suffix is the only element providing disambiguating information about syntactic number. All participants in the L2 group had extensive prior knowledge of German weak conjugation because they had studied the paradigm in their German language courses.

LACH-*EN* ("LAUGH")	
1SG	lach-*e*
2SG	lach-*st*
3SG	**lach-*t***
1PL	lach-*en*
2PL	lach-*en*
3PL	**lach-*en***

Table 1 Examples of German weak verb conjugation in present indicative tense.

Note: Verb forms used in the present experiment are marked in bold.

3.3. PROCEDURE

The research project was presented to the participants as a pilot study aimed at assessing pictures to be used as language learning materials for children. This cover story was meant to conceal the study's actual focus on grammar. Approximately one week after completing an online familiarization task (see 3.3.1.), participants were tested individually in a quiet room with consistent illumination. Informed consent was obtained prior to the experimental session, and all participants received 20 euros in reward. The session lasted approximately 90–105 minutes, and its main components were two separate eye-tracking experiments using picture-matching tasks. The session started with the first eye-tracking experiment (3.3.2.), which we report here and which focused on suffix-based prediction. This was immediately followed by the first part of a debriefing interview. The participants subsequently took an operation span test (3.3.3.), the LexTALE and a language background questionnaire. They then completed the second eye-tracking experiment, followed by the second part of the interview (3.3.4.). The second eye-tracking experiment is not reported here but represents the main focus of a separate though related study (Koch et al., 2021). It shared its overall design with the first eye-tracking experiment, but focused on German strong verbs,[1] a closed verb class that encodes

---

1   In experiment 2, we contrasted pairs such as *g**i**b-t*$_{3SG}$ versus *g**e**b-t*$_{2PL}$, whereby only the stem vowel, and not the suffix, provides a predictive number cue – yet of a very reduced phonetic salience.

number partially through stem-vowel alternations, and which presented a contrasting scenario to the experiment reported here. We provide supplementary materials on the methods and results (appendices, data, scripts) through *https://osf.io/37rfb/*.

3.3.1. Familiarization

The online familiarization task took approximately 45 minutes and introduced the participants to the visual stimuli of the eye-tracking experiment. It was designed to prevent any interpretation or comprehension difficulties during the eye-tracking experiment. The participants saw 96 pictures (one per item) that were presented successively. The pictures contained the German labels of the verbs corresponding to the depicted action, and the labels of the objects in the scenes (*Figure 1*). The instruction was to write brief picture descriptions in German.
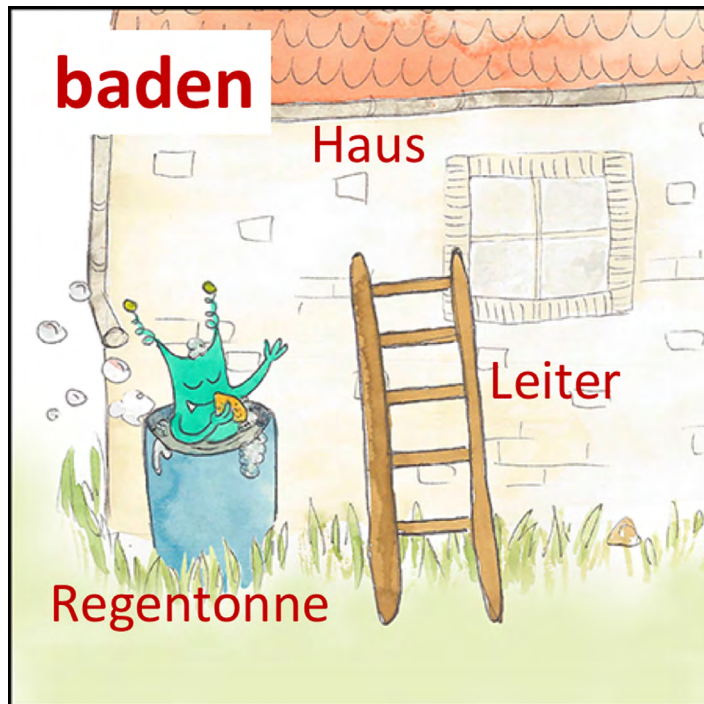

Figure 1 Example of a visual stimulus used in the familiarization task.

3.3.2. Picture-matching task with eye-tracking

During the picture-matching task, which took approximately 25 minutes, the participants saw two pictures on the screen and listened to a sentence in German that matched one of the pictures (for an illustration, see *Figure 2*). The task was introduced by means of a brief story: A girl named Anna witnessed the arrival of green and blue aliens on Earth; these aliens started doing all kinds of funny things, which Anna observed and reported with surprise. The participants listened to Anna's statements and selected the matching picture as quickly as possible by pressing the left or right button on a game controller, whilst maintaining their gaze on the screen. After five practice trials followed by a 13-point eye-tracker calibration, they completed the 96 experimental trials, with breaks after every 24 trials. Participants had to fixate on a dot in the center of the screen to enable trial onset. On test trials, the average accuracy for target picture selection was 99.28% in the L1 group ($SD = 1.29$, $range = 95.31–100$), and 99.23% in the L2 group ($SD = 1.14$, $range = 96.54–100$), indicating that none of the participants experienced comprehension troubles.

The test items were 64 German verbs[2]; all received the regular suffixes -*t* in 3SG and -*en* in 3PL PRES. Moreover, there were 32 filler items that were not analyzed. A list of all 96 verbs is

---

[2]   As test items, we used 32 weak and 32 strong verbs. Although strong verbs mark 2SG and 3SG through stem-vowel alternations, they also maintain the same, regular suffixes as weak verbs in PRES and can therefore likewise be used to study suffix-based prediction.
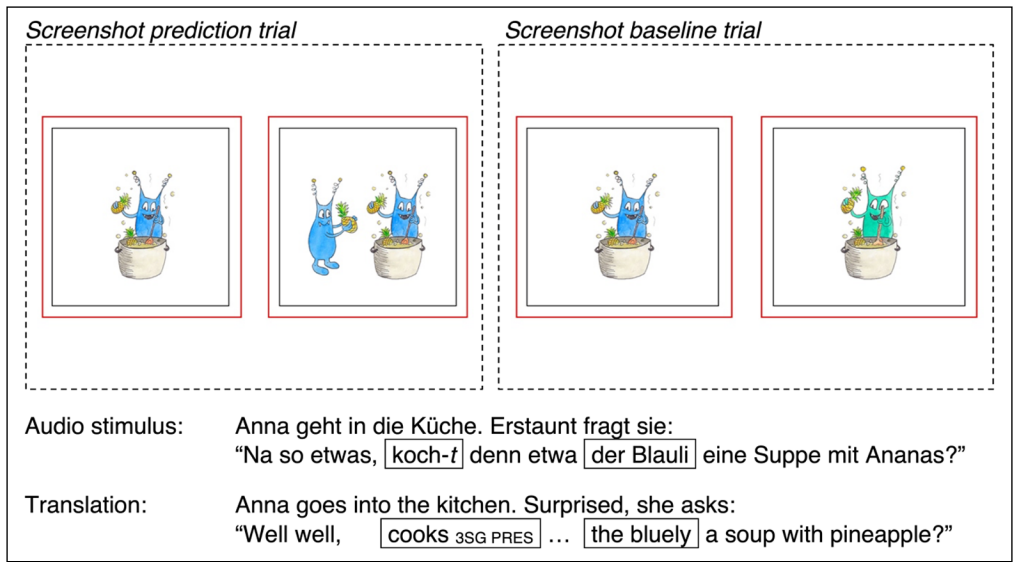
Figure 2 Illustration of experimental trials.

Note: The same audio stimulus applies to both trials. The target picture is on the left in both examples. The prediction trial is a different-number trial; alien color is identical (blue) in both pictures. The baseline trial is a same-number trial; alien color is different in the left and right pictures. The verb suffix *-t* provides a number cue for target picture in prediction but not in baseline trials. The solid, outer squares around the pictures represent areas of interest.

available in the supplementary materials. During the task, all verbs were inflected in 3SG or 3PL PRES and embedded in full, grammatical sentences describing action scenes. To maintain a low comprehension difficulty for the learners, we prioritized high-frequency verbs and cognates with Dutch.

In all trials, the onset of the auditory sentences and the display of the pictures were synchronous. Videos illustrating real-time trials are available in the supplementary materials. The visual stimuli (available on *https://doi.org/10.48316/9pmb-pf98*) depicted action scenes corresponding to a specific verb and involved blue or green aliens as agents. The picture dimensions were set to 400 × 400 pixels. The target and the distractor pictures were presented side by side, separated by a 112 pixels white margin, on a 1024 × 768 pixels screen. The corresponding interest areas for target and distractor slightly exceeded the pictures, with dimensions of 450 × 450 pixels. All looks within these areas were counted as looks towards the target or distractor picture, respectively.

All audio stimuli had a length of 9000 milliseconds, and all subcomponents had the same lengths (in milliseconds) across trials (see *Figure 3*). Each stimulus started with an introduction that served as 4500 milliseconds of preview time and that included Anna's expression of surprise, allowing the following sentence to have an interrogative format with VSO word order (the default for German declarative sentences is SVO). Placing the verb before the subject allowed us to investigate whether the participants would rely on the verb suffix to predict subject number or if they would wait for the subject NP – assumedly the default number cue – to make their choice. To lengthen the interval between verb and subject and facilitate the measurement of prediction, we inserted *denn etwa*, a modal particle that cannot be translated to English but is used to reinforce the expression of surprise in colloquial German. The subject NP was held constant and always referred to the aliens, which were named after their color. The only parameters that varied systematically in the NPs were alien color (green versus blue) and number (singular versus plural). The number information was encoded in the NP's determiner (*der*: singular, masculine, nominative; *die*: plural, nominative) and suffix (-$\emptyset_{SG}$ or -s_{PL}).

Figure 3 Sentence constituents and timings of the auditory stimuli.

Time frames	Preview time frame		Verb time frame		Subject time frame	End time frame
	Introduction	Expression of surprise	Verb	Modal particle	Subject NP	Object NP
Example sentence	*Anna geht in die Küche. Erstaunt fragt sie:*	*"Na so etwas,*	*kochen*	*denn etwa*	*die Blaulis*	*eine Suppe mit Ananas?"*
Translation	*Anna goes into the kitchen. Surprised, she asks:*	*"Well well,*	*cook* 3PL PRES		*the bluelies*	*a soup with pineapple?"*
🔊						
Audio onset	0 ms	3420 ms	4500 ms	5150 ms	6000 ms	6900 ms
Audio offset	3420 ms	4500 ms	5150 ms	6000 ms	6900 ms	9000 ms
Audio duration	4500 ms		1500 ms		900 ms	2100 ms

The sentences were spoken at a relatively slow pace by a female native speaker of High German (first author of this study). We ensured there was no co-articulation between the different sentence constituents. The recordings were further edited to bring all sentence constituents to the exact same lengths across sentences. To make the materials as natural and fluent sounding as possible, care was taken to eliminate all breaks between constituents.

As for the design of the test trials, we systematically varied the number of aliens that were visible on the two pictures (different number, same number), and the syntactic number of the target referent (singular, plural). All 64 test items were shown to the participants only once, in one of four trial types: (a) different-number, singular-target; (b) different-number, plural-target; (c) same-number, singular-target; (d) same-number, plural-target. Different-number trials represented the *prediction trial condition* in our study: The verb suffix provided the first reliable cue for target referent number (see *Figure 2*). In same-number trials, alien color was the only element that differed. In such trials, prediction was impossible because number was not contrasted and thus not informative. Therefore, same-number trials functioned as *baseline trial condition*, where the disambiguating color cue was provided through the subject NP. Items rotated between trial types according to a Latin square design: If one item was presented in trial type *a* to one participant, it would be presented in type *b* to the second participant, etcetera. Care was taken to counterbalance target alien color and target picture presentation side. Trial order was pseudo-random: Prediction trials could never appear twice in a row. To reduce the saliency of prediction and baseline trials, the disambiguating element in filler trials was never agent number or color (e.g., *Does the greenly burn **the letters/the money in the campfire?***).

3.3.3. Operation span test

To measure working memory capacity, we used the online operation span test by Klaus and Schriefers (2016; *www.socsci.ru.nl/memory/*). The participants judged equations for their correctness, while memorizing digits. To obtain scores per participant, we first determined the percentage of correctly recalled digits per block (14 blocks, ranging from two to six trials) and then calculated a mean percentage across all blocks (Zheng & Lemhöfer, 2019). Cronbach's α was .77.

3.3.4. Debriefing interview

We used a detailed, carefully structured debriefing interview (available in the supplementary materials) following the guidelines of Rebuschat (2013). The experimenter first asked general questions, inviting the participants to report if they had experienced the task as difficult or easy and if they had noticed anything special during the eye-tracking experiment. The participants were then asked what they thought the study was about and to report any strategies they had been applying. The subsequent questions were language-related and increasingly specific, starting with the role of grammar in the task, and ultimately leading towards questions about verb morphology.

Our main goal was to find out whether the participants had become aware that they could rely on verb morphology to generate predictions. We planned to code Awareness in a binary way: Participants would be coded as *aware* if the interview revealed that they had realized, at some point during the experiment, that they could focus on the verbs' suffixes to anticipate subject number; otherwise, they would be coded as *unaware*.

3.4. DATA CODING AND ANALYSIS

A laptop-mounted *EyeLink Portable Duo* recorded the participants' eye-movements with a 500 Hz sampling rate. Each sample was coded as either falling inside the target interest area (looks towards correct picture), the distractor interest area (looks towards incorrect picture) or outside of these areas. All inaccurate responses and early presses were excluded. We removed four trials for which more than 50% of the data were missing. The average percentage of eye-gaze samples that contributed to each trial was 87.74% (SD = 4.45).

Our two dependent variables were the proportion of looks towards the target picture and reaction time latency (RT, measured from the sentence onset until the button-press responses).

The proportion of target looks[3] was calculated across all samples assigned to the target or distractor interest areas, excluding blinks and saccades. We operationalized prediction of target subject number based on the verbs' morphology as more looks towards the target in prediction trials than in baseline trials, after onset of the verb and before onset of the subject NP. We shifted this critical time frame by 210 ms to the right to account for the time it takes to plan and launch eye-movements (Dahan et al., 2001) and for the audio delay introduced by the experiment computer. The prediction time frame thus ranged from 4710–6210 ms. As for the RT data, prediction should be reflected by faster RTs in prediction as compared to baseline trials. We analyzed the eye-movement data by means of cluster-based permutation analysis (see 3.4.1.) and the RT data using mixed-effects modeling (3.4.2.). Effects are reported as significant at $p < .05$. All analyses were performed in *R* (R Core Team, 2018). Data and scripts are available in the supplementary materials.

3.4.1. Cluster-based permutation analysis

To assess the presence of a prediction effect in the eye-gaze data, we used cluster-based permutation analysis (Maris & Oostenveld, 2007) using the *EyetrackingR* package (Dink & Ferguson, 2015). The participant-averaged eye-movement data were first subdivided into time bins (or 'chunks') of 50 ms (180 bins for 9000 ms). In the first step of the analysis, *t*-tests were performed for each bin to test the difference in the proportion of looks towards the target picture between prediction and baseline trials. This led to the identification of potential clusters of adjacent bins for which a significant difference was found. In the second step, we applied 2500 iterations of first randomly shuffling the bins, then repeating the *t*-tests and cluster identifications and comparing the outcomes to the clusters identified in the first step. This resulted in probabilities indicating whether or not the clusters from step one were due to chance. Any significant time clusters overlapping with the prediction time frame would tell us there was a prediction effect and inform us of its timing.

To assess prediction in the L1 and L2 groups (RQ1), we performed this analysis in each group separately. In addition, we performed a third time-cluster analysis that directly compared both groups: For each bin, we calculated the difference between the proportion of target looks in prediction and baseline trials and used the outcome as dependent variable for the *t*-tests, with GROUP (L1, L2) as independent variable.

3.4.2. Reaction-time analysis

To assess the role of mediating factors for prediction, we analyzed the RT data by means of linear mixed-effects modeling, using the *lme4* package (Bates et al., 2015; version 1.1-23). The model was fitted with restricted maximum likelihood estimation and had the following structure:

(RT) ~ 1 + TRIAL CONDITION*TARGET + TRIAL CONDITION*GROUP + TRIAL CONDITION*WORKING MEMORY + TRIAL CONDITION*TRIAL NUMBER + (1|PARTICIPANT) + (1|ITEM).

The term at the left of the tilde (~) represents the dependent variable; those to the right represent the model terms. The asterisks mark the inclusion of both the interaction and the simple effects of two variables. As fixed effects, we included the within-participant factors TRIAL CONDITION (Prediction, Baseline) and TARGET (Singular, Plural), the between-participants factor GROUP (L1, L2), and the continuous variables WORKING MEMORY and TRIAL NUMBER. Beside their simple effects, we also included two-way interactions between TRIAL CONDITION and the other variables. The random-effects terms are those that include the bar symbol (|): we included random intercepts (represented by 1) over participants and over items, thereby accounting for variation due to individual differences or item-related differences.

All fixed and random effects were included in the model on theoretical grounds, irrespective of their contributions to model fit. While TRIAL CONDITION was the crucial variable for measuring prediction, its interactions with GROUP and WORKING MEMORY would enable us to discover whether prediction was different in the L1 and L2 groups (RQ1) and whether it was influenced by working memory (RQ2). TRIAL NUMBER was included to assess any changes in RT over the course of testing.

---

3   All gaze data beyond a correct button-press were recoded as *looks towards target*. This meaningful recoding was done because once a button-press is made, the task of target identification is completed and eye-gaze data are no longer reliable indicators of sentence processing.

We also controlled for picture complexity by including the variable TARGET, which refers to the target referent being singular (one alien depicted; 3SG in audio) or plural (two aliens depicted; 3PL in audio). This was necessary because pictures that are more complex have been found to attract more looks than less complex pictures (e.g., Lukyanenko & Fisher, 2016; Schlenter, 2019).

Note that AWARENESS was not included. We planned to apply binary coding (Unaware, Aware; cf. 3.3.4.). However, the interviews revealed that all participants had become aware (cf. 4.3.), rendering the inclusion of AWARENESS in the statistical models meaningless (RQ3).

We used estimated marginal means and pairwise comparisons, using the least square means method, to inspect eight planned contrasts (cf. **Table 3**). We applied the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) to correct for multiple testing. This procedure provides an adjusted significance threshold for *p*-values, based on the number of contrasts and on the false discovery rate (set to 5%).

4. RESULTS
### 4.1. EYE-MOVEMENTS

Our first two cluster-based permutation analyses assessed predictive processing in the L1 and L2 groups separately (RQ1). Table S1 in the supplementary materials provides detailed results; **Figure 4** provides a visualization of the data. In each group, the analyses revealed a significant time cluster (ranging from 5050–6650 milliseconds in the L1 group, and from 5200–6650 milliseconds in the L2 group; both $p < .001$; cf. Table S1) that largely overlapped with the prediction time frame of the spoken sentence (4710–6210 ms). Its directionality was positive, meaning that there were more looks towards the target picture in prediction trials than in baseline trials. This confirmed the presence of robust prediction effects in both groups, starting 340 milliseconds (L1) and 490 milliseconds (L2) after prediction time frame onset and spilling over into the subject time frame.
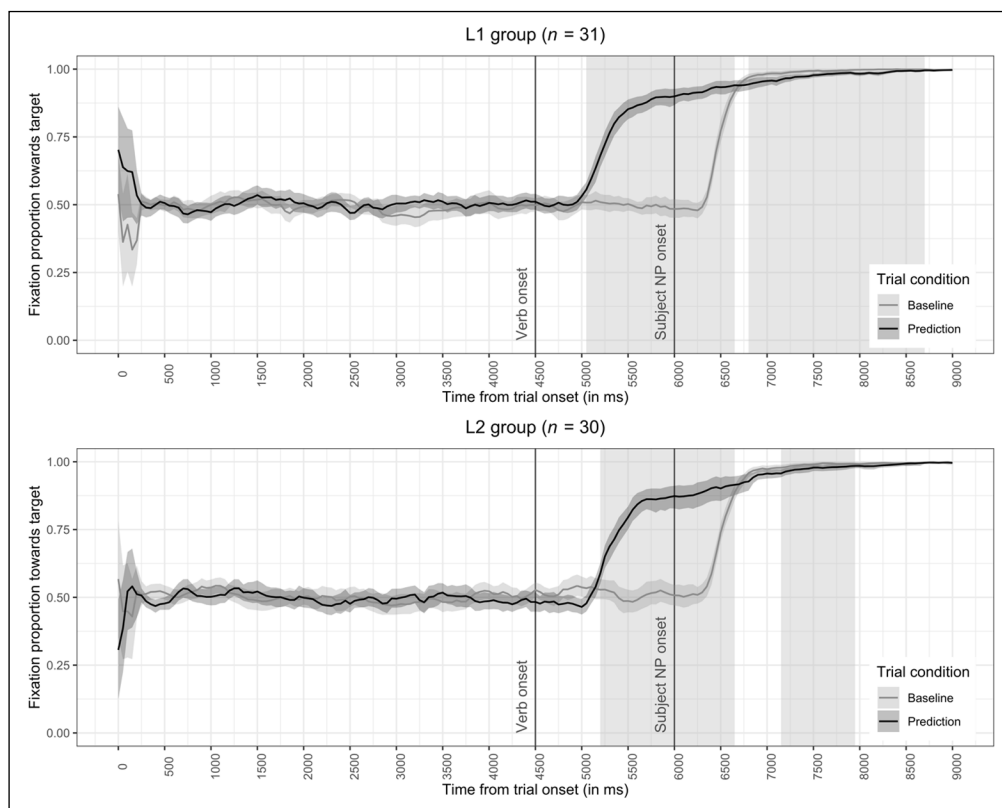
Figure 4 Proportion of fixations towards the target picture over time in the L1 and L2 groups.

Note: Error bands represent 95% confidence intervals. Shaded areas represent the time clusters during which prediction and baseline trials differed significantly.

Although the prediction effect started 150 milliseconds earlier in the L1 group, this difference was not significant according to the cluster analysis that directly compared the L1 and L2 groups (cf. Table S1). **Figure 5** visualizes this.

L1 group vs. L2 group
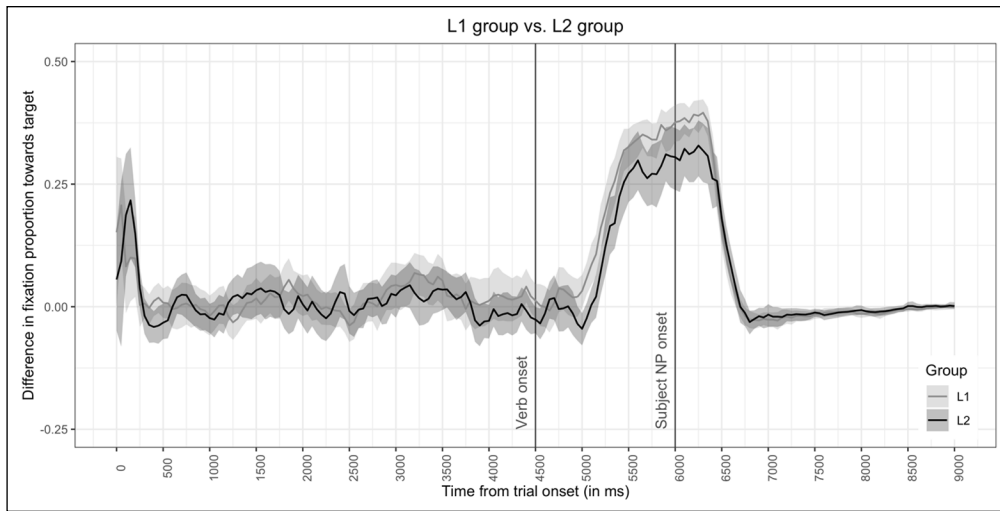
Figure 5 Difference in the proportion of target picture fixations between prediction and baseline trials in the L1 and L2 groups over time.

Note: Error bands represent 95% confidence intervals.

4.2. REACTION TIMES

The RT mixed-effects model output (provided in *Table 2*) revealed a significant effect of Trial condition. In addition, the output shows that Target, Group, Working memory, and Trial number each interacted significantly with Trial condition.

FIXED EFFECTS	RT	95% CI		SE	T	P
		LL	UL			
(Intercept)	7384	6184	8584	612	12.06	**<.001**
Trial condition	933	521	1346	210	4.44	**<.001**
Target	–14	–67	39	27	–0.51	.609
Group	210	–19	439	117	1.79	.078
Working memory	–7	–20	6	7	–1.04	.304
Trial number	0	–2	1	1	–0.12	.908
Trial condition: Target	–116	–192	–41	39	–3.01	**.003**
Trial condition: Group	211	134	288	39	5.36	**<.001**
Trial condition: Working memory	–13	–17	–8	2	–5.62	**<.001**
Trial condition: Trial number	–16	–18	–14	1	–15.26	**<.001**
RANDOM EFFECTS	**VARIANCE**	**SD**				
Participant						
(Intercept)	180810.00	425.22				
Item						
(Intercept)	6043.00	77.74				

Table 2 Parameter estimates of the reaction-time mixed-effects model.

Note: CI = confidence interval; *LL* = lower limit, *UL* = upper limit. The intercept represents the following combination of factor reference levels: Trial condition = Baseline, Target = Singular; Group = L1. Colons (:) designate interactions. Significant *p*-values are printed in bold. R^2 marginal and R^2 conditional were 0.297 and 0.566, respectively; AIC = 55174.46; REML log-likelihood = 55148.5.

We used estimated marginal means and planned contrasts to interpret the significant interactions between the categorical predictors. *Table 3* provides the estimated mean difference between specific factor levels; *Figure 6* visualizes the estimated means. The estimated means for categorical predictors cannot be inferred directly from the parameter estimates provided in *Table 2*, as the model contains both categorical and continuous predictors and interactions between these.

As for the interaction between Trial condition and Group (RQ1), there was a significant RT difference between prediction and baseline trials within both groups, reflecting prediction. In the L1 group, the model estimated mean RTs of 5915 and 6731 ms for prediction and baseline trials, respectively. In the L2 group, the estimated mean RT was 6336 ms on prediction trials versus 6944 ms on baseline trials. On prediction trials, the estimated RTs were 421 ms faster in the L1 group than in the L2 group; this was a significant difference, showing that prediction commenced earlier in the L1 than in the L2 group. On baseline trials, there was no significant group difference.

FACTOR LEVEL	CONTRAST	MD	95% CI		SE	T	P
			LL	UL			
Interaction between TRIAL CONDITION and GROUP							
Baseline	L1 vs. L2	–210	–444	24	117	–1.79	.078
Prediction	L1 vs. L2	–421	–655	–186	117	–3.59	**.001**
L1	Baseline vs. Prediction	819	768	871	27	30.97	**<.001**
L2	Baseline vs. Prediction	609	553	665	29	21.28	**<.001**
Interaction between TRIAL CONDITION and TARGET							
Baseline	Singular vs. Plural	14	–39	67	27	0.51	.609
Prediction	Singular vs. Plural	130	76	183	27	4.76	**<.001**
Singular	Baseline vs. Prediction	656	603	710	27	24.07	**<.001**
Plural	Baseline vs. Prediction	772	719	826	27	28.31	**<.001**

Table 3 Pairwise comparisons among the estimated means for TRIAL CONDITION and its two-way interactions with TARGET and GROUP (reaction-time mixed-effects model).

Note: MD = estimated reaction-time mean difference; CI = confidence interval of the estimated mean difference; *LL* = lower limit, *UL* = upper limit. *p*-values printed in bold are significant after correcting for multiple testing.
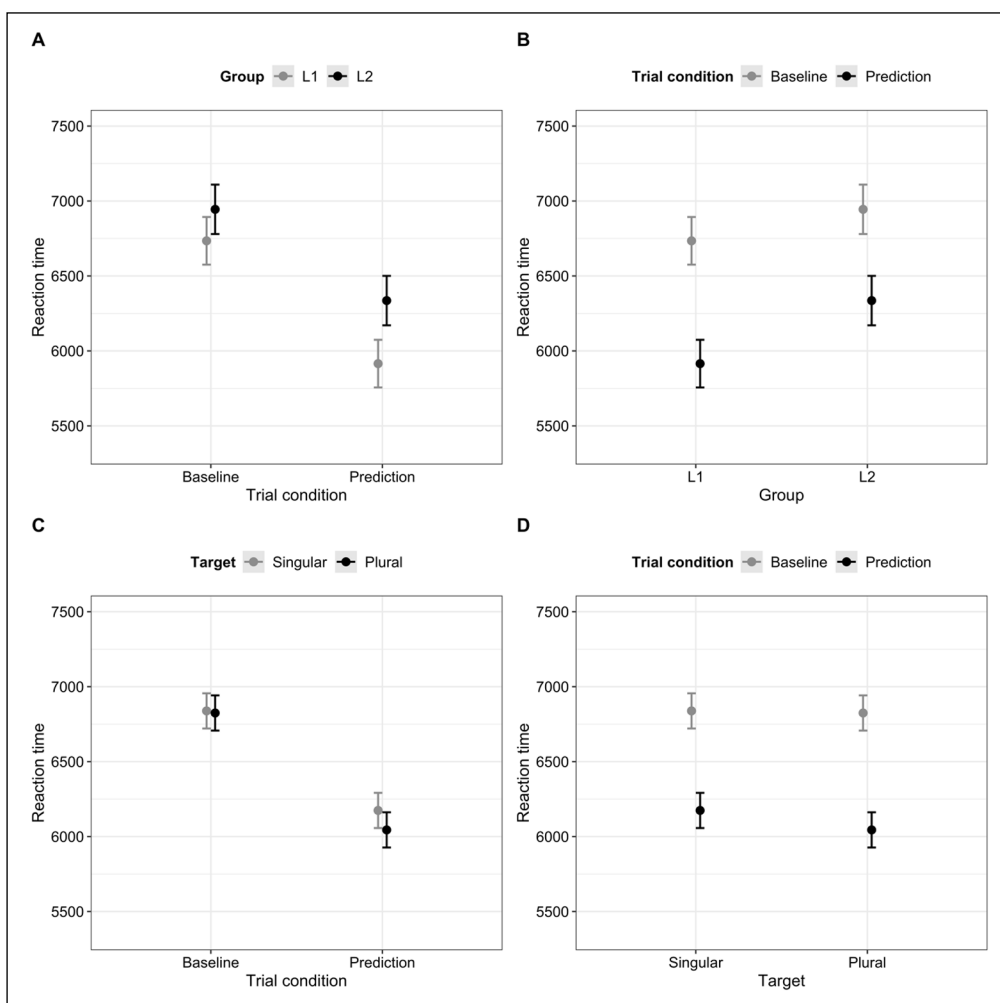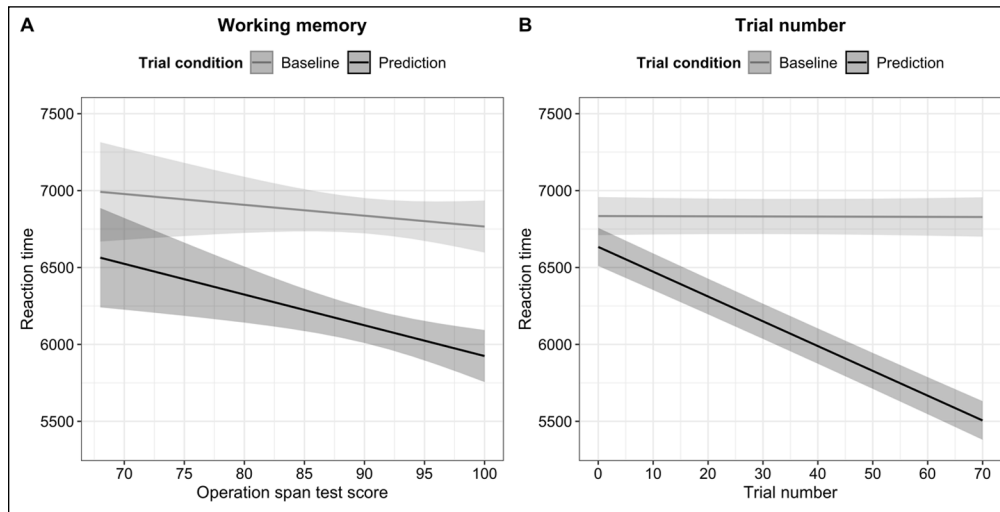

Figure 6 Estimated mean reaction times by TRIAL CONDITION and GROUP (A & B) and by TRIAL CONDITION and TARGET (C & D).

The significant interaction between TRIAL CONDITION and TARGET means the following: Overall, participants responded significantly faster on prediction trials than on baseline trials, a substantial effect that was present in both singular target trials (estimated mean RT of 6190 ms on prediction trials versus 6846 ms on baseline trials) and plural target trials (6060 ms on prediction trials versus 6833 ms on baseline trials). However, while there was no RT difference between baseline singular and plural trials, there was a significant difference on prediction trials: The estimated RTs were 130 ms faster for plural target trials than for singular target trials, reflecting a picture complexity effect (see Appendix S1 in the supplementary materials).

Figure 7 visualizes the interactions between TRIAL CONDITION and the continuous predictors. For both WORKING MEMORY and TRIAL NUMBER, their respective interaction with TRIAL CONDITION was significant, while their simple effects were not (see *Table 2*). This means that neither variable

influenced RTs on baseline trials, but they significantly influenced RTs on prediction trials. As for Working memory (RQ2), the model estimated that for every unit increase of the operation span test scores, RTs on prediction trials became faster by 20 ms. As for Trial number, the model estimated RTs on prediction trials to become 16 ms faster with every trial.

4.3. INTERVIEW OUTCOMES

The interviews revealed that, without exception and regardless of language group, all participants had become aware that they could focus on the verbs' suffixes to anticipate subject number. As there were no unaware participants, there was no need to include Awareness in the statistical models (RQ3). All participants reported that in trials with pictures that differed only in the number of agents, they could focus on the verb's morphology to identify the target picture without having to wait for the rest of the sentence. This was also true for other trial types. Generally, when the participants were asked if they had applied a specific strategy to solve the task, they reported that they first scanned the two pictures for their differences (e.g., different number/alien color/setting/objects…), after which they knew which element of the auditory sentence should provide the disambiguating cue.

5. DISCUSSION

We adopted a visual-world eye-tracking methodology to investigate to what extent L1 and advanced Dutch-speaking L2 speakers of German use the number information encoded in the suffixes of regular German verbs (-t_{3SG} and -en_{3PL}) to predict whether the upcoming subject will be singular or plural. The analyses revealed significant prediction effects in the L1 and L2 groups (RQ1). In the eye-gaze data, prediction emerged 550 milliseconds (L1) and 700 milliseconds (L2) after verb onset. Although the effect emerged somewhat earlier in the L1 group than in the L2 group, this difference was not significant. In the RTs, prediction was reflected by button-press responses that were approximately 820 milliseconds (L1) and 610 milliseconds (L2) faster in prediction trials than in baseline trials. The native speakers responded on average approximately 420 milliseconds earlier than the learners on prediction trials, which constituted a significant difference. Working memory influenced the RTs on prediction trials, with higher working memory capacity entailing faster responses (RQ2). The debriefing interviews revealed that all participants had become aware that they could focus on the verbs' form to identify the target picture in prediction trials (RQ3).

5.1. L1 AND L2 PREDICTIVE PROCESSING OF VERB NUMBER MARKINGS (RQ1)

Our study brings new evidence to the body of research that found predictive processing of morphosyntax in native speakers (e.g., Dahan et al., 2000; Hopp, 2015) and successfully extends the evidence base to a new target structure. Our findings also provide new evidence for successful morphosyntax-based prediction in L2 learners where positive evidence so far had been limited to studies investigating gender markings (e.g., Curcic et al., 2019), determiners marking animacy and distance (Andringa, 2020), or English (in)definite articles (Trenkic et al.,

2014) as predictive cues. Several studies did not find predictive processing in adult learners (e.g., gender: Lew-Williams & Fernald, 2010; case: Hopp 2015; verb number markings: Schlenter, 2019). To our knowledge, our study is the first to find evidence for the L2 predictive use of verb number markings. This outcome might have been different with less proficient learners or with a more difficult target structure. The interviews confirm that notwithstanding its morphosyntactic nature and low physical saliency, the target structure was perceived as relatively easy by all participants, possibly due to the structure's regularity and L1-L2 similarity, coupled with the contrasting picture pairs on the screen.

In line with previous research (e.g., Chun & Kaan, 2019; Grüter et al., 2012), we found prediction in the L2 to be somewhat slower than in the L1—a difference that emerged as significant in the RT data but not the gaze data. These L1-L2 processing differences cannot be attributed to a lack of comprehension of the materials in the L2 group (all participants had been familiarized extensively, and we prioritized cognates and high-frequency words), nor to a lack of knowledge of the target structure (which was mastered by all participants). Rather, the reduced prediction effect in the L2 group might be caused by overall slower lexical access and less stable lexical representations, and cognitive load may have been higher in the L2 group (cf. Kaan, 2014). Future research could explore this issue further by assessing how exactly the interaction of different individual variables leads to processing differences.

The stimulus sentences used in our experiment always had a question format. We assume that our findings also generalize to declarative sentences in which there is inversion of subject and verb, which is the case when a constituent other than the subject (e.g., accusative/dative object, adverbial or prepositional phrase) is in sentence-initial position (e.g., *Heute geht der Blauli zur Schule*; "Today goes the bluely to school", German word order). Moreover, our findings may also generalize to utterances containing imperatives. As the subject is covert in such utterances, verb morphology provides the only, and therefore, default cue for referent number (e.g., Geh_{2SG}/$geht_{2PL}$ *zur Schule!*; "$Go_{2SG/2PL}$ to school!").

In our experiment, verb form always provided a reliable number cue. That is, syntactic number and conceptual number always coincided. The morphosyntactic number markers in the verb and in the subject NP, moreover, were always grammatical. Nonetheless, cue reliability does not guarantee predictive processing, as shown by the absence of predictive number processing in Riordan et al.'s (2015) study. The fact that we found predictive number processing in a sample of L1 German speakers and that Riordan et al. did not in a sample of L1 English speakers may point to the different prominence of morphology in these two languages.

5.2. WORKING MEMORY AND PREDICTION (RQ2)

Higher working memory scores were associated with faster button-presses on prediction trials, which is in line with Huettig and Janse (2016). Research on language aptitude (cf. Li, 2016) points towards an association between working memory and explicit learning aptitude. Thus, the effect of working memory may reveal an influence of conscious, explicit language processing in prediction. It may be interesting for future research to explore the link between working memory and prediction further by comparing the effect of working memory on prediction in aware versus unaware participants and in L1 versus L2 speakers (which was not possible in the present study due to sample size).

5.3. AWARE PREDICTIVE PROCESSING (RQ3)

The debriefing interviews revealed that all L1 and L2 participants had become aware that they could focus on verb form to identify the target picture in prediction trials, and they applied this knowledge strategically. Thus, comparable to the findings of Curcic et al. (2019), the predictive processing in our sample was conscious. Awareness grew with time on task: The further the participants proceeded through the task, the faster they became at predicting target number.

The task format (picture-matching) seems to have played a role in this respect. Contrasting picture pairs may have prompted the participants to compare the two pictures and generate hypotheses. In combination with a target structure that was perceived as relatively easy, this may have led to the swift development of prediction awareness. An important point to consider here is that during the second eye-tracking experiment (cf. Koch et al., 2021), which took place 20 minutes later, shared the same design and items but had a more difficult, less salient

target structure (cf. 3.3.), only half of the participants developed prediction awareness in this experiment. Clearly, multiple factors affect whether or not a listener will develop awareness of relevant cues in the input.

The observation of a conscious form of prediction underscores the importance of thorough post-experimental interviews. Although research methods such as eye-tracking or EEG have the potential of measuring unaware language processing in real-time, the use of a time-sensitive method does not guarantee that participants will actually engage in unaware comprehension processes only (Curcic et al., 2019; Godfroid & Winke, 2015). Much depends on the experimental task at hand and the respective instructions, stimuli and target structure(s).

6. CONCLUSION

The present study provides pioneering evidence for predictive morphosyntactic processing in an L2 and also adds new evidence for morphosyntax-based predictive processing in an L1. Using regular verb suffixes as a number cue, we found significant prediction effects both in the L1 and L2 groups, though this effect commenced earlier in the L1 group. RTs showed a facilitatory influence of working memory for prediction. In light of the relative ease with which the participants in our study engaged in predictive suffix-based processing, our experiment might be suitable to further explore the impact of specific task-related parameters on prediction, such as preview time and speech rate (cf. Huettig & Guerra, 2019) or the omission of button-press responses. From a methodological perspective, our study illustrates that thorough debriefing interviews assessing awareness are indispensable for prediction research, including studies focusing on L1 processing. Taken together, our findings suggest that learners and native speakers of German exploit the rich German inflectional system predictively during online sentence processing. Our study paves the way for further exploration of predictive morphosyntactic processing, for example by using more difficult target structures or introducing more measures of individual differences.

SUPPLEMENTARY MATERIALS

Data, analysis scripts and additional supplementary materials are publicly available via the Open Science Framework site for this project: *https://osf.io/37rfb/*.

The visual stimulus materials are available via the IRIS database: *https://doi.org/10.48316/9pmb-pf98*.

ETHICS AND CONSENT

This study was approved by the Ethical Committee in Humane Sciences (ECHS) of Vrije Universiteit Brussel on December 4, 2014.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Eva M. Koch ⓘ *orcid.org/0000-0003-4648-1105*
Vrije Universiteit Brussel, Brussels, Belgium

Bram Bulté ⓘ *orcid.org/0000-0003-2761-9455*
Vrije Universiteit Brussel, Brussels, Belgium

Alex Housen
Vrije Universiteit Brussel, Brussels, Belgium

Aline Godfroid ⓘ *orcid.org/0000-0001-8011-7935*
Michigan State University, East Lansing, MI

REFERENCES

Andringa, S. (2020). The emergence of awareness in uninstructed L2 learning: A visual world eye tracking study. *Second Language Research*, *36*(3), 335–357. DOI: *https://doi.org/10.1177/0267658320915502*

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*. DOI: *https://doi.org/10.18637/jss.v067.i01*

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. DOI: *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. DOI: *https://doi.org/10.1016/j.jml.2016.10.002*

Chun, E., & **Kaan, E.** (2019). L2 prediction during complex sentence processing. *Journal of Cultural Cognitive Science*, *3*(2), 203–216. DOI: *https://doi.org/10.1007/s41809-019-00038-0*

Curcic, M., Andringa, S., & **Kuiken, F.** (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*, 42–71. DOI: *https://doi.org/10.1111/lang.12321*

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & **Hogan, E. M.** (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534. DOI: *https://doi.org/10.1080/01690960143000074*

Dahan, D., Swingley, 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. DOI: *https://doi.org/10.1006/jmla.1999.2688*

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*, 1–25. DOI: *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. DOI: *https://doi.org/10.1017/S1366728916000547*

Dink, J. W., & **Ferguson, B.** (2015). *eyetrackingR: An R library for eye-tracking data analysis*. Retrieved from *http://www.eyetracking-r.com*

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, *38*, 177–215. DOI: *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*. DOI: *https://doi.org/10.4324/9781315775616*

Godfroid, A., & **Winke, P.** (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 323–348). John Benjamins. DOI: *https://doi.org/10.1075/sibil.48.14god*

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. DOI: *https://doi.org/10.1177/0267658312437990*

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. DOI: *https://doi.org/10.1075/lab.15011.gru*

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*, 33–56. DOI: *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. DOI: *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. DOI: *https://doi.org/10.1017/S0272263116000437*

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. DOI: *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*, 196–208. DOI: *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. DOI: *https://doi.org/10.1080/23273798.2015.1047459*

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. DOI: *https://doi.org/10.1075/lab.4.2.05kaa*

Klaus, J., & **Schriefers, H.** (2016). *Measuring verbal working memory capacity: A reading span task for laboratory and web-based use*. DOI: *https://doi.org/10.31219/osf.io/nj48x*

Koch, E. M., Bulté, B., Housen, A., & **Godfroid, A.** (2021). *The predictive processing of number information in subregular verb morphology in a first and second language* [Manuscript submitted for publication]. Department of Linguistics and Literary Studies, Vrije Universiteit Brussel.

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. DOI: *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. DOI: *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. DOI: *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. DOI: *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. DOI: *https://doi.org/10.3758/s13428-011-0146-0*

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. DOI: *https://doi.org/10.1016/j.jml.2010.07.003*

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*(4), 801–842. DOI: *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. DOI: *https://doi.org/10.1016/j.cognition.2015.10.012*

Maris, E., & **Oostenveld, R.** (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. DOI: *https://doi.org/10.1016/j.jneumeth.2007.03.024*

Mitsugi, S., & **MacWhinney, B.** (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35. DOI: *https://doi.org/10.1017/S1366728914000881*

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. DOI: *https://doi.org/10.1017/S1366728915000176*

Ortega, L. (2009). *Understanding second language acquisition*. Routledge. *https://www.routledge.com/Understanding-Second-Language-Acquisition/Ortega/p/book/9780340905593*

Pickering, M. J., & **Gambi, C.** (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. DOI: *https://doi.org/10.1037/bul0000158*

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. DOI: *https://doi.org/10.1111/lang.12010*

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. DOI: *https://doi.org/10.3389/fpsyg.2015.00590*

Rommers, J., Meyer, A. S., & **Huettig, F.** (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, and Psychophysics*, *77*(3), 720–730. DOI: *https://doi.org/10.3758/s13414-015-0873-x*

Schlenter, 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. DOI: *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. DOI: *https://doi.org/10.1007/s10936-014-9322-6*

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. DOI: *https://doi.org/10.1017/S1366728913000321*

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. DOI: *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. DOI: *https://doi.org/10.1002/9781405198431*

Zheng, X., & **Lemhöfer, K.** (2019). The "semantic P600" in second language processing: When syntax conflicts with semantics. *Neuropsychologia, 127,* 131–147. DOI: *https://doi.org/10.1016/j.neuropsychologia.2019.02.010*

WHITE ROSE
UNIVERSITY PRESS
Universities of Leeds, Sheffield & York