languages

MDPI

Article

The Predictive Processing of Number Information in L1 and L2 Arabic Speakers

Alaa Alzahrani

Feather Marketing Solutions, Riyadh 35419, Saudi Arabia; alzahrani.alaaa@gmail.com

Abstract: Prior research has shown that people can predict the syntactic features of an upcoming word during sentence comprehension. However, evidence for morphosyntactic predictive processing has been limited to gender or case marking in a small subset of Indo-European languages. In the current study, we implemented the eye-tracking visual world paradigm to investigate whether L1 (n = 18) and L2 (n = 40) Arabic speakers could extract number information from singular-marked verbs to anticipate the next noun. In a between-subject design, L1 and L2 speakers heard the singular verb in the simple past form (Exp 1) and the progressive past form (Exp 2). The effect of L2 proficiency (measured using a C-test and a receptive vocabulary test) on number prediction was also examined. L1 Arabic speakers showed earlier and stronger number prediction effects regardless of verb aspect. In contrast, L2 speakers exhibited delayed (Exp 1) or limited (Exp 2) prediction, suggesting a mediating role for verb aspect. Increased L2 proficiency did not influence anticipatory eye-movements during the verb region, and only emerged as significant during the noun region. These results confirm and extend earlier research on L1 and L2 number predictive processing.

Keywords: morphosyntactic predictive processing; number prediction; L1 and L2 Arabic speakers; verb aspect; L2 proficiency

check for updates

Academic Editor: John M. Everatt

Received: 21 September 2024
Revised: 12 December 2024
Accepted: 20 January 2025
Published: 27 January 2025

Citation: Alzahrani, A. (2025). The Predictive Processing of Number Information in L1 and L2 Arabic Speakers. *Languages*, *10*(2), 25. https://doi.org/10.3390/languages10020025

1. Introduction

Fluent speakers rapidly understand spoken language as sentences unfold in time. This fast sentence processing system has been thought to be partly driven by predictive language processing: the speaker's ability to predict upcoming information based on the current and prior linguistic context (Huettig et al., 2022; Kuperberg & Jaeger, 2016; Kutas et al., 2011; Pickering & Gambi, 2018). Evidence from various methods such as event-related potentials and the eye-tracking visual-world paradigm (VWP) suggest that adult first language (L1) and second language (L2) speakers can predict lexical-semantic (e.g., Mani & Huettig, 2012; Schlenter & Felser, 2021), morphosyntactic (e.g., Hopp, 2015), and phonological features (e.g., Ito et al., 2020) of the next word in a sentence.

However, evidence for morphosyntactic prediction is mainly based on grammatical gender or case in a narrow and skewed subset of world languages, such as German (Bosch & Foppolo, 2022; Hopp, 2013, 2016; Hopp & Lemmerth, 2018), Spanish (Dussias et al., 2013; Foucart et al., 2014; Garrido-Pozu, 2022), and Italian (Bosch & Foppolo, 2022). If predictive processing is a fundamental property of sentence comprehension (e.g., Huettig et al., 2022; Kaan & Grüter, 2021; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018), then empirical support should come from a larger range of typologically different languages.

To broaden our understanding of predictive processing, the present study will investigate the predictive processing of number information in modern standard Arabic (MSA).

Unlike the commonly examined languages in the relevant literature, MSA has a dominant verb-subject-object (VSO) word order (Dryer, 2013). The verb-initial word order in MSA provides a suitable test case for investigating predictive number processing, as listeners can extract number information from the verb to predict the number of the subsequent noun. Additionally, the current study will examine the influence of speaker-related factors (L1 vs. L2 speakers and L2 proficiency level) as well as linguistic factors (verb aspect) on predictive number processing to extend prior research.

2. Literature Review

2.1. Defining Prediction

In line with most theoretical accounts of predictive processing, including the utility account (Kuperberg & Jaeger, 2016), the prediction-by-production model (Pickering & Gambi, 2018), and the parallel architecture account (Huettig et al., 2022), this study defines prediction as the pre-activation of linguistic information before exposure to the upcoming word or within a few hundred milliseconds of its encounter. This restricted definition is primarily followed in most VWP studies (Alemán Bañón & Martin, 2021; Hopp, 2016; Ito et al., 2018a, 2018b; Mitsugi & Macwhinney, 2016), and its adoption here could facilitate comparison with prior research.

2.2. L1 and L2 Predictive Sentence Processing

Decades of psycholinguistic research have demonstrated that L1 speakers can use different linguistic information, such as semantic, morphosyntactic, and phonological information, to rapidly make predictions as a sentence unfolds (for reviews, see Altmann & Mirković, 2009; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). In a pioneering VWP study, Altmann and Kamide (1999) presented English native speakers with two sentence conditions: one included semantically-constraining verbs, e.g., "the boy will eat the cake", and the other condition used non-semantically-constraining verbs, e.g., "the boy will move the cake". Participants' eye movements were recorded as they heard the auditory sentences while viewing a visual scene depicting only one edible object (the cake) and several non-edible objects (ball, car, train). When hearing a semantically constraining verb (eat), participants looked at the edible object sooner than when hearing a non-constraining verb (move), even before the auditory presentation of the object. These fixation patterns suggested that English natives used the verb's semantic information to anticipate the yet-to-be-encountered object.

Meanwhile, L2 speakers have been shown to extract semantic information to predict what is coming next (Chun et al., 2021; Mitsugi, 2018; Schlenter & Felser, 2021), with mixed results for the predictive use of morphosyntactic information (Curcic et al., 2019; Dussias et al., 2013; Frenck-Mestre et al., 2019; Garrido-Pozu, 2022; Hopp, 2016; Koch et al., 2021, 2023; Mitsugi, 2017; Mitsugi & Macwhinney, 2016; Schlenter & Felser, 2021). For instance, Schlenter and Felser (2021) reported that L2 German speakers successfully extracted morphological case information to anticipate the next word. Meanwhile, Mitsugi (2017) found that L2 Japanese speakers did not use case-marked nouns to predict the upcoming verb voice (active vs. passive). The relative difficulty of morphosyntactic prediction can be interpreted under the prediction-by-production model (Pickering & Gambi, 2018). The prediction-by-production account argues that comprehenders (language readers and listeners) are more likely to rapidly predict semantic information than syntactic information because predictive processing progresses from early (semantic) to later stages (syntax, phonology), mirroring the steps in language production (semantics, syntax, phonology) (e.g., Ito et al., 2018b, 2020). In this view, L2 speakers may struggle to recognize syntactic information in the previous input as well as predict syntactic information in the upcoming input. Another possible reason for the limited L2 prediction could be related to the challenges associated with extracting L2 morphosyntactic information due to L1-L2 differences (e.g., Dussias et al., 2013).

A growing number of studies have directly compared morphosyntactic predictive processing between L1 and L2 speakers (e.g., Garrido-Pozu, 2022; Koch et al., 2021, 2023). A common finding from these studies is that L2 speakers show later and/or weaker anticipation effects than L1 speakers. For example, Garrido-Pozu (2022) reported that L2 Spanish speakers predicted the gender of upcoming adjectives much later than the L1 Spanish group. Similarly, Koch et al. (2023) showed that L2 German participants exhibited weaker predictive number processing effects relative to native speakers. However, a number of studies have shown that L2 speakers may exhibit L1-like prediction skills when they demonstrate knowledge of the predictive cue (e.g., Hopp, 2013) or when the cue is present in both L1 and L2 systems (e.g., Dussias et al., 2013).

L1-L2 differences in predictive processing are expected, as L2 processing and performance are qualitatively more variable (Dornyei & Ryan, 2015; Tagarelli et al., 2016). This variation in L2 and L2 predictive behavior could be attributed to a range of factors (Schlenter, 2023). Compared to L1 speakers, it is widely reported that L2 speakers tend to have weaker lexical representations (Ivanova & Costa, 2008), interlingual lexical competition (Kaan, 2014), difficulty extracting syntactic information (Clahsen & Felser, 2006, 2017), a less automatic processing system (Ito & Pickering, 2021; Segalowitz & Hulstijn, 2009), and different cue weights (e.g., Bates & MacWhinney, 1989). These factors trigger unique prediction patterns across and within L2 groups, underscoring the need to investigate L2 speakers in different contexts.

2.3. Verb Number Marking as a Predictive Cue

Most predictive number processing studies have investigated whether L1 children and adults can extract and use number information from the copula verb (Brown et al., 2022; Kouider et al., 2006; Lukyanenko & Fisher, 2016; Reuter et al., 2021). These studies used a VWP design to track participants' eye movements as they viewed several objects on the screen while hearing auditory sentences with number-marked copulas, e.g., "Look, there is an apple/are some apples" or without number-marked copulas, e.g., "Look at the apple". These studies found that L1 listeners exploit number information from copula verbs (is, are, was, were) and determiners (a, an, some) to anticipate the upcoming target noun.

However, a limitation of earlier studies is that they investigated number prediction across a small set of grammatical words, such as copulas and determiners. As these grammatical words are fixed, the prediction of number information in these studies might have resulted from knowledge of the specific word rather than knowledge of an abstract grammatical number system. This possibility is especially likely in the case of less proficient L2 speakers. According to the shared syntax model, beginner L2 speakers have item-specific syntactic representations and start to develop more abstract syntactic knowledge with increasing L2 proficiency (Hartsuiker & Bernolet, 2017).

Recognizing this limitation, recent predictive number processing studies have shifted the focus to number markings on lexical verbs. These studies were conducted on Indo-European languages and showed that L1 Czech children (Smolík & Bláhová, 2022), as well as German L1 and L2 speakers (Koch et al., 2021, 2023), can utilize number-marked verbs to anticipate the number of the upcoming subject. More research is needed to examine whether L1 and L2 speakers of morphologically distinct languages show similar morphosyntactic predictive behaviors.

2.4. A Brief Description of MSA

A well-known feature of Arabic-speaking communities is diglossia (Albirini, 2016; Zughoul, 1980). Diglossia, strictly defined, refers to the coexistence of two varieties of the same language in a community, each fulfilling distinct communication functions (Ferguson, 1991). Native Arabic speakers tend to use MSA in formal communication contexts (official media, education, government, law) and utilize a regional spoken Arabic variety in everyday communication. Arabic-speaking children typically first learn a regional variety of Arabic and begin learning MSA when they start school. Arabic language programs prioritize teaching MSA to L2 learners, and learners may informally acquire a regional Arabic variety outside the classroom (e.g., Nassif, 2021; Palmer, 2007; Younes, 2014).

2.5. Number and Aspect in MSA

MSA is characterized by its verb-initial order and rich morphological system (Ryding, 2005). The singular number has different markers depending on the verb aspect. In third-person singular masculine (3SG.M) simple past verbs, the suffix "-a" marks the singular form (e.g., darasa/studied). The corresponding past progressive form starts with the particle "ka:na" ("was"), followed by the present form of the verb (i.e., -ing form), and ends with the suffix "-u" to mark the singular number (e.g., ka:na jadrusu/was studying). The majority of MSA L2 textbooks introduce the simple past verb earlier than the past progressive (e.g., Al-Batal et al., 2011; Al-Fawzan et al., 2014; Alosh & Clark, 2021). Thus, L2 Arabic speakers are likely to be more familiar with the simple past form.

Like in English, the progressive aspect differs from the simple aspect in two possible ways (Atawneh, 2001). MSA speakers may use the progressive form to indicate (a) the incompleteness of the event and/or (b) the overlap of two actions. For instance, the simple past form "darasa/studied.3SG.M" shows that the action was completed, while the past progressive "ka:na jadrusu/was.3SG.M studying.3SG.M" implies that an action was continuing or ongoing in the past. The past progressive tense can also be used to signal that an ongoing past action was interrupted by another past action, e.g., "ka:na jadrusu di ʕindama daxalt/(he) was studying when (I) came".

Importantly, MSA has specific word orders which require verb-subject agreement in number (Alshammari, 2023; Fakih, 2016; Ryding, 2005; Soltan, 2006). One structure is the verb-object (VO) order. In VO constructions, the verb should agree with the subject's number. Notably, the singular verb form sometimes might be followed by a number non-matched subject (e.g., in VSO word order), suggesting the non-markedness of this form.

2.6. Effect of Predictive Cue Form

Theoretical and empirical evidence indirectly suggests that listeners may show variability in their predictive use of the different forms of the same predictive cue. Theoretically, the competition model (Bates & MacWhinney, 1989; Macwhinney, 2012) proposes that cue validity affects processing strategies. Valid cues are linguistic features that are frequently encountered in the input and consistently signal the same relationship. For instance, if singular verbs are always followed by singular subjects, language users will perceive the singular number marker as a highly valid cue. Conversely, if singular verbs are followed by both singular and plural subjects, the singular number marker will be perceived as a less valid cue. The competition model assumes that cue validity is not a fixed property but depends on the speaker's language experience. In this view, language speakers are likely to assign different weights to the different forms of the same cue based on the perceived reliability of the cue form. A cue form with higher perceived validity will trigger predictive processing effects, while a cue form with lower perceived validity will be rarely exploited to predict the next word, as they may result in more costs for the parser (Hopp, 2016).

Another theoretical account that could explain variation in generating predictions across forms of the same cue is the utility account (Kuperberg & Jaeger, 2016). This account proposes that comprehenders change their predictive behavior based on cost–benefit analyses to maximize their language processing efficiency (e.g., Brothers et al., 2017). This cost–benefit analysis is based on the speaker's language experience. In this case, speakers with increased language experience may predictively use different forms of the predictive cue. Conversely, speakers with limited language experience might find it challenging to predictively use alternative forms of the same cue. In the latter case, the anticipatory use of unfamiliar cue forms can lead to processing errors and ultimately hinder, rather than facilitate, language comprehension. Overall, this account posits that the relative utility of predictive cues varies depending on the specific cue variant.

Additionally, there is some indirect empirical support for the influence of cue type on predictive processing. Very few L1 and L2 predictive processing studies have investigated several cues from the same domain. For instance, Brouwer et al. (2017) presented adult L1 Dutch speakers and children with two sets of sentences: one contained the neuter-gender article "het" followed by a gender-matched noun, and the other set included the neuter-gender article "de". L1 Dutch-speaking children, unlike L1 adults, showed facilitation effects (i.e., more target fixations after the onset of the noun) when they heard neuter nouns but not common nouns, suggesting an asymmetry in gender processing. One explanation for this finding is that the neuter gender is a more reliable cue because it has a restricted use, whereas the common gender article "de" is less reliable as it can sometimes precede both common and neuter nouns. These findings suggest that the validity of cues within the same grammatical domain might vary by cue reliability and speaker group (adult vs. children). Nevertheless, limited studies examined whether the L1 and L2 predictive use of a single cue varies across its morphological forms. Here, we investigated L1 and L2 prediction across two forms of the same morphosyntactic cue: singular number marking in simple past verbs vs. singular number marking in past progressive verbs.

2.7. Effect of L2 Proficiency

Several studies have examined the effect of L2 proficiency on L2 morphosyntactic predictive processing. These studies yielded mixed findings, with some studies observing stronger prediction effects with increasing L2 proficiency (Dussias et al., 2013; Garrido-Pozu, 2022; Henry et al., 2022; Hopp, 2016; Hopp & Lemmerth, 2018), and other studies reporting a limited role of L2 proficiency (Hopp, 2015). For instance, Hopp and Lemmerth (2018) revealed that advanced L2 German speakers showed nativelike gender prediction, while their less proficient counterparts demonstrated weaker anticipation effects. Dussias et al. (2013) indicated that more proficient L2 Spanish speakers predictively used gender information, whereas the lower proficiency group did not. These findings highlight the potential role of L2 proficiency in the predictive use of L2 morphosyntactic information. However, prior research has not yet explored the effect of L2 proficiency on number predictive processing.

2.8. The Present Study

The current study used a mixed design involving both between-subject and within-subject variables. The between-subject variables were speaker group (L1 vs. L2) and verb aspect (simple past vs. past progressive). The within-subject variable was L2 proficiency level. Two VWP experiments (Exp) were conducted to separately examine the effects of two verb aspects (simple past verbs vs. past progressive verbs) on predictive number processing during sentence comprehension. All participants completed only one of the two experiments. The present study aimed to answer the following research questions (RQs):

(1) Do L1 and L2 Arabic speakers extract singular number information from the suffix of lexical verbs to anticipate the number of the upcoming noun? To what extent is the prediction effect similar in L1 and L2 groups?
(2) To what extent does verb aspect influence L1 and L2 predictive number processing?
(3) To what extent does L2 proficiency influence L2 predictive number processing?

Based on the reviewed literature, it is expected that both L1 and L2 Arabic speakers would anticipate the number of the next subject. However, L1 speakers would show anticipatory looks regardless of verb aspect (cue form), while L2 speakers' prediction would vary across verb aspect. Finally, more proficient L2 speakers are expected to show an earlier and stronger prediction effect than less advanced L2 speakers.

3. Methods

3.1. Participants

A total of 72 L1 and L2 Arabic speakers completed one of the two experiments. The L1 speakers were recruited from Prolific and social media, and the L2 speakers from King Saud University. Prolific is an online research platform that recruits participants from various countries (Palan & Schitter, 2018). Compared to other online recruitment platforms, participants on Prolific have been shown to produce more accurate responses and perform better on attention check questions (Douglas et al., 2023; Peer et al., 2021). All participants were compensated for their time (6 USD for L1 speakers and 13 USD for L2 speakers).

Data from participants who scored less than 70% on either of the comprehension tasks (L1 speakers = 3, L2 speakers = 5) and those who had sampling rates less than 5 Hz (L1 speakers = 2, L2 speakers = 4) were excluded from further analysis (Prystauka et al., 2023). Of the remaining sample, 29 participants completed Exp 1 (L1 speakers = 9, L2 speakers = 20), and 29 completed Exp 2 (L1 speakers = 9, L2 speakers = 20).

As Arabic-speaking communities are diglossic (see the brief MSA description section), the current study assessed the Arabic L1 participants' level of proficiency in MSA using a self-reported questionnaire (see the background questionnaire section for details). L1 participants reported high self-perceived proficiency in all Arabic language skills (speaking, listening, reading, and writing), with ratings ranging from 8.44 to 9.28 across the skills (possible maximum score = 10).

Thirty-four out of the forty L2 participants reported their L1. The L2 Arabic speakers came from 17 different L1 backgrounds (Tagalog = 8, Urdu = 6, Bengali = 3, Filipino = 2, Malay = 2, Thai = 2, Bambara = 1, French = 1, Fula = 1, Indonesian = 1, Iranun = 1, Mandinka = 1, Pashto = 1, Persian = 1, Tausug = 1, Uzbek = 1, Wolof = 1). There were a limited number of L2 Arabic participants who shared a common L1. Only seven of these L1s distinguish between singular verbs and plural verbs: French, Fula, Pashto, Persian, Urdu, Uzbek, and Wolof (Esher et al., 2020; Mahmoodi-Bakhtiari, 2018; Robert, 2016; Schmidt, 2005; Straughn, 2011). Thus, while some of the L2 participants had a singular-plural distinction for verbs in their L1 (n = 12), most did not (n = 22)[1].

Since most of the L2 participants completed a C-test, we estimated their Arabic proficiency level based on their standardized scores on the C-test (see the C-test section for details). L2 speakers who scored one standard deviation (*SD*) below the mean fell into the lower proficiency group. Scores within one *SD* of the mean indicated intermediate proficiency, and scores one *SD* above the mean suggested advanced proficiency. Most of the L2 participants had an intermediate level of L2 proficiency (n = 26), and a few were at lower (n = 6) or higher (n = 6) L2 Arabic levels.

3.2. Materials

3.2.1. Stimuli

This study constructed two types of stimuli: auditory sentences and visual stimuli. First, the study created two sets of sentence stimuli, each containing 32 sentences. The two sets differed only in verb aspect. Set "A" used the simple past verb tense, while set "B" included the past progressive verb tense. Each set contained 16 experimental and 16 filler sentences. Each participant completed only one Exp and was therefore exposed to only one stimulus set.

Figure 1 shows an example of an experimental sentence from both experiments. As shown in Figure 1, the experimental sentence starts with the introductory phrase "hind asked", which makes the sentence-final question pragmatically appropriate. Then, the sentence includes the question marker "hal/did", which does not contain the number or gender information. After this, the 3SG.M simple past (Exp 1) or past progressive (Exp 2) verb is presented. The verb is followed by a complement, always an adverbial phrase such as "in the stadium". The complement was added to give participants time to process the verbal number morphology and direct their gaze anticipatorily to the target image (e.g., Garrido Rodriguez et al., 2023; Koch et al., 2023). Finally, the sentence ends with a clarifying statement which includes "I mean", then the singular subject "the boy", and the adjective "tall". The adjective was added to disambiguate the target object in baseline trials, which always presented pairs of identical images: one of a tall boy and one of a short boy (e.g., Koch et al., 2023). The structure of filler sentences was identical to the experimental stimuli per experiment, except for one difference: the fillers only included feminine verbs and subjects to reduce the focus on the masculine verbs and subsequently reduce participants' attention to the target structure.

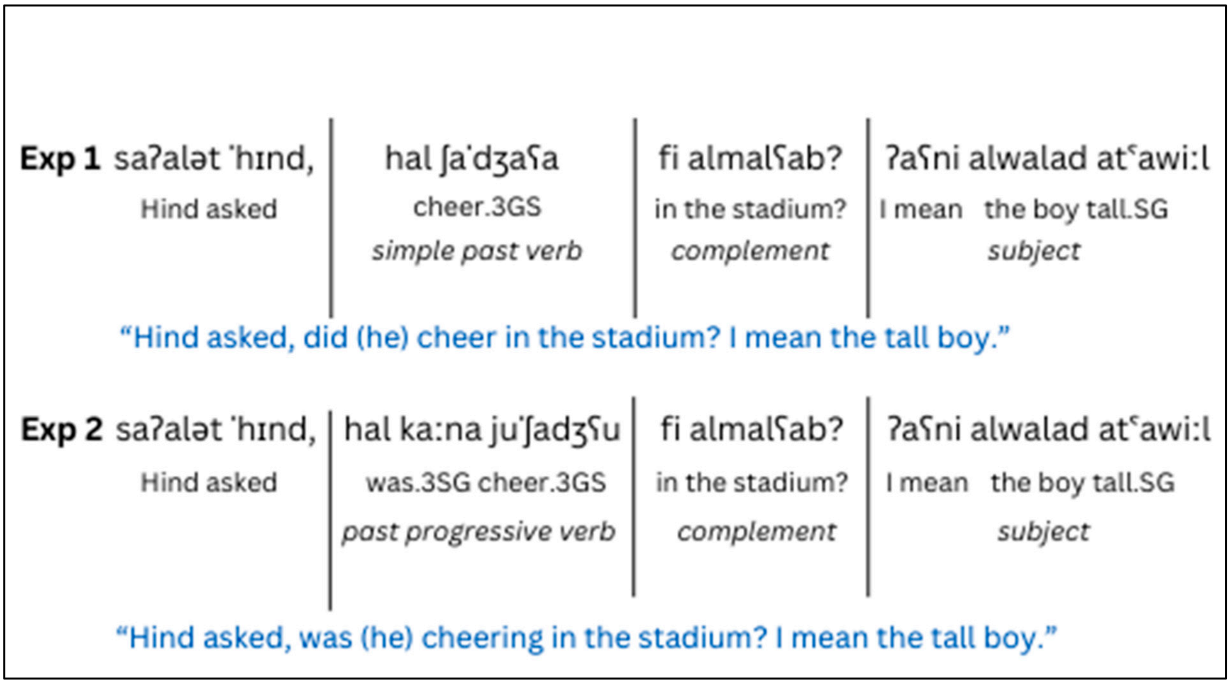

Figure 1. Sample experimental sentences for the verb ʃaˈd'aʕa 'cheer' across experiments. The Arabic sentence is in black, and its English equivalent is in blue. The vertical lines indicate the sentence regions: the introductory phrase, the verb, the complement, and the target subject.

All sentences were recorded in MSA using an AI voice generation tool (https://voicemaker.in/; accessed on 1 December 2023). An AI-generated voice was used to ensure consistency and control across the sentences as human-produced auditory stimuli exhibited variability in stress, pitch, and articulation duration. A new group of twenty-seven Arabic L1 speakers completed a speech quality rating task to assess the acceptability of the AI speech. In this task, participants listened to six AI-recorded experimental sentences and six human-recorded sentences and rated them using the Mean Opinion Scale-Expanded on an 11-point scale (Lewis, 2018). There was no significant difference in perceived intelligibility ($\chi^2(1) = 1.043$, $p = 0.307$), but there were significant differences in perceived naturalness, prosody, and social impression ($p < 0.05$). Although not human-like, the current results indicate that AI-generated speech is comprehensible.

The recorded sentences were modified in Praat (Boersma & Weenink, 2023) to ensure that all parts had the same length across the sentences (e.g., Ito et al., 2023; Koch et al., 2023). The onset, duration, and offset of all words were measured using Praat (Figure 2). *t*-tests indicated that mean word duration did not significantly differ across trial conditions within each experiment ($p > 0.05$) or between the two experiments ($p > 0.05$).

Time frame	Audio sentence	Translation	Onset	offset	Duration
Introductory phrase	saʔalət ˈhɪnd	Hind asked	0.139	1046	0.904
Verb	hal ʃaˈdʒaʕa/ hal ka:na juˈʃadʒʕu	did cheer/ was cheering	1564	2625	1.097
Complement	fi almalʕab?	in the stadium?	2817	3550	0.775
Target subject	ʔaʕni alwalad atˤawi:l	I mean the tall boy	4512	6083	1.419

TIME

Figure 2. Components in auditory sentence stimuli and their duration in ms.

Each sentence was accompanied by a visual display of four objects: the target and three distractors (Figure 3). In prediction trials, the target (one tall boy) matched the verb number, gender, and adjective, while the distractor objects had mismatched verb number (two boys or three boys) or verb gender (one girl). In other words, in prediction trials, participants saw images of one boy, two boys, three boys, and one girl. In baseline trials, the target (one tall boy) matched the verb number, gender, and the adjective, whereas the distractors mismatched the verb number (two boys or three boys) or the post-nominal adjective (one short boy). Simply put, in baseline trials, participants saw images of one tall boy, one short boy, two boys, and three boys. Visual displays for filler sentences mimicked those for the experimental sentences.

Two counterbalanced lists were created per experiment, which presented the items in a different order. Trial order was pseudo-randomized such that adjacent trials did not display the target picture in the same region and did not repeat the trial condition (prediction, baseline, filler) (e.g., Lukyanenko & Fisher, 2016). Participants were randomly assigned to one of the lists.

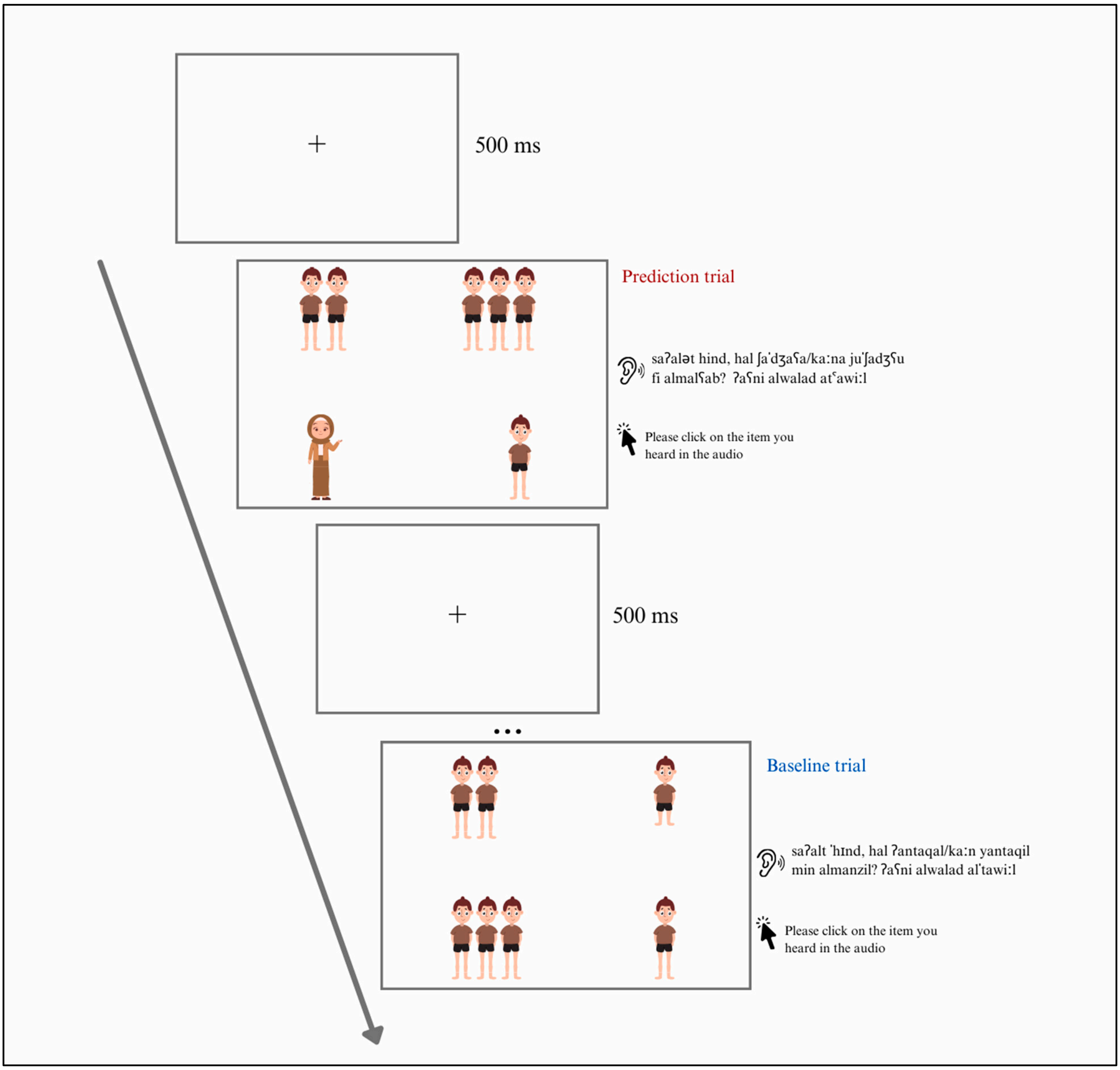

Figure 3. Trial example.

3.2.2. Comprehension Tasks

Two comprehension tasks were used to check for participants' attention during the web-based experiments. First, participants were asked in each trial to click on the object heard in the auditory sentence (e.g., Ito et al., 2023). Second, participants were provided with simple yes/no comprehension questions after half of the filler trials in each experiment. The questions always asked about the adverbial phrase in the filler sentence to avoid drawing the participants' attention to the subject. For example, if the filler sentence was "the girls shopped in the mall/in the evening", the question would be "did the girls shop in the supermarket/in the morning?". Half of the questions should be answered "yes" and the other half "no".

3.3. VWP Task Design

To establish the predictive processing effect, we contrasted target fixations in each experiment across two trial conditions: prediction trials and baseline trials. The prediction trial condition contained a 3SG.M verb and only one singular masculine object (the tall boy). This allowed participants to anticipatorily direct their gaze to the object based on verb number morphology before hearing the subject noun phrase (NP). The baseline trial condition presented a 3SG.M verb and two singular masculine objects (a tall boy and a short boy). The number information encoded in the verb was not informative in baseline trials, and participants had to listen to the disambiguating adjective cue (tall/short) at the end of the subject NP to accurately identify the target object. Each trial condition presented eight distinct sentences.

Proficiency Tests

This study administered two Arabic proficiency tests to L2 participants to overcome potential variance in test performance due to the test method (Brysbaert, 2024).

3.4. C-Test

The first L2 Arabic proficiency measure was the C-test. This test format is argued to be an objective, reliable, and valid tool to estimate overall L2 ability (e.g., Eckes & Grotjahn, 2006; Tidball & Treffers-Daller, 2008). A C-test was adapted from a previous study (Raish, 2017), which showed that it correlates well with self-reported Arabic ability ($r = 0.63$). The C-test included three short Arabic texts in which every second word was half-deleted. There were 25 gaps per text. The possible minimum score on the C-test was 0, and the maximum score was 75. The C-test was scored using a binary method (correct/incorrect). Responses were first scored automatically and then manually reviewed. Misspelled responses were considered incorrect.

Due to a technical issue, 38 out of the 40 L2 participants completed the test. The participants had an average score of 35 on the C-test ($SD = 4.63$, range = 3–61). In the present study, the C-test demonstrated good split-half reliability ($\lambda 6 = 0.93$) and internal consistency reliability ($\omega = 0.95$; $\alpha = 0.94$, 95% CI [0.91, 0.96]).

3.5. LexArabic

The second L2 Arabic proficiency measure was LexArabic (Alzahrani, 2023), a short Arabic vocabulary test inspired by LexTALE (Lemhöfer & Broersma, 2012). This vocabulary test was validated in a similar L2 Arabic learners group, showing moderate correlations with an objective general Arabic proficiency test ($r = 0.39$) and total self-ratings ($r = 0.41$) (Alzahrani, 2023). LexArabic included 90 Arabic words (60 real words; 30 nonwords), and participants had to select the words that they use or know. Test-takers could achieve a score ranging from a minimum of -100 (if only all nonwords were selected) to a maximum of 100. LexArabic was scored using the formula provided in Alzahrani (2023).

Due to a technical issue, only 36 out of the 40 L2 participants completed the test. The participants had an average score of 32 ($SD = 27.43$, range = -65–80) on LexArabic. In the current study, LexArabic showed high split-half reliability ($\lambda 6 = 0.96$) and internal consistency reliability ($\omega = 0.99$; $\alpha = 0.98$, 95% CI [0.97, 0.99]).

Background Questionnaire

This study administered the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007) to obtain participants' demographic and linguistic information. L1 and L2 speakers were asked to rate their Arabic reading, speaking, listening, and writing skills on an 11-point scale (minimum = 0 "none", maximum = 10 "perfect").

The L2 participants were additionally asked to indicate their L1, years spent learning Arabic, age of first exposure to Arabic, experience in Arabic-speaking countries and work environments, and the order in which they acquired languages. L2 speakers were also asked about their caregivers' primary language during childhood and the language of instruction from kindergarten to university.

3.6. Procedure

Participants completed the study online via Gorilla.sc (Anwyl-Irvine et al., 2020). They read task instructions and were informed to keep their webcam enabled and maintain a steady head position throughout the experiment. Then, they completed a calibration task, which was allowed to be re-taken up to three times if the previous attempt failed. Participants were automatically rejected if the calibration task was not completed successfully.

After calibration, participants completed an audio identification task in which they listened to an auditory sentence and were asked to select the sentence from four options. This was done to ensure that participants could hear the audio stimuli (Slim & Hartsuiker, 2022). Then, participants completed a two-trial practice session to familiarize them with the main task.

After the trial session, a centrally located fixation point appeared for 500 ms, followed by the simultaneous presentation of the visual display and the auditory stimulus sentence (Figure 3). Participants were instructed to click on the object that was mentioned in the audio. The visual display remained on the screen until participants clicked on the object. A green mark was shown in the bottom center of the screen if participants responded correctly; if not, they saw a red cross mark. When the display disappeared in experimental trials, a fixation point appeared in the center of the screen, indicating a new trial.

In filler trials, when the display disappeared, participants saw a comprehension question after completing half of the filler trials, with "Yes" and "No" boxes provided below. Correct responses were signaled by a green check mark in the bottom center of the screen, and incorrect responses by a red cross mark. Then, a centrally located fixation point appeared, signaling the next trial.

During the eye-tracking task, participants were given an optional break after completing half of the experiment (16 trials). The break was followed by a calibration task to ensure a proper estimation of participants' gaze throughout the task.

After the eye-tracking task, L1 participants completed the background questionnaire. Meanwhile, L2 participants completed the C-test, then LexArabic, and the background questionnaire. The experiment took around 30 min for L1 speakers and 60 min for L2 speakers.

3.7. Data Analysis

Eye-tracking data were cleaned in three steps (Prystauka et al., 2023). First, incorrectly answered trials were removed, resulting in a loss of 7.45% of the data. Second, participants with a low sampling rate (<5 Hz; n = 6) were excluded, as they would have reduced the quality of gaze data. Third, trials with high convergence values (n = 0) or low support vector machine classifier scores (n = 1) were checked and removed. Convergence and classifier scores are two quality metrics generated by Gorilla. Convergence values measure the model's confidence in finding a face. Classifier scores represent the degree to which the image captured by the model matches a face. In the current study, participants had acceptable sampling rates similar to those reported in prior research (M = 19.7 Hz, *SD* = 6.4 Hz, range 5.5–30.1 Hz in Prystauka et al. (2023)). The L1 group had a mean sampling rate of 23.78 Hz (*SD* = 5.15, range = 12.68–30.06 Hz), while the L2 group had a mean sampling rate of 18.96 Hz (*SD* = 6.91, range = 5.32–39.29 Hz).

All data analyses were conducted in R version 4.2.2. First, we utilized cluster-based permutation analysis (CPA) (e.g., Koch et al., 2023) to investigate whether there was a difference between the trial conditions (prediction vs. baseline). Second, we used linear mixed effects (LME) models (Garrido Rodriguez et al., 2023) to determine whether variables such as speaker group, verb aspect, and L2 proficiency influenced target fixations.

CPA is a non-parametric method that detects the existence of a statistical difference between conditions in time data but does not pinpoint the emergence of an effect (Ito & Knoeferle, 2022; Maris & Oostenveld, 2007; Sassenhagen & Draschkow, 2019). One important advantage of CPA is that it handles autocorrelation in eye-tracking data (Stone et al., 2021). We used the R package permutes (Voeten, 2023) to perform CPA. Four CPA models were built separately for each pair of Group (L1 speakers, L2 speakers) and Experiment (Exp 1, Exp 2). All CPA models were set with 10,000 iterations and included the dependent variable target fixation (no = −0.5; yes = 0.5), the independent variable trial type (baseline = −0.5; prediction = 0.5), and random slopes for participants and items.

Further, we built logistic LME models using the R package lme4 (Bates et al., 2015) since the dependent variable target fixation is binary (no = −0.5; yes = 0.5). In the main models, the fixed effects included trial type (baseline = −0.5; prediction = 0.5), scaled time ($M = 0$, $SD = 1$), group (–0.5 L1 speakers, 0.5 L2 speakers), and experiment (Exp 1 = −0.5; Exp 2 = 0.5). The proficiency models included the same fixed effects except for group, and also added standardized C-test scores ($M = 0$, $SD = 1$) and standardized LexArabic scores ($M = 0$, $SD = 1$). The maximal random structure justified by the design was initially included for all models (Barr et al., 2013), and only models that converged were reported.

To investigate predictive processing, we used two time windows for analysis (Garrido Rodriguez et al., 2023). The first time window (TW1) included 200 ms after the onset of the verb and ended 200 ms after the offset of the complement. We added 200 ms after word onsets and offsets to account for time to initiate saccades (Saslow, 1967). In line with the adopted definition of prediction, TW1 is the predictive window in the current study. Target fixations in prediction trials during TW1 would constitute evidence for predictive number processing. The second time window (TW2) started 200 after the auditory presentation of the target noun and finished 200 ms after the offset of the noun. The analysis of TW2 can capture another difference between the two trial conditions and examine the L2 proficiency effect in more detail. In TW2, we expect that L1 and L2 speakers will look more at the target picture during both trial conditions (prediction and baseline trials) since they have just heard the subject NP.

4. Results

This section outlines the results of our analyses. The interpretations of these results are provided in detail in the discussion section.

4.1. Background Information

Table 1 provides demographic and linguistic information for L1 participants, while Table 2 presents data for L2 participants. Between-group comparisons were conducted separately per speaker group using the Kruskal–Wallis test, as the data were non-normally distributed. None of the examined variables showed a significant difference between L1 speakers across experiments. Likewise, there was no significant difference between the L2 speakers across experiments.

Table 1. L1 participants' background and linguistic information across experiments.

	Exp 1 (n = 9)	Exp 2 (n = 9)	Combined	Between-Groups Comparisons
	Mean (*SD*)			
Age	27.78 (7.21)	30.89 (7.56)	29.33 (7.34)	χ^2 = 0.949
Speaking	9.11 (1.05)	9.00 (0.87)	9.06 (0.94)	χ^2 = 0.176
Listening	9.22 (1.09)	9.33 (0.71)	9.28 (0.89)	χ^2 = 0.009
Reading	9.11 (1.62)	9.00 (1.58)	9.06 (1.55)	χ^2 = 0.152
Writing	8.67 (1.66)	8.22 (2.22)	8.44 (1.92)	χ^2 = 0.170
Total self-ratings (/40)	36.11 (5.01)	35.56 (4.48)	35.83 (4.62)	χ^2 = 0.161

Table 2. L2 participants' background and linguistic information across experiments.

	Exp 1 (n = 20)	Exp 2 (n = 20)	Combined	Between-Groups Comparisons
	Mean (*SD*)			
Age	25.81 (2.90)	26.33 (4.65)	26.09 (3.88)	χ^2 = 0.007
Speaking	7.19 (2.01)	6.89 (1.94)	7.03 (1.95)	χ^2 = 0.031
Listening	7.56 (1.86)	7.33 (2.00)	7.44 (1.91)	χ^2 = 0.100
Reading	7.69 (1.45)	7.56 (1.98)	7.62 (1.72)	χ^2 = 0.011
Writing	7.38 (2.09)	7.61 (1.85)	7.50 (1.94)	χ^2 = 0.025
Total self-ratings (/40)	29.81 (6.73)	29.39 (7.23)	29.59 (6.89)	χ^2 = 0.043
Age started learning MSA	11.20 (5.33)	12.94 (6.56)	12.15 (6.01)	χ^2 = 0.500
Years in Arabic-speaking countries	3.06 (2.17)	6.94 (7.30)	5.12 (5.78)	χ^2 = 0.969
Years in Arabic-speaking environments	5.19 (4.49)	8.22 (8.29)	6.79 (6.85)	χ^2 = 0.097

Note: Six L2 speakers did not complete the linguistic questionnaire.

4.2. Comprehension Task

Both groups performed well on the "click the mentioned object" task. L1 speakers achieved a mean accuracy of 97% in Exp 1 (*SD* = 0.05, range = 84–100) and 95% in Exp 2 (*SD* = 0.05, range = 81–97). L2 speakers scored a mean accuracy of 95% in Exp 1 (*SD* = 0.05, range = 72–100) and 94% in Exp 2 (*SD* = 0.03, range = 88–97). Performance on the yes/no comprehension task was also great. L1 speakers exhibited a mean accuracy of 86% in Exp 1 (*SD* = 0.11, range = 71–100) and 96% in Exp 2 (*SD* = 0.06, range = 85–100). Meanwhile, L2 speakers had a mean accuracy of 88% in Exp 1 (*SD* = 0.09, range = 71–100) and 90% in Exp 2 (*SD* = 0.10, range = 71–100).

4.3. Eye-Tracking Data: Descriptive Statistics

Table 3 presents the proportion of fixations to the target picture as per speaker group, experiment, and trial condition during TW1. Across experiments, L1 and L2 Arabic speakers showed more looks at the target object in prediction trials than in baseline trials. However, this difference in target fixations between trial conditions was stronger in the L1 group than in the L2 group. Within experiments, all participants showed a larger difference in target fixations between trial conditions in Experiment 1 rather than in Experiment 2. In addition, only native Arabic speakers showed increased target fixations (their eyes fixated more on the target object image) in prediction trials relative to baseline trials in Experiment 2. In contrast, the proportion of L2 speakers' gaze on the target picture remained somewhat consistent across trial conditions in Experiment 2.

Table 3. Proportions and standard deviations of fixations to the target as a function of group, experiment, and trial condition.

Group	Exp	Baseline Trials	Prediction Trials
		***M* (*SD*)**	***M* (*SD*)**
L1 speakers	Exp 1	0.26 (0.44)	0.37 (0.48)
	Exp 2	0.25 (0.43)	0.32 (0.46)
L2 speakers	Exp 1	0.22 (0.41)	0.29 (0.45)
	Exp 2	0.25 (0.43)	0.25 (0.43)

The proportion of eye fixations on the target picture in TW1 as a function of L2 Arabic proficiency is summarized in Table 4. As shown in this table, L2 participants tended to direct more looks in prediction trials than baseline trials regardless of their proficiency level as measured in the C-test and LexArabic. Crucially, during prediction trials, L2 speakers tended to fixate more on the target picture with increasing C-test/LexArabic scores.

Table 4. Proportions and standard deviations of fixations to the target as a function of experiment, L2 Arabic proficiency level, and trial condition.

Test	Exp	Proficiency Level	Baseline Trials	Prediction Trials
			***M* (*SD*)**	***M* (*SD*)**
C-test	Exp 1	−1 *SD*	0.21 (0.41)	0.32 (0.46)
		Mean	0.23 (0.42)	0.27 (0.44)
		+1 *SD*	0.20 (0.40)	0.34 (0.47)
	Exp 2	−1 *SD*	0.12 (0.33)	0.14 (0.35)
		Mean	0.25 (0.43)	0.27 (0.44)
		+1 *SD*	0.24 (0.43)	0.21 (0.41)
LexArabic	Exp 1	−1 *SD*	-	-
		Mean	0.20 (0.40)	0.27 (0.44)
		+1 *SD*	0.31 (0.46)	0.39 (0.48)
	Exp 2	−1 *SD*	0.11 (0.32)	0.14 (0.35)
		Mean	0.26 (0.43)	0.26 (0.44)
		+1 *SD*	0.29 (0.45)	0.29 (0.45)

−1 *SD* = lower proficiency, mean score = intermediate proficiency, +1 *SD* = higher proficiency.

4.4. Eye-Tracking Data: CPA Results

Figures 4 and 5 show the time-course plot of the proportion of fixations to the target picture (i.e., singular objects) when participants heard a singular verb, e.g., na:ma/slept.3SG.M, in prediction trials compared to fixations to the same object in baseline trials. See Appendices A–D for detailed figures showing fixations on the target and other objects.

The L1 speakers, as shown in Figure 4, started to fixate more on the appropriate object during the *verb + complement* region in prediction trials and continued to increase their looks during the *noun* region and until the end of the sentence. In contrast, in baseline trials, L1 speakers increased their attention toward the target once the noun was mentioned and continued until the sentence finished.

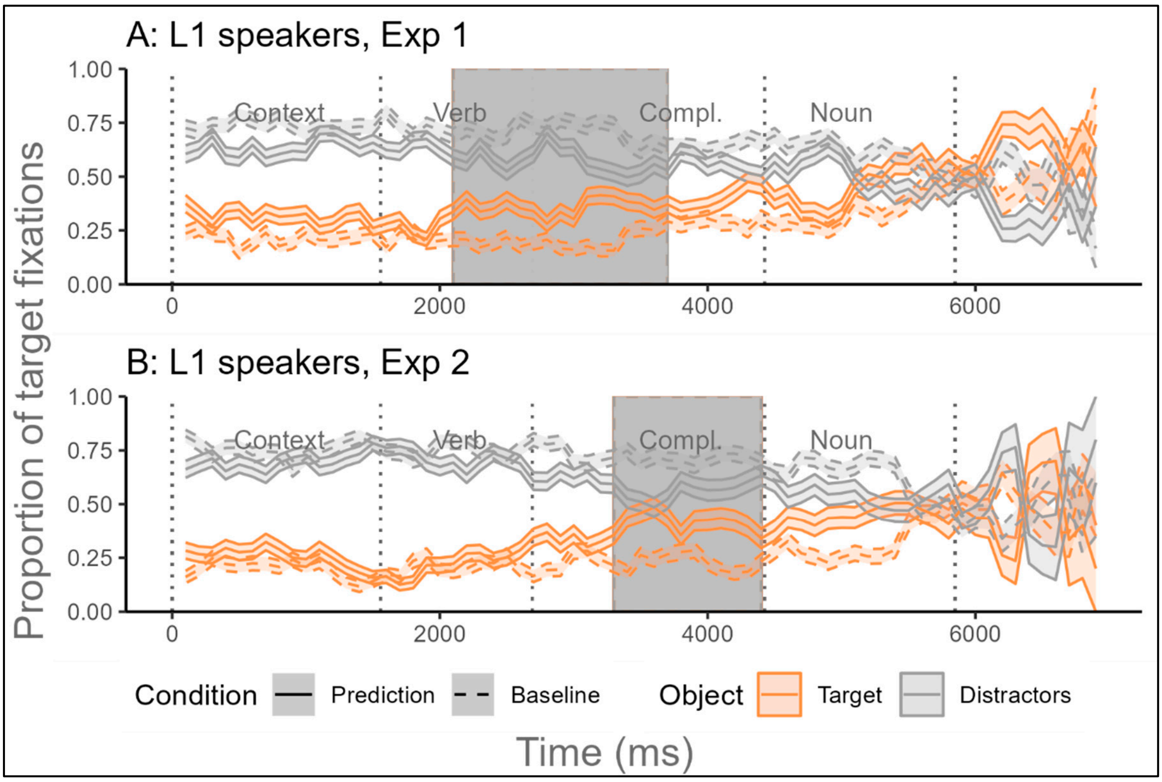

Figure 4. Time course of fixation proportions for target in the prediction (solid orange lines) and baseline trial conditions (dotted orange lines) for L1 speakers across experiments. Ribbons indicate the standard error. Dotted lines indicate the mean onset and offset of word durations in the sentences. The grey-shaded area indicates a significant CPA cluster. Compl.: complement.

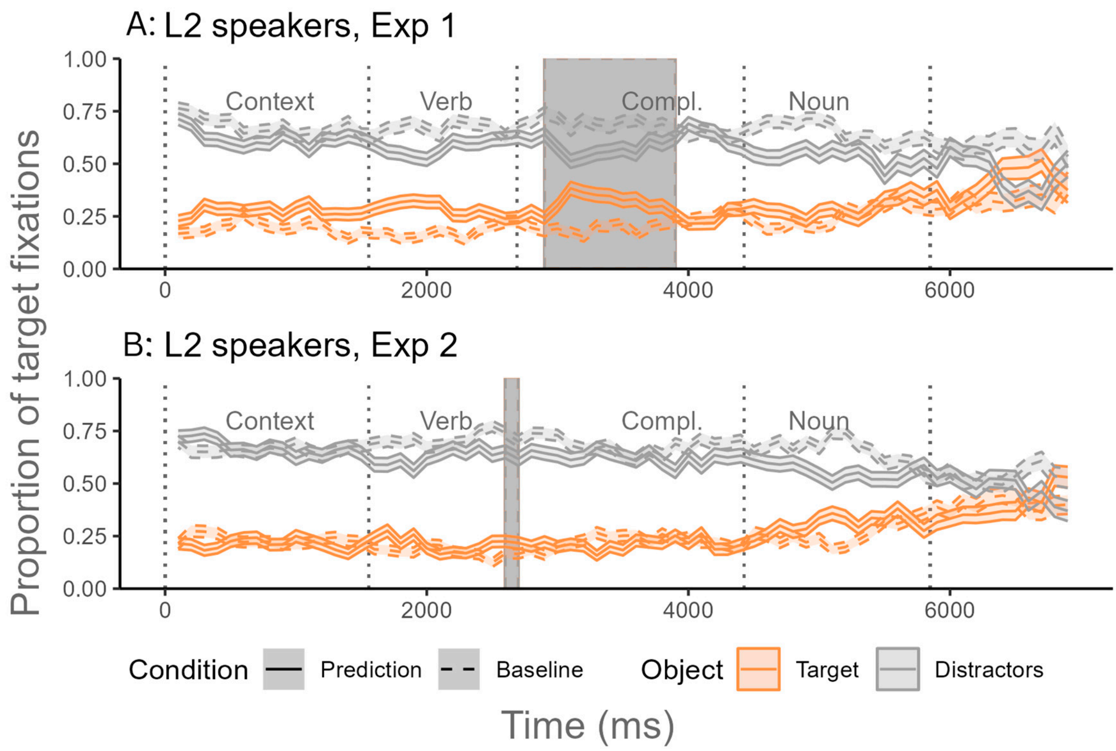

Figure 5. Time course of fixation proportions for target in the prediction (solid orange lines) and baseline trial conditions (dotted orange lines) for L2 speakers across experiments. Ribbons indicate the standard error. Dotted lines indicate the mean onset and offset of word durations in the sentences. The grey-shaded area indicates a significant CPA cluster. Compl.: complement.

The CPA results indicated a significant difference in target fixations between prediction and baseline trials among L1 speakers. This difference in fixations started from the *verb* region in Exp 1, while it emerged in the *complement* region in Exp 2.

On the other hand, the L2 participants directed more looks to the target object at the beginning of the *complement* region in prediction trials in Exp 1 (Figure 5; plot A). In contrast to L1 speakers, this increase in target looks leveled off during the *noun* region and started rising again towards the end of the sentence. In baseline trials in Exp 1, L2 speakers fixated on the target picture only after the onset of the noun. Unlike L1 speakers, the CPA results for L2 Arabic speakers showed that a divergence in target fixations between prediction and baseline trials was only significant during the *complement* region in Exp 1.

In Exp 2, the L2 speakers showed limited target fixations in prediction trials during the *verb* + *complement* region (Figure 5; plot B). In fact, the CPA results showed that L2 participants in Exp 2 directed significantly more looks at the target picture in the baseline trials than in prediction trials at some point (3400 ms) in the *complement* region. In contrast to the observed strong predictive processing among native Arabic speakers in Exp 2, L2 Arabic speakers exhibited limited anticipation.

Overall, this pattern of fixations indicates that both L1 and L2 speakers can predictively use number information on verbs to anticipate the upcoming noun, with some differences across experiments (i.e., linguistic stimuli type). Further, all participants directed their attention towards the target picture in both trial conditions after hearing the noun in the sentence and continued to do so until the end of the sentence. Crucially, the L1 speakers showed earlier (Exp 1) or stronger (Exp 2) target fixations compared to L2 speakers.

4.5. Eye-Tracking Data: LME Results

Table 5 summarizes the results of the LME models over two time regions: verb region (TW1), and noun region (TW2).

Table 5. Results from the logistic mixed effects model across two time regions.

	Verb Region			**Noun Region**		
	Estimate (95% CI)	***SE***	***z***	**Estimate (95% CI)**	***SE***	***z***
Intercept	−1.28 [−1.75, −0.80]	0.24	−5.29 ***	−2.02 [−2.49, −1.56]	0.24	−8.52 ***
Trial condition	0.55 [−0.36, 1.46]	0.46	1.19	0.64 [−0.25, 1.53]	0.45	1.40
Time	0.36 [0.31, 0.41]	0.03	13.50 ***	1.09 [0.93, 1.24]	0.08	13.90 ***
Group	−0.35 [−0.61, −0.10]	0.13	−2.72 **	−0.06 [−0.54, 0.41]	0.24	−0.26
Exp	−0.22 [−0.46, 0.01]	0.12	−1.84	−0.12 [−0.43, 0.19]	0.16	−0.74
Trial × Time	0.06 [−0.05, 0.16]	0.05	1.06	−0.12 [−0.42, 0.19]	0.16	−0.76
Trial × Group	−0.37 [−0.46, −0.27]	0.05	−7.53 ***	−0.26 [−1.04, 0.52]	0.40	−0.65
Time × Group	−0.34 [−0.44, −0.23]	0.05	−6.35 ***	−0.51 [−0.81, −0.20]	0.16	−3.25 **
Trial × Time × Group	−0.88 [−1.09, −0.67]	0.11	−8.27 ***	−0.05 [−0.66, 0.56]	0.31	−0.16

* $p < 0.05$, ** $p < 0.01$. *** $p < 0.001$. TW1 model formula: glmer (AOI ~ trial_type × grandscaledtime × group + EXP + (1 | participant_id) + (1 | EXP_item_ID), data = dat_exp_verb, family = "binomial"); TW2 model formula: glmer (AOI ~ trial_type × grandscaledtime × group + EXP + (trial_type | participant_id) + (trial_type | EXP_item_ID), data = dat_exp_object, family = "binomial").

4.5.1. RQ1: L1 and L2 Predictive Number Processing

There were significant differences between the two speaker groups in the verb region. The main effect of time ($\beta = 0.36$, $SE = 0.46$, 95% CI [0.31, 0.41]) suggests that all participants, on average, directed more attention to the target picture over time across trial conditions (prediction, baseline) and experiments (Exp 1, Exp 2) in TW1. However, the significant effect of group ($\beta = -0.35$, $SE = 0.13$, 95% CI [−0.61, −0.10]) indicates that L1 speakers fixated more on the target object during TW1 compared to L2 speakers.

The non-significant trial x time ($\beta = 0.06$, $SE = 0.05$, 95% CI [−0.05, 0.16]) indicates no significant increase in looks in prediction trials relative to baseline trials over time across group and Exp in TW1. This result might be explained by the great variability in fixation behavior between L1 and L2 participants, especially during TW1, with L2 speakers showing less predictive looks than L1 speakers (see Figures 4 and 5). This variability might have attenuated the average increase in looks in prediction trials across groups and Exp during TW1.

The significant trial–group interaction ($\beta = -0.37$, $SE = 0.05$, 95% CI [−0.46, −0.27]) in TW1 suggests that L1 Arabic participants directed more looks at the target picture in prediction trials than in baseline trials across experiments, while L2 speakers did not show a comparable difference. Likewise, the significant time–group interaction ($\beta = -0.34$, $SE = 0.05$, 95% CI [−0.44, −0.23]) reveals that L1 speakers directed more looks to the target object over time across experiments compared to L2 speakers in TW1.

A significant trial–time–group interaction ($\beta = -0.88$, $SE = 0.11$, 95% CI [−1.09, −0.67]) found that L1 speakers were more likely to fixate on the target picture in prediction trials than in baseline trials over time in TW1 regardless of Exp. This result shows that Arabic L1 participants predictively used the verbal number marking to anticipate the target object before it was mentioned during TW1 (Figure 6). In contrast, when L2 speakers heard a singular verb, they did not use verbal number information as a predictive cue to anticipate the target object. Instead, they directed more attention to the target picture over time in baseline trials compared to prediction trials.

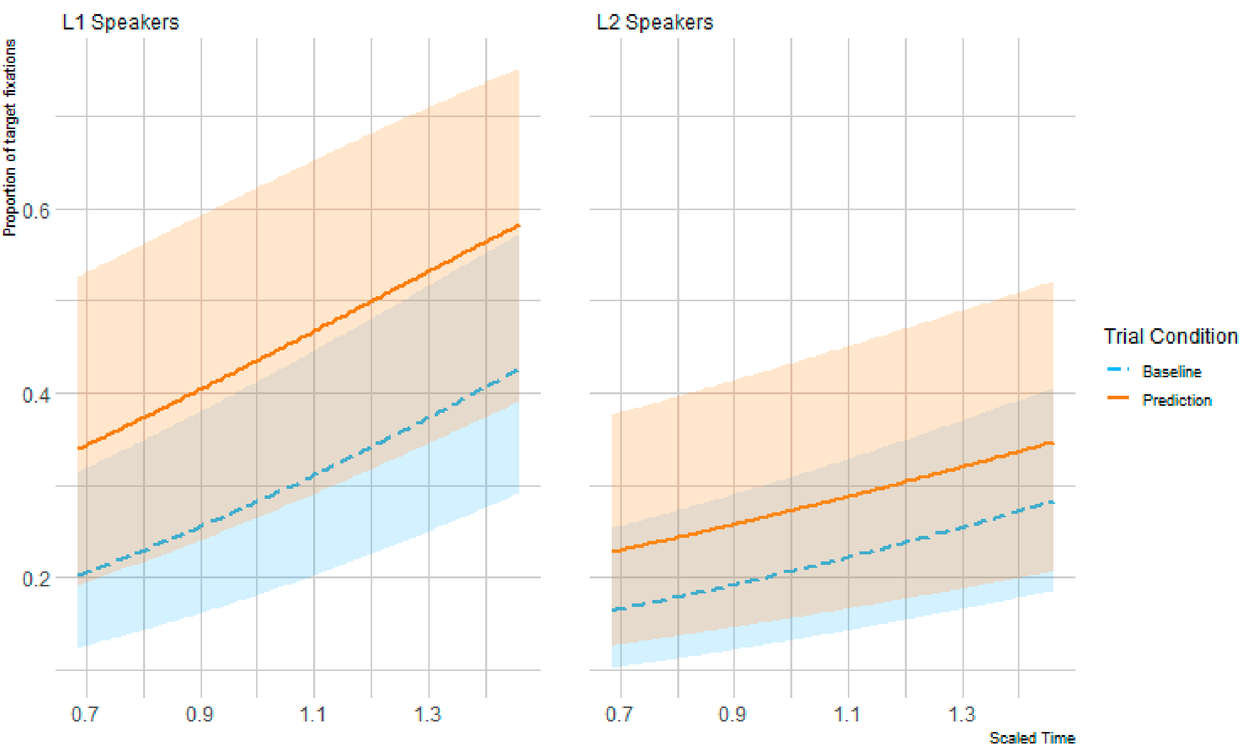

Figure 6. Predicted proportion of target fixations as a function of Group (L1 speakers, L2 speakers), trial condition, and time in TW1. Bands represent 95% confidence intervals.

Meanwhile, there was only one significant difference between the two speaker groups in the noun region. The significant time–group interaction ($\beta = -0.51$, $SE = 0.16$, 95% CI [−0.81, −0.20]) indicated that native Arabic speakers increased their target fixations as more information became available over time in TW2, unlike the L2 participants.

4.5.2. RQ2: Effect of Verb Aspect

The effect of verb aspect on L1 and L2 predictive number processing was not statistically significant in the verb model ($\beta = -0.23$, $SE = 0.11$, 95% CI [−0.44, −0.02]) and the object model ($\beta = -0.01$, SE = 0.14, 95% CI [−0.29, 0.27]). L1 and L2 speakers showed similar fixation patterns when they heard the simple past 3SG.M verb (Exp 1) and the past progressive 3SG.M verb (Exp 2) in TW1 and TW2.

4.5.3. RQ3: L2 Proficiency Effect

The L2 proficiency effect, as measured in the C-test and LexArabic, was examined in two separate models per time region. In the verb model, C-test scores ($\beta = 0.09$, $SE = 0.10$, 95% CI [−0.10, 0.28]) and LexArabic scores ($\beta = 0.15$, $SE = 0.09$, 95% CI [−0.03, 0.34]) were non-significant. In the noun model, only LexArabic scores emerged as a significant main effect ($\beta = 0.28$, $SE = 0.10$, 95% CI [0.09, 0.47]), suggesting that those with better vocabulary knowledge directed more looks towards the target image across experiments and trial conditions (Figure 7).

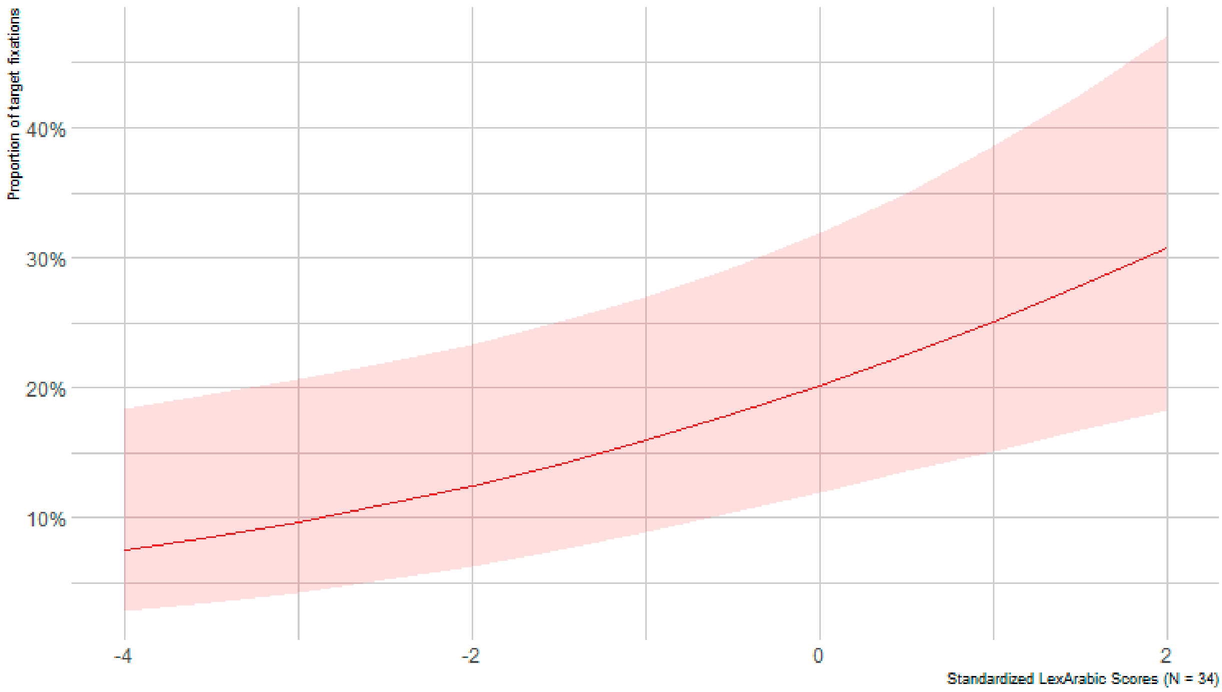

Figure 7. Predicted proportion of target fixations as a function of LexArabic scores among L2 Arabic speakers in TW2. The band represents the 95% confidence interval.

5. Discussion

This study examined whether L1 and L2 Arabic speakers could predict the number of the next noun based on the verb suffix and whether there were L1-L2 differences in anticipatory processing (RQ1). The effects of verb aspect (Exp 1 = simple past verb, Exp 2 = past progressive verb; RQ2) and L2 proficiency (C-test, LexArabic; RQ3) were also investigated. Both speaker groups anticipated number information based on verb endings, but there were differences in the speed of anticipation between the two groups. Further, L1 participants showed predictive number processing across both verb types, while L2 speakers exhibited either delayed or limited prediction. L2 proficiency did not mediate prediction and only influenced L2 speakers' target fixations in the noun region. These findings are discussed below in relation to experimental and theoretical proposals.

5.1. L1 vs. L2 Prediction

The present study extended the evidence for L1 and L2 morphosyntactic prediction to a less explored language group (Arabic speakers) and morphosyntactic domain (grammatical number). CPA results indicated that L1 Arabic speakers fixated earlier and stronger on the target image during TW1 across experiments, while L2 Arabic speakers' prediction effects were delayed (Exp 1) or not present (Exp 2). The delayed anticipatory looks among L2 Arabic speakers are in line with previous findings, suggesting that L2 comprehenders from different languages, such as Arabic (the current study) and German (Koch et al., 2021, 2023), exhibit delayed predictive number processing relative to L1 speakers. However, the finding that L2 Arabic speakers demonstrated no prediction in Exp 2 diverged from prior predictive number processing research. Across-experiment variation in prediction will be discussed below.

The observed differences in the time course of number prediction between L1 and L2 Arabic speakers could be attributed to various reasons. First, under the utility account (Kuperberg & Jaeger, 2016), speaker groups are expected to show differences in predictive behaviors based on their prior language experience. Unlike native speakers, L2 speakers are widely reported to display more variation in their L2 skills and language processing (Dornyei & Ryan, 2015; Tagarelli et al., 2016). This larger discrepancy in L2 performance might have attenuated the onset of a significant prediction effect.

Second, the L2 participants might have less established knowledge of the Arabic grammatical number system (Ivanova & Costa, 2008), leading to the observed delayed effect. Previous research has shown that only L2 German speakers who showed target-like production of grammatical gender information exhibited anticipatory effects comparable to L1 speakers (Hopp, 2013). Third, the absence of number-marked verbs in the majority of the L2 participants' L1s (n = 22) might have slowed down their ability to rapidly extract number information from verb endings. Dussias et al. (2013) reported that similarities between the L1 and the L2 in the gender system enhanced L2 speakers' predictive processing of gender information, indicating a key role for cross-linguistic influence on L2 anticipatory processing.

A fourth potential explanation for the delayed L2 prediction is that morphosyntactic information is more difficult to anticipate. According to the prediction-by-production account, comprehenders generate predictions starting from earlier (i.e., semantic encoding) to later stages (e.g., syntactic encoding) of language processing (Pickering & Gambi, 2018). As such, L2 speakers might take some time to access the later stages of the production process, hindering their ability to anticipate morphosyntactic information in time. This makes L2 morphosyntactic prediction more cognitively demanding compared to L1 prediction (Ito & Pickering, 2021; Mitsugi & Macwhinney, 2016). Fifth, as pointed out by a reviewer, it is possible that the L2 speakers did not fixate on the target due to the current design. The current study employed three image distractors in all experiments, and this could have increased the time needed to process the visual scene, diverting the L2 speakers' attention away from the target. Future research could explore the impact of distractor quantity (e.g., three vs. one) on L2 speakers' target fixations in VWP designs.

To summarize, the current findings suggest both speaker groups anticipated the number of the upcoming noun, yet there were consistent L1-L2 differences. CPA results showed that L1 Arabic speakers anticipated a singular subject when they heard singular past verbs, and this effect persisted across different forms of the past verb. Meanwhile, L2 speakers demonstrated delayed prediction of verb-based number information or no prediction across the verb forms. The reported L1-L2 variation in number predictive processing is consistent with prior findings in the literature (Koch et al., 2021, 2023).

5.2. Effect of Verb Aspect

The present study examined whether L1 and L2 Arabic speakers would show differential prediction effects when the number cue was presented on verbs with varying aspects: simple past and progressive past. CPA results revealed that L1 speakers demonstrated predictive effects regardless of verb aspect, while L2 speakers' prediction was influenced by verb aspect. Previous studies have considered different cues from the same domain (Brouwer et al., 2017). The present study thus contributed to the literature by revealing a potential role of cue form in predictive number processing.

L2 Arabic speakers were influenced by verb aspect, as shown in their distinct predictive behaviors across experiments. The L2 speakers directed more attention to the singular object after hearing the singular verb in Exp 1, but limited prediction was found in Exp 2 (Figure 4). As the same stimuli were used across experiments, this prediction pattern is not likely to have stemmed from differences in linguistic stimuli. Further, L2 participants across experiments were similar in their Arabic linguistic experience (Table 1), suggesting minimal influence of participant differences on the observed prediction effects.

Variability in L2 Arabic speakers' predictive use of number-marked verbs could be explained by the concept of cue validity. The Competition Model (Bates & MacWhinney, 1989) proposes that the availability and reliability of cues in spoken and written language affect the validity of the cue, which in turn affects the strength of the cue. In this view, the recruited L2 participants might have encountered the singular marker on more simple past verbs than past progressive verbs, especially since the simple past form is introduced earlier in most MSA teaching textbooks (e.g., Al-Batal et al., 2011; Al-Fawzan et al., 2014; Alosh & Clark, 2021). This higher exposure could have increased the perceived reliability of number markings on the simple past form, facilitating its predictive use among L2 speakers. On the other hand, the past progressive is not only relatively less frequently encountered but also requires attending to number information across two constituents ("ka:na yantaqil", was.3SG.M, moving.3SG.M). This requires more processing time and memory resources for L2 Arabic speakers to exploit number information from the past progressive verb.

Another explanation for the distinct L2 prediction effects across verb aspects is prediction utility. The utility approach (Kuperberg & Jaeger, 2016) argues that speakers might not generate predictions when processing costs outweigh the benefits. Here, the L2 Arabic speakers might not have utilized number information from the progressive past verb due to their reduced exposure/familiarity to this verb form. This limited engagement in predictive number processing could minimize processing costs for the L2 participants and ultimately facilitate their sentence comprehension (e.g., Grüter et al., 2020).

Taken together, verb aspect could modulate morphosyntactic predictive processing, especially for less experienced users like L2 speakers. This finding highlights the importance of investigating prediction effects across variants of the same cue, as it allows us to capture a more nuanced picture of between-individual differences.

5.3. L2 Proficiency

The current study found no significant effect of L2 proficiency (C-test scores, vocabulary test scores) on L2 predictive number processing. The influence of L2 proficiency emerged later, with more advanced L2 Arabic speakers (higher LexArabic scores) directing more target looks after the auditory presentation of the noun (β = 0.15, SE = 0.09, 95% CI [−0.03, 0.34]). Even though some studies similarly reported a limited role for L2 proficiency in morphosyntactic predictive processing (Hopp, 2015), most found a significant effect (Dussias et al., 2013; Garrido-Pozu, 2022; Henry et al., 2022; Hopp, 2016; Hopp & Lemmerth, 2018).

Several reasons could have contributed to the negligible influence of L2 proficiency in the current study. First, the used L2 proficiency measures might not have been suitable for accurately capturing overall L2 Arabic ability. Prior studies that reported a significant L2 proficiency effect mostly administered standardized language tests, which cover a more comprehensive range of L2 skills (e.g., speaking, listening, reading, writing). Second, the examined predictive cue (number marking) might be less susceptible to L2 proficiency compared to the previously investigated morphosyntactic cues. Existing studies reporting a positive significant L2 proficiency effect have been primarily conducted on the predictive use of gender (Dussias et al., 2013; Garrido-Pozu, 2022; Hopp, 2016; Hopp & Lemmerth, 2018) and case (Henry et al., 2022), which might be more influenced by L2 proficiency due to their unique properties. To illustrate, in MSA, the grammatical number system stands apart from gender and case: it directly reflects the quantity of the referent regardless of syntactic role, while gender can be semantically arbitrary for nouns, and case often indicates syntactic relationships between words (Ryding, 2005). Third, it is possible that the recruited L2 speakers might have similar L2 Arabic levels, and this limited variation might have reduced the likelihood of capturing a significant L2 proficiency effect. Although these explanations might account for the obtained null L2 proficiency effect, the actual reason remains unknown.

5.4. Limitations and Directions

This study is limited in several ways. First, the sample size and number of items per trial condition are smaller than required (see Prystauka et al., 2023, for sample size recommendations). This sample size could have influenced the magnitude of the results. The sample size was constrained by available resources, resulting in a limited number of participants and items in the study. It is thus recommended to replicate the current study with a larger sample size and number of items to ensure the generalizability of the obtained results. Second, this study only examined the singular marker, and current results might not extend to other number markers in MSA: the dual and the plural marking. Future research may consider investigating the effect of verb aspect across the different number classes in MSA to gain a better understanding of this effect.

Third, the present study used AI-generated auditory stimuli to ensure better control across stimuli. As indicated by a reviewer, the prediction effects elicited by AI-based auditory stimuli may diverge from those triggered by human-generated stimuli, since previous studies have shown that listeners adapt their predictions to the speaker (Brothers et al., 2019; Corps et al., 2023; Sala et al., 2024). It is worthwhile to examine this topic in future research.

Fourth, the L2 speakers' performance on the C-test might not be optimal due to reasons we cannot explain. More work is needed to explore more appropriate Arabic proficiency test formats for psycholinguistic research. Another concern that should be addressed in future studies is the predictive power of the available Arabic proficiency tests, to allow a more robust investigation of the L2 proficiency effect in Arabic processing research. Fifth, the L2 participants might not have ideally represented the different levels of Arabic L2 proficiency, and this could have influenced the reported L2 proficiency effect. Further research is recommended to investigate a more balanced sample of L2 Arabic speakers to assess the current results. Sixth, this study did not examine the effects of diglossia on predictive processing among L1 and L2 Arabic speakers. Future research should consider exploring the potential influence of diglossia in Arabic predictive processing as empirical evidence suggests that Arabic diglossia impacts language acquisition and processing (Abou-Ghazaleh et al., 2018; Andria et al., 2022; Asadi & Asli-Badarneh, 2023).

Finally, this study found differential number prediction effects across verb aspect, but this finding might not readily extend to other grammatical systems. It might be worthwhile to examine the predictive use of a challenging feature such as case marking across two grammatical constituents, i.e., accusative case on nouns and adjectives. The rich morphological system of MSA makes it a particularly fertile ground for investigating the interplay between morphology and prediction in language comprehension. Subsequent research could leverage this intricate system to explore, for example, how several morphological markers sharing the same grammatical function (e.g., accusative case) influence anticipatory processing.

5.5. Conclusions

This study investigated L1 and L2 Arabic speakers' predictive number processing using lexical verbs. Compared to L1 speakers, L2 speakers showed delayed (Exp 1) or limited (Exp 2) anticipatory looks based on number-marked verbs, and this effect varied by verb aspect. The current findings provide some evidence that L2 speakers' predictive processing is mediated by the specific features of the target structure. This underscores the importance of including multiple forms of the predictive cue when investigating morphosyntactic predictive processing, especially for L2 speakers. This would allow us to gain a better understanding of the factors mediating predictive processing in sentence comprehension.

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not obtain Institutional Review Board approval as the author was not affiliated with a university while conducting the study. Nevertheless, this study was conducted in compliance with the ethical regulations governing research involving human subjects in the Kingdom of Saudi Arabia, issued by the National Committee for Bioethics Research on Human Subjects in Saudi Arabia (KSU-REC). In accordance with KSU-REC and the Helsinki Declaration, this study ensured the protection of participants' rights, welfare, and confidentiality. Participants were informed about the questionnaire sections and voluntarily agreed to participate. The participants indicated that they understood their right to withdraw at any time, and their written consent was obtained before completing the experiment. The primary research protocol, which took the form of an eye-tracking experiment, did not cause any harm to the participants. Participants were identified by random ID numbers to ensure data confidentiality.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data, items, and R scripts are available on the Open Science Framework: https://osf.io/t9cpd/?view_only=808d8ea322b84b7caa16757afabd58c9.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

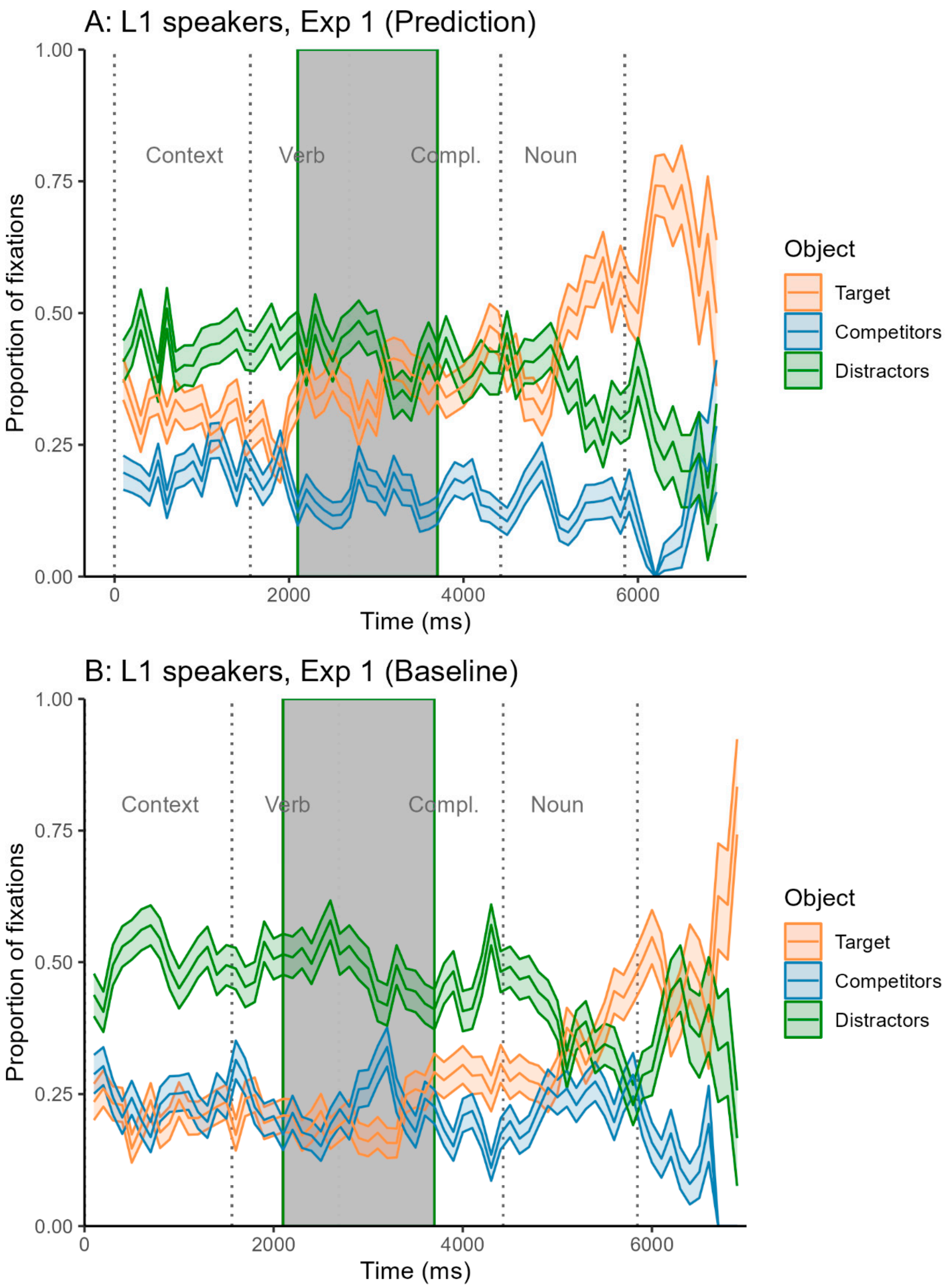

Figure A1. Time course of fixation proportions for target in the prediction (upper half figure) and baseline trial conditions (lower half figure) for L1 speakers in Experiment 1. Ribbons indicate the standard error. Dotted lines indicate the mean onset and offset of word durations in the sentences. The grey-shaded area indicates a significant CPA cluster. Compl.: complement. A competitor shares the target object's number but differs in either adjective (e.g., "the tall boy" vs. "the short boy") or gender (e.g., "the tall boy" vs. "the tall girl"), while a distractor has a different number than the target object (e.g., "the tall boy" vs. "the two/three tall boys"). In prediction trials, the competitor has the same number as the target object but a different gender. In baseline trials, the competitor matches the target object's number but has a different adjective.

Appendix B

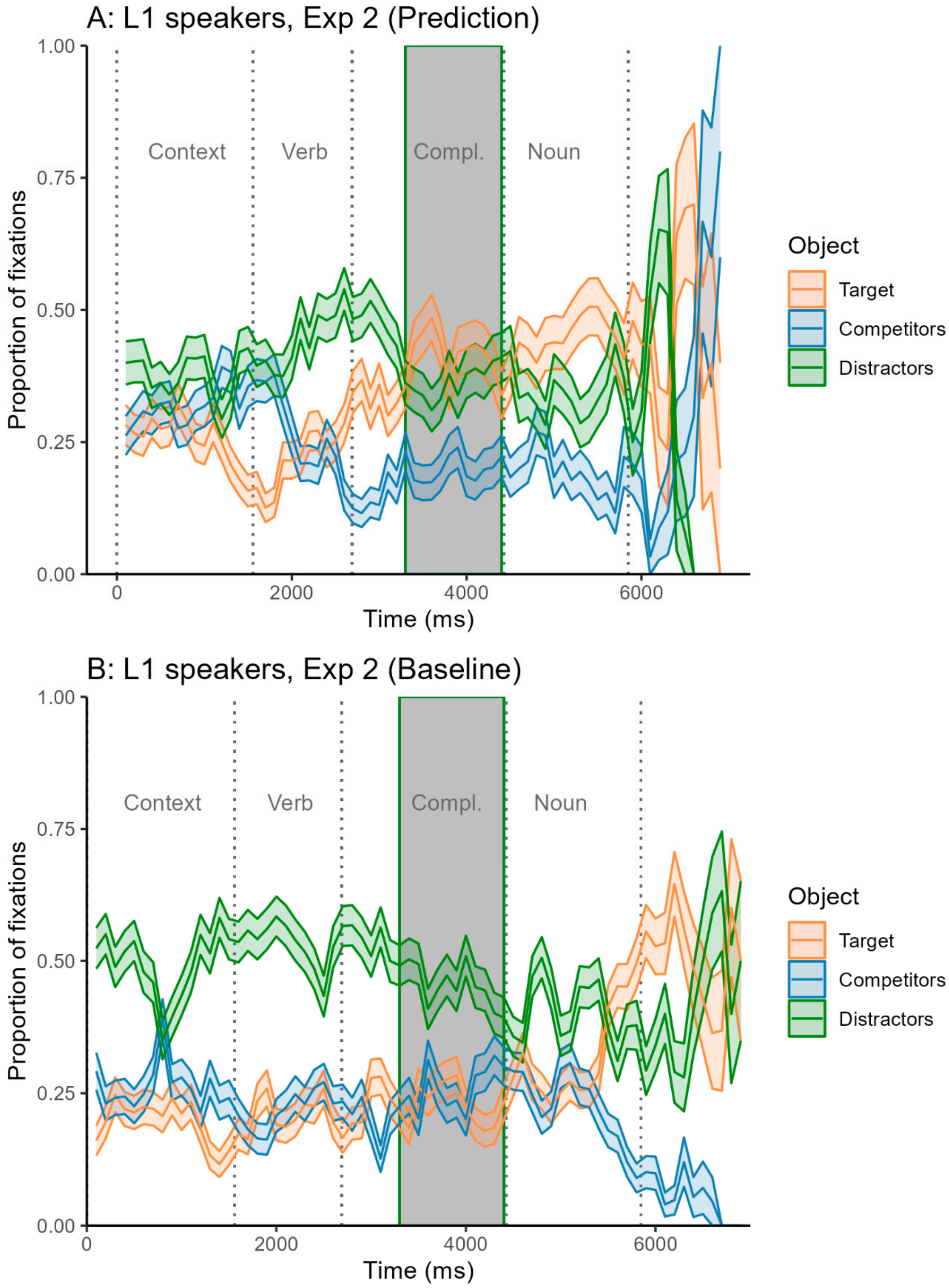

Figure A2. Time course of fixation proportions for target in the prediction (upper half figure) and baseline trial conditions (lower half figure) for L1 speakers in Experiment 2. Ribbons indicate the standard error. Dotted lines indicate the mean onset and offset of word durations in the sentences. The grey-shaded area indicates a significant CPA cluster. Compl.: complement. A competitor shares the target object's number but differs in either adjective (e.g., "the tall boy" vs. "the short boy") or gender (e.g., "the tall boy" vs. "the tall girl"), while a distractor has a different number than the target object (e.g., "the tall boy" vs. "the two/three tall boys"). In prediction trials, the competitor has the same number as the target object but a different gender. In baseline trials, the competitor matches the target object's number but has a different adjective.

Appendix C

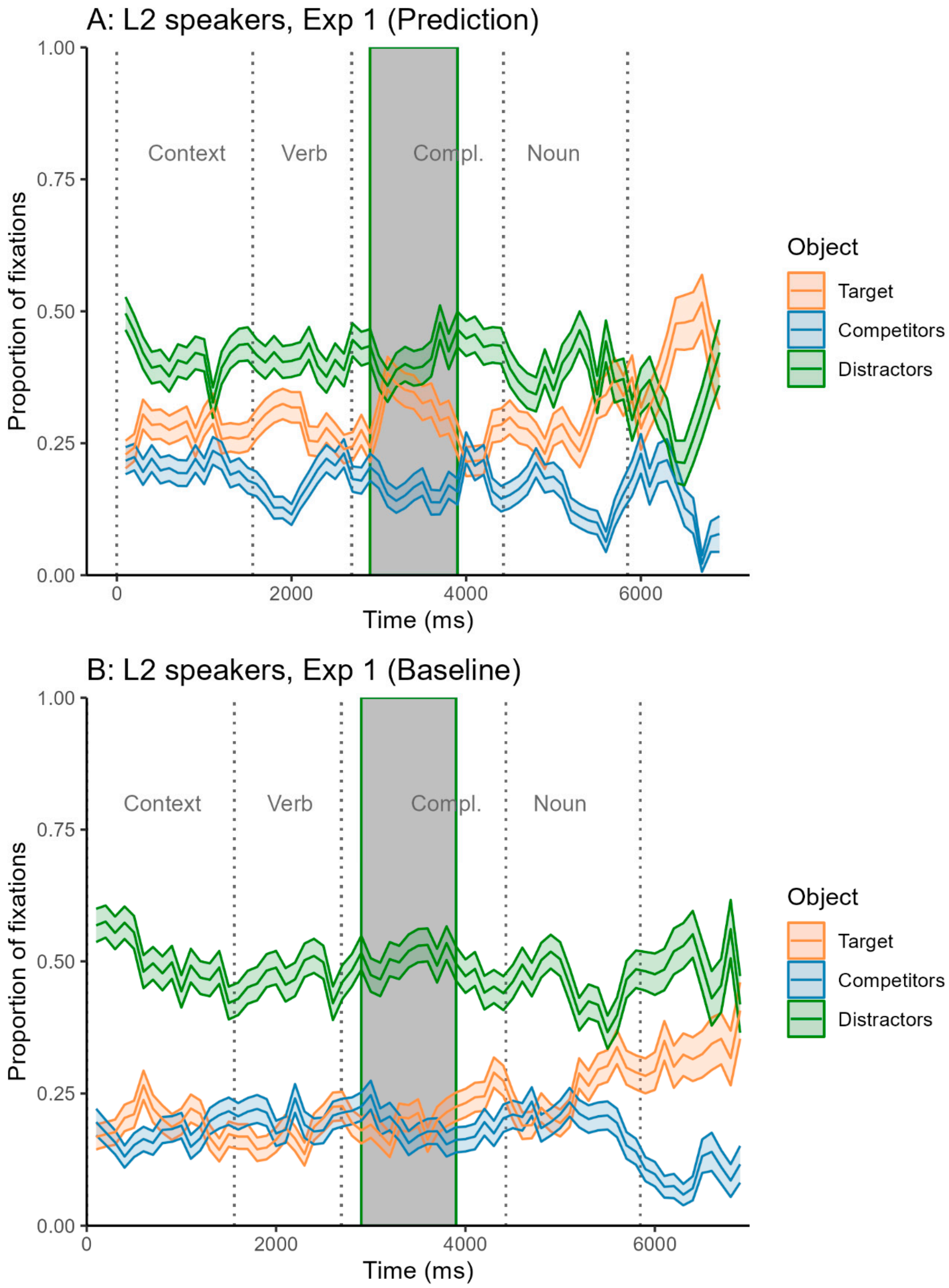

Figure A3. Time course of fixation proportions for target in the prediction (upper half figure) and baseline trial conditions (lower half figure) for L2 speakers in Experiment 1. Ribbons indicate the standard error. Dotted lines indicate the mean onset and offset of word durations in the sentences. The grey-shaded area indicates a significant CPA cluster. Compl.: complement. A competitor shares the target object's number but differs in either adjective (e.g., "the tall boy" vs. "the short boy") or gender (e.g., "the tall boy" vs. "the tall girl"), while a distractor has a different number than the target object (e.g., "the tall boy" vs. "the two/three tall boys"). In prediction trials, the competitor has the same number as the target object but a different gender. In baseline trials, the competitor matches the target object's number but has a different adjective.

Appendix D

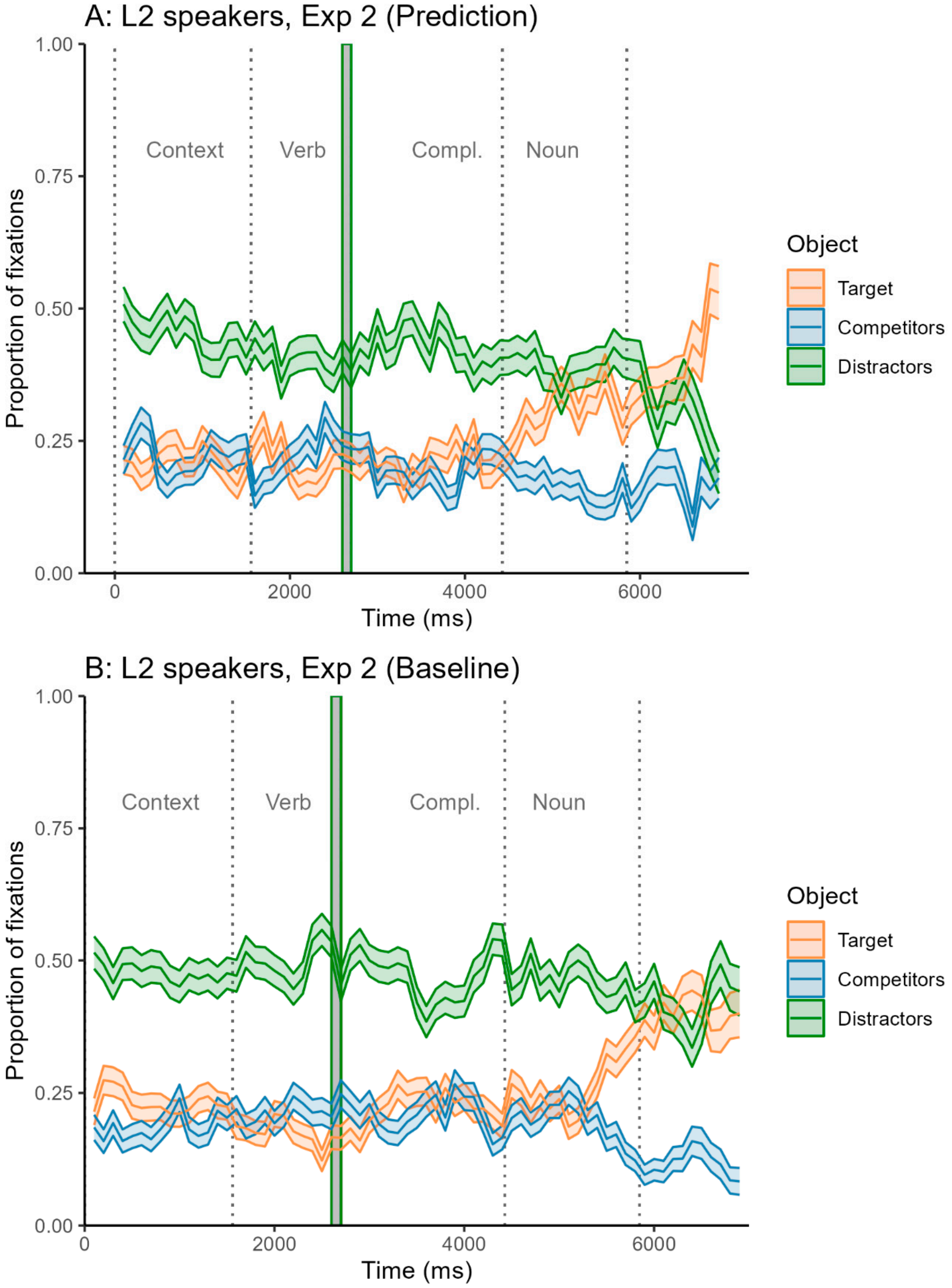

Figure A4. Time course of fixation proportions for target in the prediction (upper half figure) and baseline trial conditions (lower half figure) for L2 speakers in Experiment 2. Ribbons indicate the standard error. Dotted lines indicate the mean onset and offset of word durations in the sentences. The grey-shaded area indicates a significant CPA cluster. Compl.: complement. A competitor shares the target object's number but differs in either adjective (e.g., "the tall boy" vs. "the short boy") or gender (e.g., "the tall boy" vs. "the tall girl"), while a distractor has a different number than the target object (e.g., "the tall boy" vs. "the two/three tall boys"). In prediction trials, the competitor has the same number as the target object but a different gender. In baseline trials, the competitor matches the target object's number but has a different adjective.

Notes

[1] As suggested by a reviewer, we ran a separate LME model to test whether the presence of L1 number marking (L1s which distinguish between singular and plural verbs vs. L1s which lack this distinction) influenced L2 number predictive processing. We found that L2 participants who had L1 verb number morphology did not demonstrate a significant increase in target fixations in prediction trials than in baseline trials in the verb region (Estimate = 0.14, SE = 0.20, z = 0.70, p = 0.479) or the object region (Estimate = 0.42, SE = 0.24, z = 1.77, p = 0.077) relative to L2 participants who lack L1 verb number morphology.

References

Abou-Ghazaleh, A., Khateb, A., & Nevat, M. (2018). Lexical competition between spoken and literary Arabic: A new look into the neural basis of diglossia using fMRI. *Neuroscience*, *393*, 83–96. [CrossRef] [PubMed]

Al-Batal, M., Brustad, K., & Al-Tonsi, A. (2011). *Al-Kitaab fii ta'allum al-'arabiyya: A textbook for beginning Arabic, part one*. Georgetown University Press.

Albirini, A. (2016). *Modern Arabic sociolinguistics: Diglossia, variation, codeswitching, attitudes and identity*. Routledge.

Alemán Bañón, J., & Martin, C. (2021). The role of crosslinguistic differences in second language anticipatory processing: An event-related potentials study. *Neuropsychologia*, *155*. [CrossRef]

Al-Fawzan, A. B. I., Al-Tahir, H. M., & Abdul Khaliq, M. F. M. (2014). *Arabic at your hands* (2nd ed.). Arabic For All.

Alosh, M., & Clark, A. (2021). *Ahlan wa sahlan: Functional modern standard Arabic for beginners*. Yale University Press.

Alshammari, A. R. (2023). Analyzing word order variation and agreement asymmetry in SVO and VSO structures of standard Arabic: Towards a unified account. *Cogent Arts & Humanities*, *10*(2). [CrossRef]

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. [CrossRef] [PubMed]

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609. [CrossRef]

Alzahrani, A. (2023). LexArabic: A receptive vocabulary size test to estimate Arabic proficiency. *Behavior Research Methods*, *56*, 5529–5556. [CrossRef] [PubMed]

Andria, S., Madi-Tarabya, B., & Khateb, A. (2022). Behavioural and electrophysiological analyses of written word processing in spoken and literary Arabic: New insights into the diglossia question. *European Journal of Neuroscience*, *56*(6), 4819–4836. [CrossRef]

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. [CrossRef]

Asadi, I. A., & Asli-Badarneh, A. (2023). The impact of diglossia-effect on reading acquisition among Arabic-speaking children: A longitudinal study. *Journal of Psycholinguistic Research*, *52*, 1919–1937. [CrossRef]

Atawneh, A. (2001). Contrastive study of aspect in English and Arabic. *International Journal of Arabic-English Studies*, *2*. [CrossRef]

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. [CrossRef] [PubMed]

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting-linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. [CrossRef]

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney, & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–76). Cambridge University Press.

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer [Computer program]* (Version 6.3.18). Available online: http://www.praat.org/ (accessed on 1 December 2023).

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*. [CrossRef]

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. [CrossRef]

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, *159*, 50–65. [CrossRef] [PubMed]

Brown, V. A., Fox, N. P., & Strand, J. F. (2022). "Where are the . . . Fixations?": Grammatical number cues guide anticipatory fixations to upcoming referents and reduce lexical competition. *Journal of Experimental Psychology: Learning, Memory, and Cognit*, *48*(5), 643–657. [CrossRef] [PubMed]

Brysbaert, M. (2024). Designing and evaluating tasks to measure individual differences in experimental psychology: A tutorial. *Cognitive Research*, *9*(11). [CrossRef]

Chun, E., Chen, S., Liu, S., & Chan, A. (2021). Influence of syntactic complexity on second language prediction. In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 70–89). John Benjamins. [CrossRef]

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. [CrossRef] [PubMed]

Clahsen, H., & Felser, C. (2017). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. [CrossRef]

Corps, R. E., Liao, M., & Pickering, M. J. (2023). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition*, *26*(1), 231–243. [CrossRef]

Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*, 42–71. [CrossRef]

Dornyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS ONE*, *18*(3), e0279720. [CrossRef]

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer, & M. Haspelmath (Eds.), *WALS Online* (v2020.3). Zenodo. [CrossRef]

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. [CrossRef]

Bosch, J. E., & Foppolo, F. (2022). Predictive processing of grammatical gender in bilingual children: The effect of cross-linguistic incongruency and language dominance. *Lingue e Linguaggio*, *21*(1), 5–27.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290–325. [CrossRef]

Esher, L., Floricic, F., & Maiden, M. (2020). Finite verb morphology in the romance languages. In *Oxford research encyclopedia of linguistics*. Oxford University Press. [CrossRef]

Fakih, A. H. A. (2016). Agreement in standard Arabic VSO and SVO word orders: A feature-based inheritance approach. *Theory and Practice in Language Studies*, *6*(1), 21–33. [CrossRef]

Ferguson, C. (1991). Diglossia revisited. *Southwest Journal of Linguistics*, *10*, 214–234.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469. [CrossRef]

Frenck-Mestre, C., Kim, S. K., Choo, H., Ghio, A., Herschensohn, J., & Koh, S. (2019). Look and listen! The online processing of Korean case by native and non-native speakers. *Language, Cognition and Neuroscience*, *34*(3), 385–404. [CrossRef]

Garrido-Pozu, J. J. (2022). Predictive processing of grammatical gender: Using gender cues to facilitate processing in Spanish. *Lingua*, *278*. [CrossRef]

Garrido Rodriguez, G., Norcliffe, E., Brown, P., Huettig, F., & Levinson, S. C. (2023). Anticipatory processing in a verb-initial mayan language: Eye-tracking evidence during sentence comprehension in tseltal. *Cognitive Science*, *47*(1), e13292. [CrossRef] [PubMed]

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. [CrossRef]

Hartsuiker, R., & Bernolet, S. (2017). Syntactic representations in late learners of a second language: A learning trajectory. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 205–224). John Benjamins. [CrossRef]

Henry, N., Jackson, C. N., & Hopp, H. (2022). Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research*, *38*(3), 397–422. [CrossRef]

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. [CrossRef]

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. [CrossRef]

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. [CrossRef]

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. [CrossRef]

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*. [CrossRef] [PubMed]

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. [CrossRef]

Ito, A., & Pickering, M. J. (2021). Automaticity and prediction in non-native language comprehension. In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 26–46). John Benjamins Publishing. [CrossRef]

Ito, A., Corley, M., & Pickering, M. J. (2018a). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. [CrossRef]

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*, 107291. [CrossRef]

Ito, A., Nguyen, H. T. T., & Knoeferle, P. (2023). German-dominant Vietnamese heritage speakers use semantic constraints of German for anticipation during comprehension in Vietnamese. *Bilingualism: Language and Cognition*, *27*(1), 57–74. [CrossRef]

Ito, A., Pickering, M. J., & Corley, M. (2018b). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. [CrossRef]

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288. [CrossRef] [PubMed]

Kaan, E. (2014). Predictive sentence processing in L2 and L1. *Parsing to Learn*, *4*(2), 257–282. [CrossRef]

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 1–24). John Benjamins. [CrossRef]

Koch, E., 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. [CrossRef]

Koch, E., Bulté, B., Housen, A., & Godfroid, A. (2023). The predictive processing of number information in subregular verb morphology in a first and second language. *Applied Psycholinguistics*, *44*(5), 750–783. [CrossRef]

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. [CrossRef]

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. [CrossRef] [PubMed]

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*. Oxford University Press. [CrossRef]

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. [CrossRef] [PubMed]

Lewis, J. R. (2018). Investigating MOS-X ratings of synthetic and human voices. *Voice Interaction Design*, *2*(1), 22.

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. [CrossRef] [PubMed]

Macwhinney, B. (2012). The logic of the unified model. In A. Mackey, & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). Routledge.

Mahmoodi-Bakhtiari, B. (2018). Morphology. In A. Sedighi, & P. Shabani-Jadidi (Eds.), *The oxford handbook of persian linguistics* (pp. 273–299). Oxford Academic. [CrossRef]

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producer. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. [CrossRef]

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940–967. [CrossRef] [PubMed]

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. [CrossRef] [PubMed]

Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics*, *38*(4), 953–983. [CrossRef]

Mitsugi, S. (2018). Generating predictions based on semantic categories in a second language: A case of numeral classifiers in Japanese. *International Review of Applied Linguistics in Language Teaching*, *58*(3), 323–349. [CrossRef]

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. [CrossRef]

Nassif, L. (2021). Codeswitching between Modern Standard and Colloquial Arabic as L2 sociolinguistic competence: A cross-sectional study from an integrated approach curriculum. *Applied Pragmatics*, *3*(1), 26–50. [CrossRef]

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. [CrossRef]

Palmer, J. (2007). Arabic diglossia: Teaching only the standard variety is a disservice to students. *Journal of Second Language Acquisition and Teaching*, *14*, 111–122.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. [CrossRef] [PubMed]

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. [CrossRef]

Prystauka, Y., Altmann, G. T., & Rothman, J. (2023). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods*, *56*(4), 3504–3522. [CrossRef]

Raish, M. (2017). *The measurement of the complexity, accuracy, and fluency of written Arabic*. Georgetown University.

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. [CrossRef]

Robert, S. (2016). Tense and aspect in the verbal system of Wolof. In Z. Guentcheva (Ed.), *Aspectuality and Temporality: Theoretical and Empirical Issues* (pp. 171–230). John Benjamins.

Ryding, K. (2005). *A reference grammar of modern standard Arabic*. Cambridge University Press.

Sala, M., Vespignani, F., Casalino, L., & Peressotti, F. (2024). I know how you'll say it: Evidence of speaker-specific speech prediction. *Psychonomic Bulletin & Review*, *31*, 2332–2344. [CrossRef]

Saslow, M. G. (1967). Latency of saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033. [CrossRef] [PubMed]

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*, e13335. [CrossRef] [PubMed]

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*, 253–267. [CrossRef]

Schlenter, J., & Felser, C. (2021). Second language prediction ability across different linguistic domains. In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 48–68). John Benjamins Publishing Company.

Schmidt, R. L. (2005). *Urdu: An essential grammar*. Routledge.

Segalowitz, N., & Hulstijn, J. H. (2009). Automaticity in bilingualism and secondlanguage learning. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*, *55*(7), 3786–3804. [CrossRef]

Smolík, F., & Bláhová, V. (2022). Here come the nouns: Czech two-year-olds use verb number endings to predict sentence subjects. *Cognition*, *219*, 104964. [CrossRef] [PubMed]

Soltan, U. (2006). Standard Arabic subject-verb agreement asymmetry revisited in an Agree-based minimalist syntax. In C. Boeckx (Ed.), *Agreement systems* (pp. 239–265). John Benjamins Publishing Company. [CrossRef]

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841. [CrossRef]

Straughn, C. A. (2011). *Evidentiality in Uzbek and Kazakh*. The University of Chicago.

Tagarelli, K. M., Ruiz, S., Vega, J. L. M., & Rebuschat, P. (2016). Variability in second language learning. *Studies in Second Language Acquisition*, *38*(2), 293–316. [CrossRef]

Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, *18*(3), 299–313. [CrossRef]

Voeten, C. C. (2023). *permutes: Permutation tests for time series data* (R package version 2.8). Available online: https://cran.r-project.org/package=permutes (accessed on 25 December 2023).

Younes, M. (2014). *The integrated approach to Arabic instruction*. Routledge.

Zughoul, M. R. (1980). Diglossia in Arabic: Investigating solutions. *Anthropological Linguistics*, *22*(5), 201–217.