Feleke, T. L., & Lohndal, T. (2024). The role of differential cross-linguistic influence and other constraints in predictive L2 gender processing. *Glossa Psycholinguistics*, 3(1): 14, pp. 1–43. DOI: https://doi.org/10.5070/G60111325

The role of differential cross-linguistic influence and other constraints in predictive L2 gender processing

Tekabe Legesse Feleke, UiT The Arctic University of Norway, Department of Language and Culture, AcqVA Aurora Research Center, Tromso, Norway, tekabe.l.feleke@uit.no

Terje Lohndal, Norwegian University of Science and Technology, Department of Language and Literature, Trondheim, Norway; UiT The Arctic University of Norway, Department of Language and Culture, AcqVA Aurora Research Center, Tromso, Norway, terje.lohndal@ntnu.no

Previous studies on the use of morphosyntactic gender cues for linguistic prediction show that non-native speakers' use of grammatical gender information is influenced by various factors. In the present study, we examined the influence of differential cross-linguistic influence (DCLI), knowledge of L2 lexical gender, gender congruency, and L2 fluency. To this end, we investigated L1 Oromo L2 Amharic speakers as well as L1 Amharic speakers, using the Visual World Paradigm (VWP) and supplementary offline experiments. We investigated two groups of L2 Amharic speakers, i.e., L1 Eastern Oromo L2 Amharic and L1 Western Oromo L2 Amharic speakers. The Eastern Oromo dialect patterns with Amharic in terms of gender agreement unlike the Western Oromo dialect which does not have grammatical gender. Analyses of the participants' proportion of eye fixations show that early exposure to the gendered Eastern Oromo dialect facilitates predictive L2 gender processing. L2 fluency, the speakers' knowledge of L2 lexical gender and specific properties of the gender cues modulate predictive L2 gender processing. However, there is no significant influence of lexical gender congruency. The study has ecological significance as it presents empirical data from understudied languages.

1. Introduction

Humans employ prediction for different aspects of their life including language processing. Linguistic prediction is an anticipatory process employed by the human brain to ease the processes of language comprehension (Kamide, 2008; Kuperberg & Jaeger, 2016). This linguistic prediction can be triggered by various anticipatory cues (phonological forms, semantic references, discourse markers and morpho-syntactic features) available in the input. Numerous previous studies reported that monolingual adults deploy these linguistic cues to predict what follows in an utterance (see Altmann & Kamide, 1999; Foucart, 2021; Kaan, 2014; Kamide, 2008; Kim & Grüter, 2021; Koch et al., 2021; Lozano-Argüelles et al., 2020). However, the evidence is mixed when it comes to L2 learners (see Kaan, 2014; Lago et al., 2021; Schlenter, 2023 for reviews). Pertinent to our study, studies indicate that native and L2 learners exploit grammatical gender information differently (see Aumeistere et al., 2022; Grüter et al., 2012; Grüter & Kaan, 2021; Hopp, 2013, 2016; Hopp & Lemmerth, 2018). Most studies reported either quantitative or qualitative difference between L1 and L2 predictive gender processing (see Kaan, 2014; Lago et al., 2021; Schlenter, 2023). This difference is usually attributed to constraints such as cross-linguistic influence, L2 fluency, knowledge of L2 lexical gender, cue availability, and other individual-level factors (Hawkins, 2009; Jakubowicz & Roulet, 2004; Kaan, 2014; Lago et al., 2021; Schlenter, 2023).

Regardless of a recent growing interest in L2 gender processing, the interplay between L2 predictive gender processing and these constraints has not been adequately investigated (see Bordag & Pechmann, 2007; Curcic et al., 2019; Hopp, 2022; Kaan, 2014; Lago et al., 2021). In the existing studies, these factors were treated with varied degree of emphasis. A handful of studies reported the influence of the L1 gender system on predictive L2 gender processing (e.g., Foucart & Frenck-Mestre, 2011; Lemmerth & Hopp, 2019; Paolieri et al., 2010; Sabourin & Stowe, 2008). Relatively, the influence of the knowledge of L2 lexical gender on predictive L2 gender processing has received more attention (see Clahsen & Felser, 2006; Grüter et al., 2012; Hopp, 2010, 2013, 2016, 2018; Prévost & White, 2000). Hence, most processing-based models of L2 predictive gender processing relied on individual differences associated with knowledge of lexical gender (see Clashen & Felser, 2006). The relationship between L2 fluency and predictive L2 gender processing is the most intricate one. Some studies reported native-like prediction competency of speakers with advanced L2 fluency (e.g., Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018). Others failed to replicate the effect of L2 fluency (e.g., Dijkgraaf et al., 2017; Hopp, 2015; Kim & Grüter, 2021; Mitsugi, 2020). The observed inconsistency is partly a reflection of the diverse background of the target L2 learners. Most previous studies investigated L2 learners that differ in terms of L2 fluency, exposure, and several other dimensions (Andersson et al., 2019; Dyson, 2016).

In the present study we investigate the extent to which predictive L2 gender processing is moderated by differential cross-linguistic influence, knowledge of L2 lexical gender, and L2 fluency. Deviating from the traditional L1 *vs*. L2 comparisons, we explore differential cross-linguistic influence of dialects of the first language on predictive L2 gender processing. Operationally, we conceptualize 'differential cross-linguistic influence' as effects of different magnitude that are imposed by different L1 dialects. To this end, we investigate bidialectal Oromo speakers whose second language is Amharic. Oromo is a Lowland East Cushitic language (Clamons, 1992, 1993; Feleke & Lohndal, 2023; Yimam, 1988) whereas Amharic is a Semitic language (Feleke et al., 2020; Feleke, 2021, 2023; Hetzron, 1972). Both languages belong to the Afro-asiatic language family. Oromo and Amharic are spoken in Ethiopia, East Africa. Amharic is the lingua franca of the country (Meyer, 2006). Oromo is spoken in the Oromia Regional State, one of the states in Ethiopia. L1 Oromo speakers are exposed to Amharic starting from elementary school since Amharic is taught as a second language in schools across the country.

We investigate two groups of late L2 Amharic speakers: L1 Western Oromo L2 Amharic speakers and L1 Eastern Oromo L2 Amharic speakers. These two Oromo dialects mark gender differently. The Eastern dialect has a contrasting masculine and feminine gender, and gender is realized via exponents on various elements such as determiners, verbs, and adjectives. The Western Oromo dialect does not have grammatical gender (Clamons, 1992, 1993; Feleke & Lohndal, 2023; Feleke, 2024a). Like the Eastern Oromo dialect, Amharic has a contrasting masculine-feminine gender (see Kramer, 2015). We examine whether the grammatical gender similarity between the Eastern Oromo dialect and Amharic facilitates predictive Amharic gender processing. Furthermore, we explore the extent to which the facilitation effect is constrained by the target factors. Specifically, our study aims at addressing the following three objectives; (a) to determine the magnitude of influence of the Oromo dialects on processing Amharic gender predictively; (b) to explore the role of the knowledge of L2 lexical gender in augmenting predictive L2 gender processing; finally, (c) to investigate the impact of L2 fluency and gender congruency on predictive L2 gender processing. Since the target Oromo dialect speakers share the same culture, education background and exposure to the second language, we effectively controlled for the discrepancies in participants' background, which was not the case in many previous studies.

We employ webcam-based eye tracking as the main data gathering tool, combined with supplementary offline experiments. Series of analyses on the participants' proportion of eye fixations show that L2 Amharic speakers can use grammatical gender predictively. Early exposure to the gendered Eastern Oromo dialect facilitates predictive L2 gender processing. Moreover, L2 fluency and L1 speakers' knowledge of L2 lexical gender modulate predictive L2 gender processing. Lexical gender congruency does not have a significant impact on predictive L2 gender processing. The results have ecological and theoretical implications. The arguments that favor predictive L2 gender processing usually come from studies conducted on well-studied languages of the Western world, and they bias towards these languages (see Blasi et al., 2022; Kutlu & Hayes-Harb, 2023; Niemi & Laine, 1989). In this regard, our study contributes empirical data from languages that have not appeared in the psycholinguistics literature and plays a crucial role in addressing the ecological imbalance. The study also informs the debate regarding the representational and processing-based nature of predictive L2 gender processing. Some previous studies argued that the ability to process L2 gender predictively is acquired only if gender is instantiated in the first language (see Grüter & Rohde, 2013; Hawkins, 2009). Others rejected the issue of representation and linked the L2 processing difficulty to individual-level factors (e.g., Clashen & Felser, 2006; Grüter et al., 2012; Hopp, 2018; Prévost & White, 2000). Moreover, our study reflects on the interaction between the morphosyntax of first and second languages by systematically exploiting the grammatical gender similarity between the L1 and L2.

Crucially, our study deviates from the traditional L1 vs. L2 dichotomy and investigates the potential impact of dialects of the same L1 on predictive L2 gender processing. In other words, it sheds light on the link between L2 predictive gender processing and the dialectal variation in the first language. Previous studies indicate that changes or variations in the dialects of a first language can affect the speakers' prediction behavior (see Lundquist et al., 2016; Lundquist & Vangsnes, 2018). In most cases, there are rich lexical and morpho-syntactic variations in what is generally described as 'first language' (Feleke, 2024b). Studies that previously investigated the interaction between native and nonnative grammars during predictive L2 processing have not examined these variations.

This paper is structured as follows. Follow the above brief introduction, in Section 2 we introduce the theoretical assumptions associated with predictive L2 gender processing and the gender systems of the target languages–Oromo and Amharic. Section 3 deals with the research questions and predictions. In Section 4, we briefly describe the methods used. In Section 5, we present the results, and then the discussion of the results follows in Section 6. Finally, in Section 7, we present conclusions and theoretical implications of our findings.

2. Theoretical background and the language context

2.1 Predictive L2 gender processing

Previous studies offer various definitions of *prediction*, which often differ from study to study. The present study adopts a restrictive view of prediction that has been commonly used by studies previously conducted on sentence processing (see Kim & Grüter, 2021; Kuperberg & Jaeger, 2016). In these studies, prediction is defined as a pre-activation of the upcoming nouns. This means that only effects observed before the onset of the critical words, in our case the target nouns, provide empirical evidence of predictive processing. Based on this assumption, previous eye tracking studies reported that language users can move their eyes to a particular object in the display even before the object is mentioned. Their relative proportion of looks to the objects is then interpreted as an index of how listeners predictively resolve reference, based on partial input as a sentence is unfolding (see Huettig et al., 2011). There is plenty of evidence that shows language users engaging in proactive, forward-looking processing during sentence comprehension (see Dahan et al., 2000; Foucart et al., 2014; Grüter et al., 2012; Hopp, 2016; Hopp & Lemmerth, 2018; Kaan et al., 2010; Lew-Williams & Fernald, 2010). As a result, prediction in language processing has been investigated in a wide range of linguistic domains: semantics (e.g., Casillas & Frank, 2013; Grüter et al., 2020; Kamide et al., 2003; Weber et al., 2016;), morpho-syntax (e.g., Hopp, 2016; Lemmerth & Hopp, 2019; Paorlieri et al., 2019) and discourse (e.g., Boudewyn et al., 2015; Kim & Grüter, 2021; Scheutz & Eberhard, 2004).

Previous studies show that native speakers of gendered languages use gender information on adjectives and varieties of determiners to anticipate upcoming nouns (e.g., Dussias et al., 2013; Halberstadt et al., 2018; Hopp, 2013, 2016; Hopp & Lemmerth, 2018; Paorlieri et al., 2010). It has been argued that the capacity to process gender predictively starts early in life for native speakers of gendered languages. Some previous studies also argued that simultaneous bilinguals are effective predictors, like monolingual speakers (Foucart et al., 2014; Grüter et al., 2012; Hopp, 2016; Kaan et al., 2010; Lew-Williams & Fernald, 2010) implying that early exposure to the second language scaffolds the mastery of predictive L2 gender processing. However, there have been inconsistent reports when it comes to the predictive behavior of late L2 learners. The earlier assumption was that late L2 learners have a *Reduced Ability to Generate Expectation – RAGE* (see Grüter & Rohde, 2013; Hawkins & Casillas, 2008). Predictive processing was conceived of as a capacity that is conditioned by the instantiation of the native language. According to this hypothesis, even advanced L2 learners may not be able to predict since they do not reach *native-like* automatization that is compulsory to predict. This argument has recently been challenged as the increasing number of studies reported predictive processing by both native and late L2 learners (see Bañón & Martin, 2021; Dussias et al., 2013; Fourcart et al., 2014; Hopp, 2013; Kaan, 2014; Kim & Grüter, 2021). These later studies contended that the same factors that influence L1 predictive processing also affect L2 processing. In other words, the competence to use prediction during L2 gender processing is constrained by the same individual-level factors (Bordag & Pechmann, 2007; Kim & Grüter, 2021; Schlenter, 2023). Therefore, L1-L2 difference in predictive behavior is attributable to modulating factors such as cross-linguistic influence, L2 fluency, memory load, and cue availability and reliability.

The discords associated with L2 predictive gender processing can be subsumed into two accounts of gender processing: processing-based and representational. The processing-based account associates the L2 predictive gender processing difficulty to the variability in the L2 lexical access and to various individual and language-level factors. It holds the view that L2 grammars do not differ from native grammars in terms of representation and that the variability in processing L2 morphosyntax predictively follows from difficulties in mapping the target morphophonological forms to syntactic features in situations of real-time processing pressure (see Prévost & White, 2000). Related to this, the Weaker Link Hypothesis (see Gollan et al., 2008) assumes that individual difference in gender assignment obscures the possibility of using gender for prediction since non-target lexical gender leads to a prediction error, implying that lexical gender and grammatical aspects of gender have a causal relationship. Similarly, the Non-selective Lexical Access Hypothesis (e.g., Salamoura & Williams, 2007) assumes that gender representations in the bilingual mental lexicon are interrelated or accessed non-selectively. Hence, L2 learners whose L1 marks grammatical gender typically show more target-like gender agreement processing than L2 learners whose L1s lack gender. According to this hypothesis, L2 learners whose L1s realize gender agreement may not be invariably *native-like* in processing because target predictive gender processing can be limited to certain gender markings. Similarly, the Missing Surface Inflection Hypothesis (Prévost & White, 2000) predicts that L2 inflectional variability stems from failures in retrieving inflectional forms under real-time processing pressure. L2 learners usually rely on default forms of gender because they fall short of mapping the appropriate target lexical form into the given grammatical context. The Shallow Structure Hypothesis (Clashen & Felser, 2006), Lexical Bottleneck Hypothesis (Hopp, 2018) and Lexical Gender Leaning Hypothesis (Grüter et al., 2012) make similar predictions. Put together, these models show that a strong lexical gender representation is a prerequisite for successful predictive processing of L2 gender agreement.

The representational accounts (Hawkins, 2009; Howard, 2011; Jakubowicz & Roulet, 2004) assume that there is an inherent difference between L1 and L2 learners that emerges from input type and learning strategies. For instance, during L1 acquisition, nouns are acquired in continuous speech, together with other constituents such as determiners and adjectives. However, during L2 learning, nouns are taught separately in a written form, in a way that disrupts the smooth association between nouns and other elements. Therefore, according to this account, L2 learners whose L1s do not encode grammatical gender do not project syntactic gender features for agreement and hence need to rely on non-grammatical cues for computing gender concord in the L2. For L2 learners, the predictive effects of gender marking could be limited to some memorized exceptions of listed noun-gender associations, because feature-based syntactic agreement between determiners and nouns is not available. This account also appeals to age-related grammatical impairments arguing that L2 grammars do not encode abstract syntactic gender features necessary for computing agreement relations unless these features are instantiated in the L1 (see Hawkins, 2009; Hawkins & Casillas, 2008).

2.2 Cross-linguistic influences and other constraints

Predictive L2 gender processing can be influenced by various factors. For example, it was argued that the availability of a higher number of gender cues in the input aids the use of gender for prediction (Fowler & Jackson, 2017; Garrido-Pozú, 2022). It was also indicated that training of gender assignment boosts the use of prediction for L2 gender processing (see Hopp, 2016). Mode of input presentation was also stated as one of the contributing factors. For instance, in the study conducted by Hopp & Lemmerth (2018) on cross-linguistic lexical and syntactic influence in grammatical gender, the elicited production task showed that all bilingual children assign target gender to nouns, irrespective of whether the nouns belong to the same or different gender class in German and Russian. In the visual world paradigm, however, only simultaneous bilinguals could predict. Simultaneous bilingual children showed predictive gender processing, irrespective of the gender (in)congruency between L1 and L2. In contrast, the successive bilingual children showed predictive gender processing only for lexically congruent nouns. Based on these findings, the study argued that the asynchronous acquisition of L2 in successive bilinguals implicates that L2 gender is first accessed through the L1 lexicon.

Moreover, studies reported that syntactic gender similarity between the L1 and L2 enhances predictive L2 gender processing (see Bordag & Pechmann, 2007; Bosch & Foppolo, 2022; Dussias et al., 2013; Morales et al., 2016). Conversely, a mismatch in gender agreement between the L1 and L2 poses difficulty for processing L2 gender agreement presumably since predictive gender processing requires easy and rapid access to grammatical gender knowledge (Dussias et al., 2013; Hopp, 2013; Lew-Williams & Fernald, 2007, 2010). For example, in successive experiments, Foucart and Frenck-Mestre (2011) investigated the effect of similarity between the first and the second language on grammatical gender processing in L2, focusing on German advanced learners of French and French native controls. The study explored gender agreement violations between determiner and noun, postposed adjective and noun, and preposed adjective and noun. The ERP results showed a similar P600 effect for native and non-native speakers for agreement violations when agreement rules were similar in the L1 and L2 whereas no effect was found for L2 learners when agreement rules varied across languages. The study concluded that syntactic processing in the L2 is affected by the similarity of syntactic rules in the L1 and L2. Recently, Bosch and Foppolo (2022) also investigated linguistic prediction based on grammatical gender in Italian-German bilingual children, using a visual world eye tracking paradigm. Children listened to sentences while looking at objects that either matched or mismatched in grammatical gender, and that varied with respect to cross-linguistic gender congruency. The study reported rapid predictive processing, as the children anticipated nouns based on the grammatical gender of the determiners. Furthermore, the children exhibited a gender congruency effect, i.e., they showed negative cross-linguistic influence when the grammatical gender of the two languages did not overlap, leading to delayed anticipation.

There is also evidence that, like syntactic overlap, lexical gender similarity between L1 and L2 plays a crucial role in L2 predictive gender processing. For example, Paolieri et al. (2020) investigated whether processing a word in one language is affected by the gender of its translation equivalent in another language. A group of Catalan–Spanish bilinguals performed a translation-recognition task while event-related potentials (ERPs) were recorded. The ERP data showed a reduced N400 for the congruent condition, implying facilitation between the lexical gender of the first and second languages. Moreover, Morales et al. (2016) investigated gender-congruency effects using the visual world paradigm. The study reported a reduced eye gaze latency in the congruent condition. Paolieri et al. (2019) also investigated the role of lexical gender in processing gender information. In the study, Russian–Spanish bilinguals showed gender congruency effects when they translate concrete nouns in isolation or in noun-phrases. Several other studies also reported lexical gender congruency effects (see Bobb et al., 2015; Dussias et al., 2013; Hopp, 2016; Hopp & Lemmerth, 2018).

The ability to predict during L2 gender processing can also be constrained by L2 fluency (see Blumenfeld et al., 2016; Hopp, 2016; Hopp & Lemmerth, 2018). However, the relationship between L2 fluency and L2 predictive gender processing is the complex one. Several studies previously reported greater or native-like prediction competency of speakers with advanced L2 fluency (see Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018). Equally, a substantial number of studies failed to replicate the L2 fluency effect (see Dijkgraaf et al., 2017; Hopp, 2015; Kim & Grüter, 2021; Mitsugi, 2020). Some other studies indicated that advanced L2 learners are different from both native and intermediate L2 learners in predictive use of gender cues (Dijkgraaf et al., 2019; Kaan et al., 2016). These show that the relationship between L2 fluency and L2 predictive gender processing is an area that awaits further investigation. The problem lies partly in measuring the fluency *per se*. Previous studies that examined the link between L2 fluency and L2 predictive gender processing used fluency tests that varied from perception test, receptive vocabulary to more standardized tests (see Segalowitz, 2016 and Tavakoli, 2016 regarding the difficulty in measuring L2 fluency).

Gender assignment errors is another factor that increases difficulties in computing gender agreement. If the target gender assignment has not been established, it usually leads to prediction errors. In turn, this affects the reliability of grammatical gender cues and hinders the use of prediction during L2 gender processing. Related to this, Hopp (2012) explored lexical and syntactic aspects of gender processing in L2 production and comprehension in advanced English learners of German and German native speakers through a picture naming task and a visual world paradigm. Results showed a strong correlation between variability in gender assignment during production and variability in using gender as a predictive cue in comprehension. Only L2 learners that had target-like performance in gender assignment in the production task were able to use gender as a predictive cue in comprehension. This finding shows that the variability in using gender predictively is more likely related to lower levels of activation and access to gender nodes rather than deficits of mental representation in the L2 grammar. Predictive L2 gender processing may also be influenced by non-linguistic factors such as motivation (Luna & Peracchio,

2002) and memory load (Gordon et al., 2001; Ito et al., 2018). To the best of our knowledge, not many studies have investigated the roles these factors play in L2 predictive gender processing.

2.3 Gender in Oromo and Amharic

Gender is often defined as classes of nouns that are reflected on other associated words (Hockett, 1958). The link between nouns and the associated words is expressed via gender agreement. Languages differ in the way they mark gender (see Corbett, 1991). Different dialects can also have different gender systems as is the case in Oromo. The Oromo gender system varies across dialects (see Clamons, 1992, 1993; Feleke & Lohndal, 2023; Feleke, 2024a). The Eastern dialect has a phonology-based gender assignment pattern, i.e., nouns that end in consonants and in the low central vowel /a/[1] are masculine whereas nouns that have other vowel /e, i, o, u/ endings are feminine (cf: (1a, c, e) and (1g)). In the dialect, the vowel endings serve as declension class markers (see Feleke, 2023; Feleke & Lohndal, 2023; Feleke, 2024a). Conversely, the Western dialect does not have grammatical gender (see (1b), (1d), (1f) & (1h)) since the feminine gender has been entirely neutralized (see Clamons, 1992, 1993; Feleke & Lohndal, 2023; Feleke, 2024a). Therefore, in the Eastern dialect, nouns agree in gender with elements such as adjectives, verbs, and determiners. In the Western dialect, however, this agreement does not exist due to the neutralized feminine gender.

(1) a. Ablee tam ǧaala-tt-a? *(Eastern)*
knife *which.F* like-2.SG-IPFV
'Which knife do you like?'

b. Albee[2] kam ǧaala-tt-a? *(Western)*
knife which.M like-2.SG-IPFV
'Which knife do you like?'

c. Iddoo-n tun tiyya. *(Eastern)*
place-NOM this.F mine.F
'This place is mine.'

d. Eddoo-n kun kiyya. *(Western)*
place-NOM this.M mine.M
'This place is mine.'

[1] We presented only short vowels here, but the same explanation applies to the long central vowel /aa/ and to the non-central vowels: /ee/, /ii/, /oo/ and /uu/.

[2] 'knife' is *ablee* in the Eastern dialect, but *albee* in the Western dialect. 'Place' is *iddoo* in the Eastern dialect, but *eddoo* in the Western dialect.

e. Furtuun-n gudd-oo ɗa. *(Eastern)*
key-NOM big-F COP
'The key is big.'

f. Furtuun-n gudd-aa ɗa. *(Western)*
key-NOM large-M COP
'The key is big.'

g. Gurbaa-n ɗuf-ø-e. (*Eastern)*
boy-NOM come-3.M.SG-PFV
'The boy came.'

h. Gurbaa-n ɗuf-ø-e. (*Western)*
boy-NOM come-3.M.SG-PFV
'The boy came.'

Amharic has a default masculine gender (see Kramer, 2015). Most Amharic nouns are masculine in gender, but Amharic has maintained the masculine-feminine gender distinction. Hence, the Eastern Oromo dialect patterns with Amharic in terms of gender agreement. For example, in both the Eastern dialect and Amharic, nouns agree in gender with verbs ((2c), (2d), (2g) & (2h)), adjectives ((2c) & (2d)) and determiners ((2a), (2b), (2e) & (2f)).

(2) a. Mana kamii guddaa ɗa? *(Eastern)*
house which.M big.M. COP
'Which house is big?'

b. Yətɨǧǧa-w bet tɨllɨk' nəw? *(Amharic)*
which-M house big COP.M
'Which house is big?'

c. Simbirroo diim-tuu-n barart-t-e. *(Eastern)*
bird red-F-FOC fly-3.F.SG-PFV
'The red bird flew.'

d. K'əyy-w-a wəf bərrər-ečč. *(Amharic*)
red-DEF-F bird fly.PFV-3.F.SG
'The red bird flew.'

e. Ǧiʔa kana nin-ǧaala ɗa. *(Eastern)*
moon this.M FOC-like COP
'I like this moon.'

f. Yɨh-ičči-n c'ərək'a ɨ-wədd-allə-hu. *(Amharic)*
this-F-FOC moon 1.SG-like-AUX-1.SG
'I like this moon.'

g. Daawwitti-n c'abt'-ø-e. *(Eastern)*
mirror-FOC break-3.F.SG-PFV
'The mirror is broken.'

h. Məstwət-u tə-səbbər-ə. *(Amharic)*
mirror-DEF.M PASS-break.PFV-3.M.SG
'The mirror is broken.'

In Amharic, only definite nouns agree in gender with attributive adjectives, i.e., there is no gender agreement between indefinite Amharic nouns and attributive adjectives. Predicative adjectives never agree in gender with nouns in Amharic (see (2b)). Moreover, when there is a sequence of adjectives, only the first adjective necessarily agrees in gender with noun in Amharic; the agreement between the noun and the rest of the adjectives is optional (see Kramer, 2015). In the Eastern Oromo dialect, adjectives always agree in gender with the noun[3]. In the dialect, only the proximal demonstrative agrees in gender with the noun, and the distal demonstrative never agrees in gender with the noun. In Amharic, however, both distal and proximal demonstratives agree in gender with the noun. Besides, Amharic adjectives are strictly prenominal ***tɨllɨk*** *bet* 'a big house'. Interrogative pronouns have a flexible position ***yətu*** *bet new*? or *bet* ***yətu*** *new*? 'Which is a house?'. Demonstrative pronouns also have a flexible position in Amharic: ***yɨh*** *bet new* or *bet* ***yɨh*** *new* 'This is a house', but the postnominal position is used in very rare contexts, for example when a special emphasis is placed on the noun.

Table 1 presents examples of the three gender agreement domains that we investigate: noun-adjective, noun-interrogative pronoun, and noun-demonstrative pronoun gender agreement. Examples (3a) and (3b) show that, in the Eastern Oromo dialect, the feminine noun *abaaboo* and the masculine noun *mana* agree in gender with the adjective – *guddoo* and *guddaa* respectively. In the Eastern dialect, there are also distinct feminine *tamii* and masculine *kamii* interrogative pronouns. Likewise, there are distinct masculine and feminine proximal demonstrative pronouns – *tun* and *kun* respectively. However, in the Western Oromo dialect ((3c) & (3d)), only the masculine adjective *guddaa* is acceptable, and the feminine counterpart *guddoo* does not exist. Moreover, in the Western Oromo dialect, there is only one interrogative pronoun, *kami*, and one proximal demonstrative pronoun *kun* since the feminine counterparts have been neutralized. Amharic gender agreement (see (3e) & (3f)) patterns with the gender agreement of the Eastern Oromo dialect. Hence, in Amharic, the feminine noun *s'iggerəda* and masculine noun *bet* agree in gender with the adjective – *tɨllɨk'wa*, and *tɨllɨk'u* respectively. Moreover, Amharic has distinct feminine *yətwa* and masculine *yətu* interrogative pronouns. Likewise, it has distinct masculine *yɨh and* feminine *yɨčči* proximal demonstrative pronouns.

[3] There are a few exceptions, though. For example, adjectives such as *adii* 'white', *gaarii* 'generous' and *kelloo* 'yellow' do not agree in gender with nouns.

Table 1: The gender agreement domains under investigation.

	Gen.	N-ADJ	N-INT	N-DEM	Varieties
(3a)	F	abaaboo guddoo rose big.F 'A big rose'	tamii bašoo ɗa? which.F cat COP 'Which is a cat?'	tun haamtuu ɗa this.F sickle COP 'This is a sickle.'	*Eastern*
(3b)	M	mana guddaa house big.M 'Big house'	kamii farda ɗa? which.M horse COP 'Which is a horse?'	kun rasaasa Ø this.M gun COP 'This is a gun.'	
(3c)	M	abaaboo guddaa rose big.M 'A big rose'	kamii adurree[4] ɗa? which.M cat COP 'Which is a cat?'	kun haamtuu ɗa this.M sickle COP 'This is a sickle'	*Western*
(3d)	M	mana guddaa house big.M 'Big house'	kamii farad ɗa? which.M horse COP 'Which is a horse?'	kun rasaasa Ø this.M gun COP 'This is a gun.'	
(3e)	F	tɨllɨk'-wa s'iggereda big-DEF.F rose 'The big rose'	dɨmmət yət-wa nat? cat which-DEF.F COP 'Which is a cat?'	yɨčči maćɨd nat which.F sickle COP.3.F.SG 'This is a sickle.'	*Amharic*
(3f)	M	tɨllɨk-u bet tall-DEF.M boy 'The big house'	fərəs yet-u nəw? horse which-DEF.M COP 'Which is a horse?'	yɨh t'əmənǧa nəw this.M gun COP.3.M.SG 'This is a gun.'	

[4] There is lexical variation between the two Oromo dialects. For example, 'cat' is *bašoo* in the Eastern dialect, but *adurree* in the Western dialect.

From a sociolinguistic standpoint, Oromo is spoken in the Oromia Regional State, and it serves as the working language of the region. Furthermore, Oromo serves as a medium of instruction in the schools across the region. The Western dialect is the dominant school lect, but there is a high degree of flexibility regarding dialect use in the schools; for example, the schoolteachers sometimes mix their local dialect during classroom instruction (see Feleke, 2024a). In general terms, most Oromo communities except the speakers of the Western dialect, are bidialectal speakers since they use a different dialect in the school and other contexts. As a native lect, the Eastern dialect is spoken in the eastern part of the Oromia Regional State whereas the Western dialect is spoken in the western part of the Region. Native speakers of the Eastern dialect are exposed to the Western dialect via schooling (see Feleke & Lohndal, 2023; Feleke, 2024a). However, native speakers of the Western dialect do not have exposure to the Eastern dialect. Amharic is the working language of the Federal Government of Ethiopia. It is spoken as a first language in the Amhara Regional State and in most cities and towns across the country (Meyer, 2006). Amharic is also spoken as a second language almost all over the country, and it is taught as a second language in the schools in Ethiopia. The grade level in which Amharic is introduced varies from state to state. In the Oromia region, which is the target of the present study, Amharic is introduced as a subject at grade five[5]. This means that L1 Western L2 Amharic speakers must acquire gender marking in schools after the exposure to Amharic since their native Western dialect does not encode grammatical gender. This is not the case for the speakers of the Eastern Oromo dialect because they have prior exposure to a gender system that is compatible with the gender system of Amharic.

3. Research questions and predictions

In this study, we aim to answer three questions. The first question (a) is about the anticipatory behavior of the first and second language speakers, and whether the anticipatory behavior is affected by differential cross-linguistic influence. The last two questions (b & c) are related to the influence of individual-level factors on predictive L2 gender processing. Each question is explained in what follows, along with associated predictions.

a. Is there differential cross-linguistic influence from a previously acquired dialect?

In connection with this question, we test the two assumptions held in previous studies. First, we examine the supposition that previously acquired languages influence predictive L2 processing. Based on Dussias et al. (2013), Hopp (2013), Lew-Williams and Fernald (2007) and others, we predict facilitation from the Eastern Oromo dialect due to the gender agreement similarity between Amharic and the Eastern Oromo dialect. In other words, we anticipate a prolonged

[5] This is true for all public schools, but in some private schools Amharic is introduced at grade one.

early eye fixation among the L1 Eastern dialect speakers, but a reduced late eye fixation among the L1 Western dialect speakers because of the lack of gender agreement in the Western Oromo dialect. Furthermore, following Dussias et al. (2013), Hopp (2016), Hopp and Lemmerth (2018), Morales et al. (2016) and Paolieri et al. (2019, 2020), we predict an early eye fixation in the gender congruent condition than in the gender incongruent condition. Second, we tap into the representation of L2 grammatical gender. If we do not find any evidence of prediction among the bilingual groups, this suggests that L2 learners have a representational deficit associated with L2 predictive processing as argued by Hawkins (2009), Howard (2011), and Jakubowicz & Roulet (2004). However, if there is evidence of prediction among the L2 groups, this suggests that the processing-based models are on the right track as previously argued by Grüter et al. (2012), Hopp (2018), Prévost and White (2000), and Salamoura and Williams (2007).

b. Is there a link between L2 fluency and predictive L2 processing?

Although there are inconsistencies in the literature regarding the role of L2 fluency in predictive L2 processing (see Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018), studies that recognize a positive relationship between L2 fluency and L2 predictive processing seem to be higher in number. Hence, we predict a strong and positive association between L2 fluency and predictive L2 gender processing.

c. Is there a link between knowledge of L2 lexical gender and predictive L2 processing?

Finally, following Hopp (2012, 2013, 2015, 2016), we predict a strong association between the participants' knowledge of L2 lexical gender and predictive L2 processing. We predict that L2 learners that have advanced knowledge of L2 lexical gender could proactively fixate at the target noun more frequently than those that do not have adequate knowledge of L2 lexical gender. **Table 2** presents the summary of our predictions.

Table 2: Summary of the research questions and predictions.

No	Research questions	Predictions
1	Is there differential cross-linguistic influence from a previously acquired dialect?	• Facilitation from the Eastern Oromo dialect • Predictive processing by both L2 Amharic groups (*processing-based*) or else prediction only by the L1 Amharic and L1 Eastern dialect speakers (*representation-based*)
2	Is there an association between L2 fluency and predictive L2 processing?	• A strong positive correlation between L2 fluency and predictive L2 gender processing
3	Is there a link between knowledge of L2 lexical gender and predictive L2 processing?	• A strong association between knowledge of lexical gender and predictive L2 processing

4. Methods

We measured L2 fluency, knowledge of L2 lexical gender, and the proportion of eye fixations of the three target groups – L1 Amharic, L1 Eastern L2 Amharic, and L1 Western L2 Amharic speakers. A self-reported perceptive fluency test was used as a proxy for the L2 fluency measure (see Derwing et al., 2006). The test was incorporated into a background questionnaire which also consisted of other items related to the participants' language history, demographic information, and language preference. Regarding the L2 fluency test items, the participants were asked to rate their perceived degree of Amharic fluency on a Likert scale that ranged from 0 to 10 where 0 stands for 'I do not speak Amharic' and 10 stands for 'I am a fluent speaker of Amharic'. They were also asked to rate their Amharic fluency based on the four Amharic skills (listening, speaking, reading, and writing) using the same scale (0–10). The mean score was used as a measure of L2 fluency. The L2 fluency test items were presented using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). The participants did the questionnaire prior to the eye tracking experiment. The visual world paradigm and measure of the participants' knowledge of L2 lexical gender are discussed in Sections 4.1 and 4.2, respectively.

4.1 The Visual World Paradigm (VWP)

We used the webcam-based visual world paradigm as an alternative to infrared eye tracking (see Bott et al., 2017; Prystauka et al., 2023; Vos et al., 2022; Yang & Krajbich, 2021). Using Gorilla Experiment Builder, the participants were presented with a visual display showing two pictures, either in the *same* or in the *different* gender condition. In the same gender condition, the target and competitor pictures have the same gender and in the different gender condition, the target and the competitor pictures have different gender. Each target-competitor pair was presented along with an audio instruction. The participants selected the target picture based on the instruction. Meanwhile, their eye movements were recorded. The task was administered in a temporary laboratory that we established at Haramaya University.

4.1.1 The participants

We tested one hundred eighty (180) participants: 61 L1 Eastern L2 Amharic speakers; 58 L1 Western L2 Amharic speakers, and 61 L1 Amharic speakers. They are 1st and 2nd year students at Haramaya University. The L2 Amharic groups are born and grow up in the areas where either the Eastern or the Western Oromo dialect is spoken as a native dialect. The L1 Amharic speakers are from Addis Ababa, the capital, and they are native speakers of standard Amharic. They do not have active exposure to any other language except to English which is the medium of instruction across secondary schools and universities in the country. The target L2 speakers are recruited based on four inclusion criteria: (1) being a late L2 learner of Amharic (acquired Amharic in schools); (2) being able to freely communicate in Amharic; (3) born and grown up either in the eastern or the western part of the Oromia Region of Ethiopia, and (4) scored 15 or above on a picture naming receptive vocabulary test.

For the fourth parameter, we administer a modified British Picture Vocabulary Scale (BPVS3) (see Dunn and Dunn, 2009). The participants are presented with pictures (starting with two and gradually increased to eight) of different objects, and instructed to choose the correct one, based on audio instructions provided in Amharic. The instructions are recorded from an adult female native speaker (age = 28) of the standard Amharic. Each picture is assigned a label A, B, C… etc. The pictures are displayed on a white screen using a personal computer and PowerPoint. After listening to the audio instruction, the participants select the label that correctly represents the named picture and provide their response on the answer sheets, using pen or pencil. The test consists of 20 items and was administered in the temporary laboratory we established at Haramaya University. The number of correctly selected labels is taken as the picture naming fluency measure. The test is administered prior to the administration of the background questionnaire. Since we are interested in L2 speakers that have basic knowledge of Amharic vocabulary, only those who scored 15/20 and above on the picture naming task are recruited for the eye tracking experiments. All participants have normal vision, based on the self-reported responses. They consented electronically before the experiment. As a compensation, each participant received 25 Ethiopian birr. In total, 201 participants started the experiment; of these, 8 (4%) dropped out before the calibration due to lack of interest and in some cases due to an attested exposure to additional languages. Thirteen participants (6.5%) dropped out after failing the calibration. The remaining 180 (90%) participants successfully completed the experiment. The participants were recruited via research assistants. **Table 3** presents the profile of the participants that completed the experiment.

Table 3: The participants.

Language groups	Total	Mean age (range)	Sex	L2 Fluency	Gender knowledge
L1 Amharic	61	21 (19–27)	(M = 17)	8.5 (5–10)	15.5 (13–16)
L1 Eastern	61	22 (18–25)	(M = 50)	7.0 (3–10)	11.9 (9–16)
L1 Western	58	21 (18–25)	(M = 51)	7.2 (3–10)	11.4 (7–15)

4.1.2 Stimuli and procedures

4.1.2.1 The stimuli

Sixteen (16) pictures of inanimate objects were selected from electronic and print sources. Four of the nouns designated by the pictures have masculine gender both in the Eastern Oromo dialect and in Amharic (M-M), four nouns have feminine gender both in the Eastern dialect and in Amharic (F-F), four nouns have masculine gender in the Eastern dialect but feminine gender in Amharic (M-F), and the remaining four have feminine gender in the Eastern dialect, but masculine gender in Amharic (F-M). Hence, half of the pictures designate gender congruent nouns and the remaining half gender incongruent nouns. The frequency of the nouns designated by the pictures were matched across the congruent and incongruent conditions, using SKELL/SKETCH ENGINE[6] (Kilgarriff et al., 2015). There was no significant difference between the frequency of the congruent nouns and the incongruent nouns, independent sample t-test, $t = .743, p = .463$. During the test design, the gender congruent pictures were paired with each other, and the gender incongruent pictures were always paired with the gender incongruent pictures. Therefore, in the gender congruent condition, both the target and the competitor pictures have the same gender in the Eastern dialect and in Amharic. Conversely, in the gender incongruent condition, both the target and the competitor pictures have different gender in the Eastern dialect and in Amharic. The 16 pictures were presented in a 2*2*3 design: two gender conditions (same, different), two stimuli conditions (target, competitor), and three gender agreement conditions (noun-interrogative, noun-adjective, noun-demonstrative). These resulted in 192 possible target-competitor combinations, and total of 96 test items ((16*2*2*3)/2). In each item, the target picture appeared either to the right or to the left of the computer screen. Each target-competitor picture pair was presented in the same color. There were additional 14 distractors. In the distractor condition, pairs of pictures of numerals were displayed on the computer screen with audio instructions such as ‘Among the following, which is number one?’. In sum, there were 110 test items, 96 experimental and 14 distractor items. The items were randomized and equally divided into two in such a way that the gender agreement domains and distractor items are equally distributed between the categories. Half of the test items (55) was administered to half of the participants in each study group, and the remaining half was administered to the rest of the participants. The pictures were equally sized into 400 by 400 pixels.

Table 4 presents samples of the Amharic audio instructions. The audio instructions were recorded by the adult female native speaker of the Standard Amharic (age 28). We provided the adult speaker with written list of Amharic sentences and asked her to read the sentences with natural pace and accent. We segmented the recorded sentences using Audacity. We used silence to align the gap between the segments. The recorded instructions contain the three gender agreement conditions. In the noun-interrogative pronoun condition ((4a) & (4b)), the cues are gender suffixes attached to the interrogative pronoun. In the noun-adjective condition ((4c) & (4d)), the cues are gender suffixes on the first adjectives, and in the noun-demonstrative pronoun condition ((4e) & (4f)), the cues are gender suffixes attached to the demonstrative pronoun. In all

---

[6] See at https://www.sketchengine.eu/skell/.

conditions, we were interested in the time window between the onset of the words that carry the gender cues and the target nouns. In each condition, there was a gender-neutral adjective which served as a buffer between the gender cues and the target nouns. In Section 2.3, we showed that in Amharic, only definite nouns agree in gender with adjectives. Besides, when there is a sequence of adjectives in a sentence, only the first adjective obligatorily agrees in gender with noun.

Table 4: Sample Amharic audio instructions.

No.	Cond.	G	Carrier phrase	Cues	Buffer	Targets	Post-targets	Gloss (condensed)
(4a)	N-INT	F	kənnəzzih wɨst’	***yət-wa***	***k’əyy***	*s’iggerəda*	ɨndəhonəčč asayy	…which one is a red rose?
(4b)	N-INT	M	kənnəzzih wɨst	***yət-u***	***k’əyy***	c’amma	ɨndəhonə asayy	…which one is a red shoe?
(4c)	N-Adj	F	kənnəzzih wɨst’	***k’əyy -wa***	***k’onğo***	*s’iggerəda*	yət ɨndalləčč asayy	Show where a nice red rose is
(4d)	N-Adj	M	kənnəzzih wɨst’	***k’əyy -u***	***k’onğo***	c’amma	yət ɨndallə asayy	Show where a nice red shoe is
(4e)	N-DEM	F	kənnəzzih wɨst’	***yɨhččin***	***k’əyy***	*s’iggerəda*	bətikɨkkɨl assay	Select this red rose
(4f)	N-DEM	M	kənnəzzih wɨst’	***yɨhɨn***	***k’əyy***	c’amma	bətikɨkkɨl assay	Select this red shoe
			1700ms	**1100ms**	**900ms**			
				Target window				

The audio stimuli were sequenced as follows: the duration between the onset of the carrier phrases and the words that carry the gender cues was 1700ms. The duration from the onset of the carrier phrases to the onset of the buffers was 2800ms (1700 + 1100ms). The duration from the onset of the carrier phrases to the onset of the target nouns was 3700ms (2800ms + 900ms). The critical time window for the anticipatory processing extends from the onset of the words that carries gender cues to the onset of the target nouns (1100 + 900 = 2000).

4.1.2.2 Procedures

The experiment was programmed in the Gorilla Experiment Builder platform (Anwyl-Irvine et al., 2020). We used the Gorilla platform which employs the recent version of WebGazer (available since November 11, 2021). The study started with written instructions explaining the purpose and the general procedure of the experiment. Then, the participants were directed to the eye-tracking task. The task consisted of 2000ms preview time followed by a fixation cross. The fixation cross was displayed for 500ms. Then, the eye tracking window appeared in which each pair of the target and competitor pictures was displayed along with the audio instruction. The pair of the pictures was displayed (see **Figure 1**) indefinitely until one of the pictures was selected by the participants, using the mouses-click. The experiment was conducted on 4 PCs with the monitor size of 1366 × 768, 1280 × 720, 1024 × 768, and 1536 × 864 pixels. Maximum of four participants were tested simultaneously, wearing headphones.

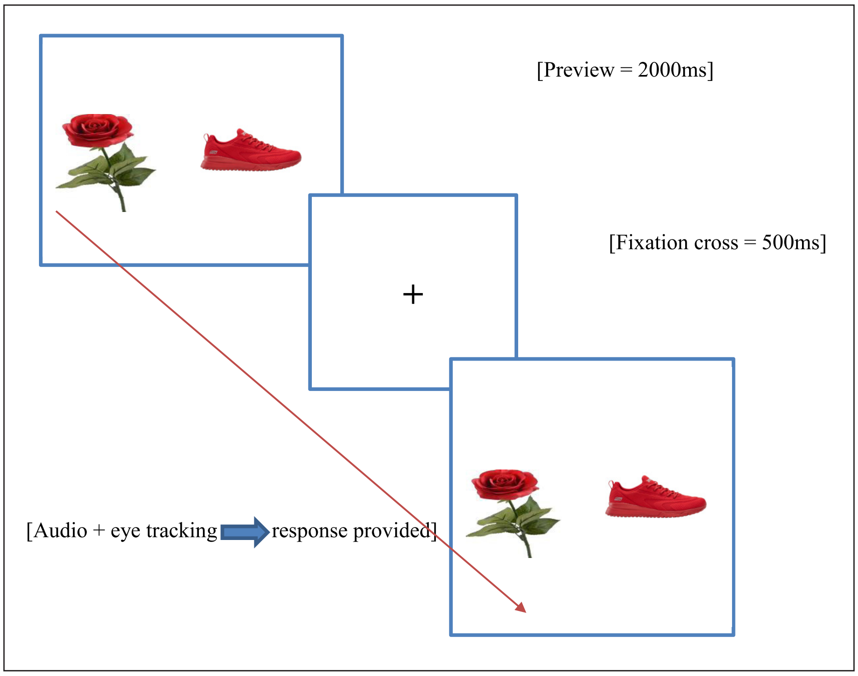


Figure 1: The sequence of stimuli displays in the visual world paradigm, *s'iggerəda* 'rose' in the left and *c'amma* 'shoe' in the right.

The participants were instructed to choose the pictures based on the audio instructions. They were also told that the task would last approximately 25 minutes. They were not instructed to sit at a particular distance from the screen, but before the experiment it was assured that the laboratory had sufficient light. The experiment began with a practice session of 10 items from different materials. Then, the participants were directed through a calibration phase of 5 points presented consecutively across the entire screen. They were instructed to look at and click on each point which was visible for 3 seconds. They were told to look at the points without moving their heads. Each participant had to pass the calibration, which we set to 60% (3/5) of the gaze prediction. Following the validation, the participants received visual feedback (plots) on their performance. If 60% of the threshold was reached for the validation plot, the participants could proceed to the experiment. If not, they were looped back to the start of the calibration phase. If the calibration was not successful within 5 attempts, the study was aborted. The calibration was repeated once half-way in the experiment, after 28 trials, to compensate for the small head movements. The total study duration of the experiment was around 25 minutes.

4.2 The picture description production task

After the eye-tracking task, participants performed a gender assignment task. The task was used to determine the participants' knowledge of L2 lexical gender. The same 16 pictures from the eye-tracking experiment were used in the task. Each picture, for example *rose*, appeared at the center of the computer screen. Then, the participants were asked to describe orally the advantage of the object designated by picture, using a complete Amharic sentence. After seeing the picture of *rose*, for example, the participants produced sentences such as in (5).

(5)	S'ɨggerəda	fɨk'ɨr	lə-məgləs'	tɨt'ək'mal-ləčč.
	rose	love	to-express	serve-3.F.SG
	'A rose serves to express love'.			

The gender assigned to the target noun (rose) was determined based on the gender agreement that was established between the noun *rose* and the verb *serve*. Since Amharic allows gender agreement between nouns and verbs, we exploited this agreement relationship to probe into the participants' knowledge of L2 gender assignment. The participants were given a score of 1 for the correct agreement and 0 otherwise. The L2 gender assignment score is the number of responses that contain the correct gender agreement between the noun and the verb.

4.3 Preprocessing steps

The Gorilla engine provides *raw* and *normalized* coordinates of the eye fixation, and we used the normalized coordinates in our analyses. Gorilla also provides *face-conf,* which is a data quality metrics. Face-conf indicates how strongly the image under the model resembles a face. Its values vary from 0 to 1, and values greater than 0.5 are indicative of a good model fit. We excluded values less than 0.6 on the face-conf metrics which together constituted 8 (4.4%) participants. The sampling rate of our original sample (N = 180) varied from 3.88Hz to 59.42 Hz (M = 36Hz). We excluded additional two participants whose sampling rate was less 5Hz. In total, 10 (5.5%) participants were excluded. The average sampling rate in the resulting groups (170 participants) was 37Hz. This means that, in average, 37 data points were sampled every 1000ms, and one data point was acquired every 27ms.

Our analyses were based on the proportion of eye fixations within the preselected time window, 2000ms from the onset of the word that carries the gender cues to the onset of the target nouns. While determining the cues onsets, we included 200ms extra time to compensate for the time taken to launch the eye movement. Then, we performed two-step statistical analyses – first on the time course of prediction, then on the factors that modulate predictive L2 processing. We inspected the time course of prediction using divergence point analysis (Stone et al., 2021) and cluster-based permutation test (Ito & Knoeferle, 2023; Maris & Oostenveld, 2007; Meyer et al., 2021). In the divergence point analysis, the language group was independent variable, and the dependent variable was the proportion of looks to the target and competitor pictures. We used non-parametric bootstrapping divergence point in which the existing dataset is resampled multiple times to generate "new" datasets, and a statistical test is applied after each resample. The resampling was stratified by subject, timepoint, and stimuli conditions (target and competitor). A new divergence point was estimated after each resampling. Then, a distribution of divergence points was generated whose mean was taken as the overall divergence point. We bootstrapped means and confidence intervals (CIs) to determine the significant point of divergence.

For further condition-specific fine-grained analysis, we performed cluster-based permutation test following the manual provided by Ito and Knoeferle (2023). In this case, the agreement conditions (noun-adjective, noun-demonstrative and noun-interrogative) and language groups were independent variables, and the proportion of looks to the target and competitor pictures was the dependent variable. Cluster-based permutation analysis (CPA) is also a non-parametric test, and it detects the difference between two conditions in a preselected time window – 2000ms in our case. A significant result indicates a robust effect of condition. We performed the cluster analysis based on LME, using the *clusterperm.lmer* function in the permutes R package. In both divergence and cluster-based analyses, we grouped the timepoints into 30ms bins because one data point was acquired approximately every 27ms which can be rounded to 30ms.

The analysis of the time course of prediction was followed by the analysis of the factors that modulate the predictive processing. For the latter analysis, we performed logistic mixed effects regression to detect the effect of the gender conditions (same and different), language groups (L1 Amharic, L1 Eastern and L1 Western), knowledge of lexical gender, and L2 fluency on the proportion of looks to the target pictures. We performed a similar but separate analysis for the congruency effect. The separate analysis was necessary since gender congruency could be established only between the Eastern Oromo dialect and Amharic. For the regression analysis, we grouped the timepoints into 50ms bins since we wanted to reduce the potential effect of autocorrelation by grouping the timepoints into larger bins. To reduce noise in the data, both time course and factor analyses were performed only on nouns whose lexical gender was properly assigned by the participants. We also excluded eye fixations outside the participants' screen dimension. The analyses were conducted in R (version 4.2.3, 2023).

5. Results

In Section 4.1.2, we indicated that there are *different* (predictable) and *same* (unpredictable) gender conditions in our experimental design. To inspect the time course of prediction, we performed a divergence point analysis for each gender condition. The analysis was done in R

(R Core Team, 2023). We defined our interest period from the onset of the words containing the gender cue to 2000ms after the onset of the target noun. We included a post-target 2000ms time window since the sustained target-over-competitor preference may emerge after the onset of the target noun, for the groups that may not predict. For the data visualization and statistical analysis, we followed the tutorial provided by Stone et al. (2021). We were particularly interested in whether the estimated divergence point is prior to the onset of the target nouns, indicating prediction of the upcoming nouns. Based on Stone et al. (2021), we first ran a one-sample t-test on target fixation proportions against chance over aggregated items. We set as a divergence point the first time point in a run of at least 10 consecutive time points with significant *t*-values. Next, we created new datasets by means of non-parametric bootstrapping, where the data were resampled 2000 times by replacement within the categories of subject, timepoint and stimuli condition (*target vs. competitor*); a new divergence point was estimated after each resampling.

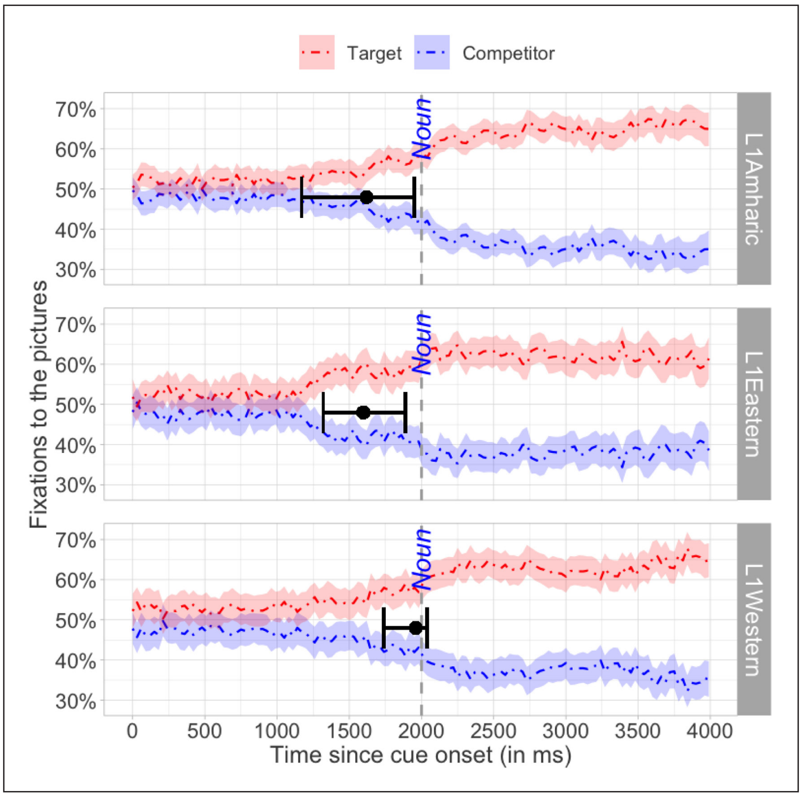


Figure 2: Divergence points and 95% confidence intervals superimposed on the fixation curves. Points with error bar indicate the bootstrap mean and its 95% percentile and confidence interval which reflects divergence point and its temporal uncertainty. The broken vertical line represents the onset of the target noun. L1 Eastern Oromo dialect speakers show the earliest predictive onsets 1594ms [1320, 1890] followed by L1 Amharic speakers 1619ms [1200, 1950]. L2 Western dialect speakers showed a reduced anticipatory gaze 1995ms [1740, 2090].

Figure 2 shows the fixation proportion for the target and competitor pictures of the *different* condition along with the estimated divergence point and 95% confidence interval. For the L1 Amharic speakers, the earliest divergence onset was 1619ms [1200, 1950] after the onset of the gender cues and for the L1 Eastern dialect speakers, the earliest divergence onset was 1594ms [1320, 1890] after the onset of the gender cues. For the L1 Western Oromo dialect speakers, the earlies divergence onset was 1995ms [1740, 2090] after the onset of the gender cues. The mean difference in the divergence point, when L1 Amharic speakers are compared to L2 (L1 Eastern + L1 Western), was 210ms [–570, 330]. Since the confidence interval included zero, there was no reliable difference between the L1 Amharic group and the L2 Amharic groups. The mean difference in the divergence points when L1 Western compared to the L1 Eastern was 385ms [–660, –60]. Since the confidence interval did not include zero, there was a reliable difference between the two groups, indicating earlier onset predictive looks by the L1 Eastern dialect speakers.

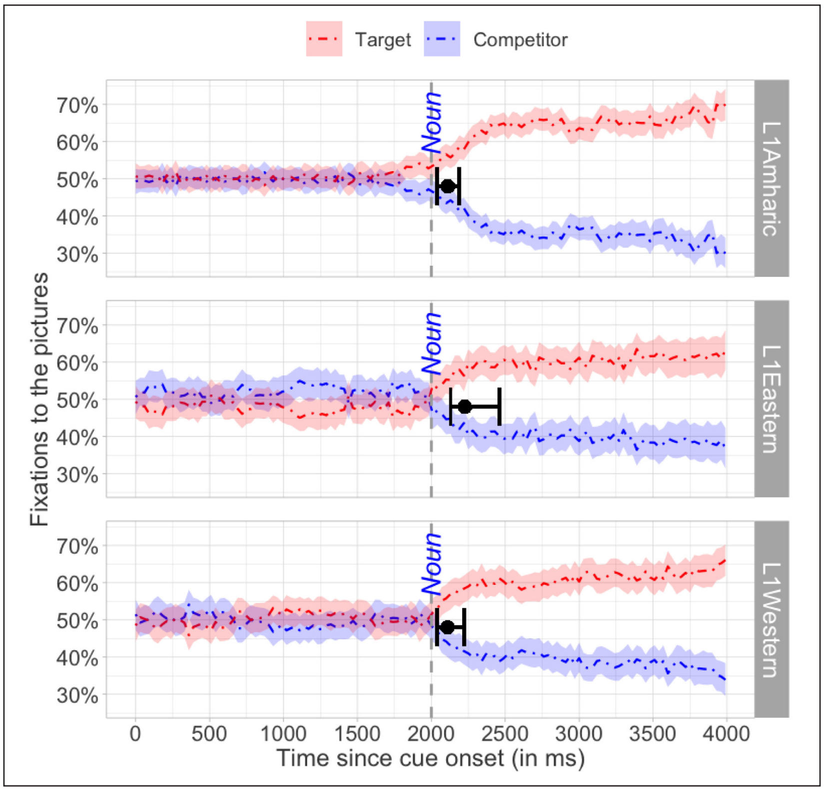


Figure 3: Divergence points and 95% confidence intervals superimposed on the fixation curves. Points with error bar indicate the bootstrap mean and its 95% percentile and confidence interval which reflects divergence point and its temporal uncertainty. The broken vertical line represents the onset of the target noun. None of the language groups show predictive onsets in the nonpredictive *same* condition.

Figure 3 shows the fixation proportion for the target and competitor pictures of the *same condition* along with the estimated divergence point and 95% confidence interval. For the L1 Amharic speakers, the earliest divergence time onset was 2112ms [2050, 2180] after the onset of the gender cues and for the L1 Eastern Oromo speakers, the earliest divergence time onset was 2227ms [2100, 2450] after the onset of the gender cues. For the L1 Western Oromo speakers, the earlies divergence time onset was 2109ms [2040, 2260] after the onset of the gender cues. The mean difference in the divergence points when the L1 Amharic speakers were compared to L2 (L1 Eastern + L1 Western) was 35ms [–150, 60]. Since the confidence interval included zero, there was no reliable difference between the L1 Amharic and the L2 Amharic groups. The mean difference in the divergence points when L1 Western compared to the L1 Eastern was 10ms [–60, 360]. Since the confidence interval included zero, there was no reliable difference between the two groups.

The results of the divergence point analysis indicate that L1 Amharic and L1 Eastern dialect speakers show an earlier divergence look to target vs. competitors in the *different* gender condition than the L1 Western dialect speakers. However, in our analysis, we collapsed the three gender agreement conditions, ignoring the likelihood of agreement-specific eye fixation differences. It could be that the L2 Amharic speakers' predictive processing which otherwise stands out at the agreement-specific levels was masked by our approach. To inspect if this was the case, we performed a separate agreement-specific cluster-based permutation analysis on the L1 Western and the L1 Eastern dialect speakers' proportion of eye fixations for the *different* gender condition. For the data visualization and statistical analysis, we followed the tutorial provided by Ito and Knoeferle (2023). Prior to the analysis, we computed a *log-ratio* which indicates a fixation bias towards the target over the competitor using *log ((Look to the target + 0.5) / (Look to the competitor + 0.5))*. First, we identified clusters of small time-bins (30ms) in which the effect was significant, using by-subject and by-item linear mixed effect regression (Baayen et al., 2008), and we computed *cluster-mass statistics* – the sum of all the individual test statistics. Then, we created null hypothesis distribution which each cluster in the original dataset was compared against. After reshuffling the original dataset, we repeated this step several times (*nperm = 1000*). Afterwards, we compared each cluster in the original data set with the null hypothesis distribution. We defined our period of interest from 500ms before the onset of words containing the gender cues to 2000ms after the onset of the target noun. We included the time window prior to the onset of the gender cue to account for a possible baseline effect. We included the post-target 2000ms since the sustained target-over-competitor preference may emerge after the onset of the target noun, for the groups that may not predict.

Figure 4 shows the proportion of eye fixations of the L1 Eastern Oromo dialect speakers across the three gender agreement conditions (noun-adjective, noun-demonstrative, noun-interrogative). In the noun-adjective gender agreement condition, we found a significant difference between the stimuli (target and competitor) conditions (1180–4000, cluster mass statistics = 2906, $p <$ 0.001) which indicates that the target pictures were inspected more often than the competitor pictures. Similarly, in the noun-demonstrative pronoun gender agreement condition, we found a significant difference between the stimuli conditions (1630–1750ms, cluster mass statistics = 31.5, $p < 0.001$). Likewise, in the noun-interrogative pronoun gender agreement condition, we found a significant difference between the two stimuli conditions (1840–4000ms, cluster mass statistics = 1176, $p < 0.01$) which implies that the target pictures were inspected more often than the competitors. These results entail that the L1 Eastern Oromo dialect speakers employ prediction in all agreement conditions.

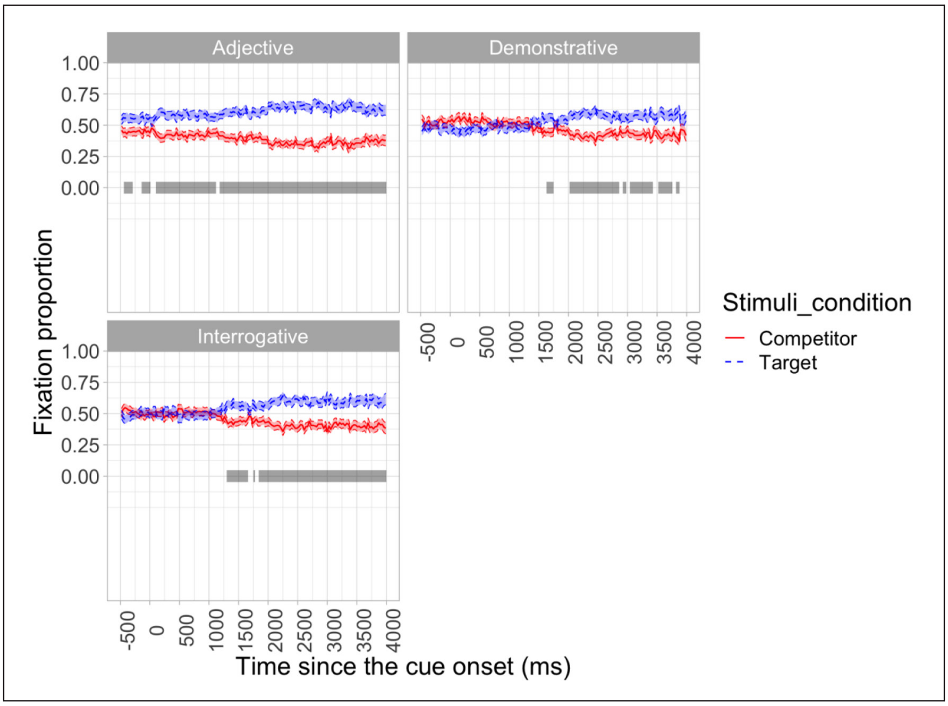


Figure 4: Proportion of eye fixations of the L1 Eastern dialect speakers across the agreement conditions; the thick grey horizontal line indicates a significant difference between the stimuli conditions (target and competitor).

Figure 5 shows the proportion of eye fixations of the L1 Western dialect speakers in the three gender agreement conditions. We found a significant difference between the stimuli conditions (1240–4000ms, cluster mass statistics = 2945, $p < 0.001$) in the noun-adjective gender agreement condition which suggests that the target pictures were inspected more often than the competitors. Similarly, in the noun-demonstrative pronoun gender agreement condition, we found a significant difference between the stimuli conditions (1660–4000ms, cluster mass statistics = 2104, $p < 0.001$). In the noun-interrogative pronoun gender agreement condition, there was no significant difference between the two stimuli conditions; looks to the target pictures significantly differed from looks to the competitors 2000ms after the onset of the gender cues (2010–4000ms, cluster mass statistics = 877, $p < 0.001$). This means that the L1 Western Oromo dialect speakers employ prediction only in the noun-adjective and in the noun-demonstrative pronoun gender agreement conditions.

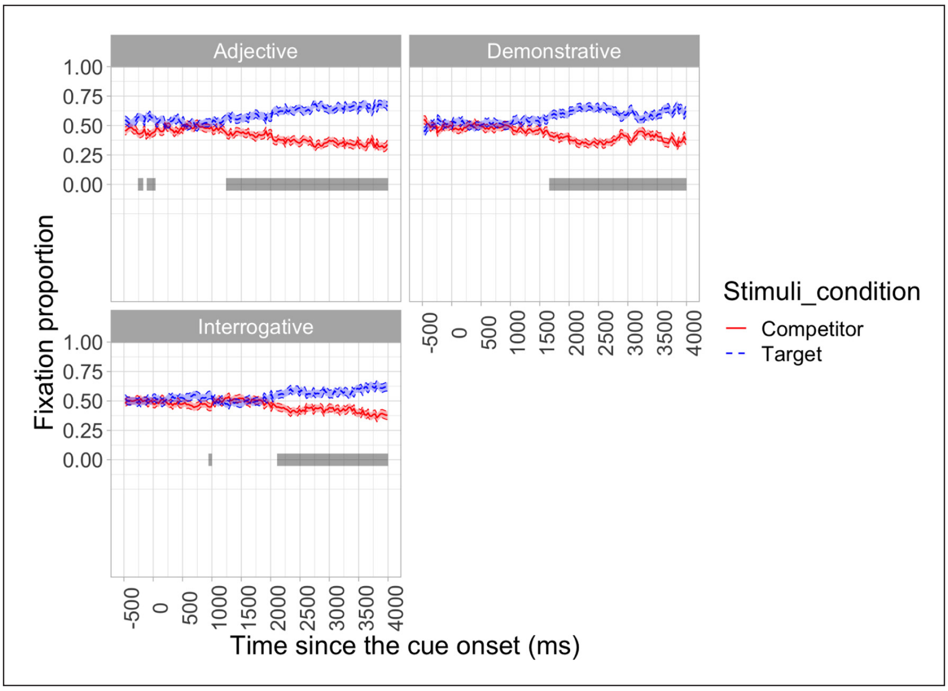


Figure 5: Proportion of eye fixations of the L1 Western dialect speakers across the agreement conditions; the thick grey horizontal line indicates a significant difference between the stimuli conditions (target and competitor)

We employed a logistic mixed effects regression model to examine the effect of the individual level factors on predictive L2 gender processing. We performed a logistic regression with subject and item intercepts and random slopes for language groups and gender conditions, using the lmerTest package (Kuznetsova et al., 2017). Specifically, we analyzed the effects of gender conditions (same and different), language groups (L1 Amharic, L1 Eastern and L1 Western), knowledge of lexical gender, and L2 fluency on the proportion of eye fixations. The dependent variable (the proportion of eye fixations) was coded as ‘1’ if the look was to the target picture and as ‘0’ if the look was to the competitor picture. The continuous independent variables (L2 fluency and knowledge of lexical gender) were centered around the mean. The categorical variables were coded as follows: we employed Helmert coding for the language groups and compared the L1 Amharic speakers with the L2 Amharic groups (L1 Eastern + L1 Western), and the L1

Eastern dialect speakers with the L1 Western dialect speakers. We deviation-coded the gender conditions and compared each level of the variable with the overall mean. We coded the *same* level as –0.5 and the *different* level as 0.5. We started model filling with the most complex one and gradually simplified until there were no issues of convergence based on Bates et al. (2015). We estimated the model parameters using restricted maximum likelihood estimation (REML). We used the likelihood ratio to determine the main effect of each independent variable. To this end, we created a null model by removing the mixed effect of interest but retaining all other fixed and random effects. Then, we compared the null model with the full model – the maximally best fitting model. We relied on the odds ratios to show the relative influence of the levels of the categorical variables. The odds ratios are presented in **Table 5**.

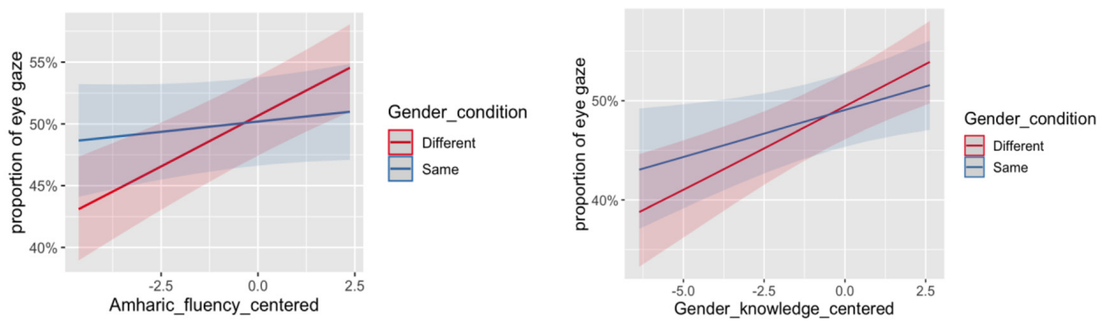
Table 5: Effects of gender conditions, knowledge of lexical gender and L2 fluency.

	Target Fixations		
Predictors	**Odds Ratios**	**CI**	**p**
(Intercept)	1.03	0.91–1.18	0.620
Language group [L2 Amharic]	0.81	0.70–0.93	0.004
Language group [L1 Western]	0.88	0.78–0.98	0.022
Gender condition [Different]	1.01	0.87–1.19	0.859
Amharic fluency	1.03	1.00–1.06	0.033
Knowledge of lexical gender	1.06	1.02–1.09	0.001
Gender condition [Different] × Amharic fluency	1.03	1.03–1.04	0.001
Gender condition [Different] × Knowledge of lexical gender	1.03	1.02–1.04	0.001
Amharic fluency × Knowledge of lexical gender	1.01	1.00–1.02	0.196

Observations: 507045, Marginal R^2 / Conditional R^2: 0.004 / 0.055.

The regression analysis results revealed that there was a significant effect of language group ($\chi^2 = 11.3$, $p < 0.01$); the proportion of eye fixations of the L2 Amharic group (L1 Eastern + L1 Western) was significantly smaller than the proportion of eye fixations of the L1 Amharic speakers (OR 0.81, $p < 0.01$). Similarly, the proportion of eye fixations of the L1 Western dialect speakers was significantly smaller than the proportion of eye fixations of the L1 Eastern dialect speakers (OR 0.88, $p < 0.01$). We also found significant effects of L2 fluency ($\chi^2 = 8.04$, $p < 0.001$) and knowledge of lexical gender ($\chi^2 = 8.51$, $p < 0.001$). In other words, higher scores on the L2 fluency and on the knowledge of lexical gender resulted in an overall increase in the proportion of fixations to the target pictures.

There was no significant effect of gender conditions ($\chi2 = 0.04$, $p = 0.840$), but we found a significant interaction between L2 fluency and gender conditions ($\chi^2 = 368$, $p < 0.001$). **Figure 6(a)** shows that as the participants' L2 fluency scores increase, the proportion of their looks to the target pictures increases significantly in the *different* gender condition (OR 1.03, $p < 0.001$). Likewise, there was a significant interaction between the knowledge of lexical gender and the gender conditions ($\chi^2 = 365$, $p < 0.001$). **Figure 6(b)** shows that as the participants' scores on the knowledge of lexical gender increase, the participants' proportion of looks to the target pictures also increases in the *different* gender condition (OR 1.03, $p < 0.001$). There was no significant interaction between the knowledge of lexical gender and L2 fluency ($\chi^2 = 1.40$, $p = 0.240$).



(a) (b)

Figure 6: The interaction between L2 fluency, knowledge of lexical gender and gender conditions.

Since the Western Oromo dialect does not have grammatical gender, the gender congruency effect was computed only for the L1 Eastern dialect speakers and the L1 Amharic speakers. We used deviation coding and compared each level of the categorical variables to the overall mean. We coded the *incongruent* condition of the congruency conditions as –0.5 and the *congruent* condition as 0.5. As to the gender conditions, we coded the *different* condition as 0.5 and the *same* condition as –0.5. Regarding the language groups, we coded L1 Amharic group as –0.5 and the L1 Eastern dialect group as 0.5. Then, we performed logistic mixed effects regression with subject and item intercepts and random slopes for the language groups and gender conditions. The results of the regression analysis revealed that there was no statistically significant effect of congruency of lexical gender ($\chi^2 = 0$). However, as can be seen from **Table 6**, the proportion of eye fixations of the *congruent* condition was higher than the overall mean proportion of the eye fixations (OR 1.38, $p < 0.001$). Similarly, there was no significant effect of language groups ($\chi2 = 0$). However, we found a significant interaction between the language groups and gender congruency ($\chi^2 = 10.40$, $p < 0.001$); in the *congruent* gender condition, the proportion of eye fixations of the L1 Eastern dialect speakers increased significantly (OR 1.38, $p < 0.001$). Nevertheless, there was no significant interaction between the gender congruency and gender conditions ($\chi^2 = 0.02$, $p = 0.900$); the increased proportion of eye fixations in the *congruent* condition was the same for the predictable *different* and the unpredictable *same* gender conditions.

Table 6: Effect of lexical gender congruency on L2 predictive processing.

	Target Fixation		
Predictors	**Odds Ratios**	**CI**	**p**
(Intercept)	0.99	0.88–1.11	0.820
Gender congruency [Congruent]	1.38	1.15–1.66	0.001
Language group [L1Eastern]	0.90	0.81–1.01	0.082
Gender condition [Different]	1.03	0.89–1.21	0.670
Gender congruency [Congruent] × Language group [L1Eastern]	1.05	1.02–1.08	0.001
Gender congruency [Congruent] × Gender condition [Different]	0.98	0.72–1.33	0.891

Observations: 365,820, Marginal R^2 / Conditional R^2: 0.008 / 0.055.

6. Discussion

In this study, we aimed at addressing three research objectives: (1) determining the extent to which predictive L2 gender processing is influenced by differential cross-linguistic influence, (2) examining the interaction between L2 fluency and predictive L2 gender processing, and (3) exploring the relationship between knowledge of L2 lexical gender and predictive L2 gender processing. This section discusses each point based on the results reported in Section 5 and the arguments provided in previous studies.

1.1 Predictive processing and the role of differential cross-linguistic influence

We have seen that the L1 Eastern dialect speakers have earlier onset of predictive looks than the L1 Western dialect sparkers, implying that cross-linguistic influence is a crucial factor in L2 predictive gender processing. We can take this as evidence of facilitation from the Eastern Oromo dialect. Several previous studies recognized the pivotal role that cross-linguistic influence from the first language plays in predictive L2 gender processing (e.g., Bosch & Foppolo, 2022; Dussias et al., 2013; Garrido-Pozú, 2022; Hopp, 2016; Hopp & Lemmerth, 2018; Morales et al., 2016). Data from the Oromo-Amharic bilingual speakers further elucidate that morphosyntactic similarity between the Eastern Oromo dialect and Amharic not just enhances the predictive processing of Amharic gender agreement, but also enforces predictive processing. Quantitatively, the L1 Eastern dialect speakers predicted faster than the L1 Amharic speakers. Such a strong enforcement of grammars of the first language was previously reported by Meir et al. (2020) in the study they conducted on the processing of case marking by Russian-Hebrew bilingual speakers. In their study, Russian-Hebrew bilingual speakers exploited Hebrew case markers more effectively than the monolingual Hebrew speakers. The strong enforcement during the L2 case processing was associated with the robust case marking in Russian as opposed to the sporadic case marking in Hebrew. When it comes to the Oromo-Amharic bilingual speakers, we have seen that in Amharic, gender is assigned based on semantic parameters. In contrast, the Eastern Oromo dialect has robust phonological endings or declension classes that make the gender of the nouns unambiguously identifiable. The Eastern Oromo dialect also has a richer gender morphology than Amharic. For example, only attributive adjectives agree with nouns in gender in Amharic. However, in Oromo, every adjective agrees with noun in gender.

We have seen that lexical gender congruency between Amharic and the Eastern Oromo dialect does not significantly affect predictive L2 gender processing. This means that lexical gender congruency does not conspire with the syntactic gender congruency in the process of facilitating predictive L2 gender processing. Many previous studies argue that lexical gender congruency plays a crucial role in predictive L2 gender processing (e.g., Hopp, 2016; Lago et al., 2021; Morales et al., 2016; Paolieri et al., 2020). The Oromo-Amharic bilingual data do not replicate this claim. As we indicated in the introduction and elsewhere, speakers of the Eastern Oromo dialect are exposed to the neutralized Western Oromo dialect. We suspect that the facilitative role of the lexical gender congruency between Amharic and the Eastern Oromo dialect is probably suppressed by the speakers' exposure to the neutralized Western Oromo dialect.

1.2 The role of L2 fluency and knowledge of L2 lexical gender

In the above discussions, we argued that cross-linguistic influence is the most important factor that mediates L2 predictive gender processing. However, cross-linguistic influence is not the only important factor. Knowledge of L2 lexical gender influences L2 predictive gender processing too. We have seen that knowledge of lexical gender interacts with the gender condition. In other words, speakers that have better knowledge of lexical gender frequently gaze at the target pictures in the predictable *different* gender condition than speakers that have little knowledge of lexical gender. This clearly shows that there is a link between the mastery of L2 lexical gender and L2 predictive gender processing. This is not the first time this has been found. Several previous studies recognized the importance of the knowledge of L2 lexical gender in predictive L2 gender processing (see Clashen & Felser, 2006; Grüter et al., 2012; Hopp, 2013, 2016; Lemmerth & Hopp, 2018, 2019; Prévost & White, 2000). Indeed, most lexicalist models of L2 gender processing such as the Shallow Structure Hypothesis (Clashen & Felser, 2006), the Lexical Bottleneck Hypothesis (Hopp, 2018), the Missing Surface Inflection Hypothesis (Prévost & White, 2000), and the Lexical Gender Learning Hypothesis (Grüter et al., 2012) heavily rely on the integration between lexical and syntactic representations. As it stands, it seems that the knowledge of L2 lexical gender contributes to L2 predictive gender processing. In our case, it is a secondary factor of importance after differential cross-linguistic influence.

We have also seen that there is a strong association between L2 fluency and predictive L2 gender processing. Speakers that have advanced Amharic fluency frequently gaze at the target pictures in the predictable *different* gender condition than speakers that have lower Amharic fluency. The evidence suggests that a native-like fluency in L2 leads to an automated processing of L2 gender agreement. The result strengthens previous studies that reported a positive association between L2 fluency and L2 predictive gender processing (e.g., Blumenfeld et al., 2016; Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013; Hopp, 2016; Hopp & Lemmerth, 2018).

More importantly, from the results of the cluster-based permutation analysis we have learned that L1 Eastern dialect speakers effectively employ the gender cues to anticipate the upcoming nouns in all agreement conditions. The inconsistency in the noun-pronominal order does not affect the predictive processing of the L1 Eastern dialect speakers. In Amharic, interrogative pronouns and demonstrative pronouns have flexible word order in a sentence. This flexibility in the position of the gender cues only influences the predictive processing of the L1Western dialect speakers; the L1 Western dialect speakers do not predict in the most flexible noun-interrogative pronoun gender agreement condition. The result partly coincides with previous studies that stress the importance of cue availability or reliability (Falk & Bardel, 2011; Garrido-Pozu, 2022; Grüter & Rohde, 2013; Hawkins, 2009; Jakubowicz & Roulet, 2004; Kaan, 2014; Kaan & Grüter, 2021). Meanwhile, it also indicates that L2 learners can withstand input variability in the L2 if they are native speakers of languages or dialects that have rich gender morphology. It could be that, once the gender system of the first language is instantiated in the early childhood, cue availability or reliability ceases to become an essential factor in predictive L2 processing.

In general, data obtained from the Oromo-Amharic bilingual speakers imply that bilingual speakers predict only in a conducive condition, i.e., in the condition that is attenuated by facilitation from the first language, knowledge of L2 lexical gender, L2 fluency, and the availability of the gender cues. Some previous studies also argue that predictive processing is deployed only in a conducive condition (see Schlenter, 2023). By implication, prediction during L2 gender processing is *optional*; it aids the processing automatization only if the condition is favorable. In the unfavorable condition, learners may rely on other cues such as semantic relationship (see Prévost & White, 2000). Apparently, the joint contribution of several factors creates a suitable condition that fosters the deployment of prediction during L2 gender processing. The optionality of prediction during L2 gender processing stands against previous representation-based models of L2 predictive processing such as the Reduced Ability to Generate Expectation (RAGE) (Grüter & Rohde, 2013) and the Representational Deficit Hypothesis (Hawkins, 2009). These models assume that L2 learners whose first languages do not encode grammatical gender may fail to acquire the capacity to predict. Data from the Oromo-Amharic speakers confirm that the lack of prediction is not an inherent trait of L2 learners. L1 Western L2 Amharic speakers whose first language or dialect does not encode grammatical gender could employ gender cues to predict, given that the condition is conducive. If there exists an inherent deficit associated with L2 prediction, this would not have been possible. Rather, the results indicate that whether to predict or not during L2 gender processing is largely determined by language-specific and individual level factors. In this regard, we illustrated that L1 Western dialect speakers may or may not be able to predict depending on their level of L2 fluency, knowledge of L2 lexical gender, and the availability of the gender cues. Many studies have previously arrived at the same conclusion but based on well-studied languages (see Kaan, 2014; Schlenter, 2023).

7. Conclusion and theoretical implications

The take-home message of our study is that late L2 speakers can exploit gender cues to anticipate upcoming nouns during sentence processing. At the same time, their prediction capability can be constrained by cross-linguistic influence from the L1, knowledge of lexical gender, L2 fluency, and specific properties of a grammatical feature. Differential cross-linguistic influence is the main factor that modulates L2 predictive gender processing. The combination of morpho-syntactic similarity and other factors creates the maximal conducive context for the effective use of prediction in L2 gender processing.

These findings have ecological and theoretical relevance. Most theories of L2 gender processing are derived from a few well-studied languages of the Western world. In this regard, the present study plays a crucial role in bringing into light some of the many understudied languages. More importunately, the study has defied the traditional approach in cross-linguistic influence in L2 acquisition by shedding light on the importance of variation within the first language. It investigates differential cross-linguistic influences of dialects of the first language. We have argued that the same first language can differentially influence processing a language that is acquired at a later age. This is an important revelation against the traditional approach that coarsely compares first language to second language. Our study suggests that an adequate description of the potential influence of the previously acquired language requires a fine-grained examination of the variability in both first and second languages.

Moreover, whether the difference between L1 and L2 predictive gender processing is *processing-related* or *representational* has been subject to debate. Some earlier studies argued that late L2 learners have a representational deficit implying that prediction in late L2 gender processing is unachievable unless grammatical gender is instantiated in the first language (see Grüter & Rohde, 2013; Hawkins, 2009). However, most recent studies argue that L2 learners do not have an inherent representational deficit associated with predictive L2 gender processing (see Kaan, 2014; Lago et al., 2021; Schlenter, 2023). Data obtained from the Oromo-Amharic late bilingual speakers suggest that these later studies are on the right track. We have seen that both the speakers of the Western and the Eastern Oromo dialects are able to predict in a conducive context. The only difference between the two groups is that the prediction capability of the Western dialect speakers is more constrained by domain-specific properties of the second language and by individual-level factors. In short, our findings pattern with the processing-based models of predictive L2 gender processing. L2 learners do not seem to have inherent problem with prediction, but their processing ability can be hampered by language-specific and individual-level factors. Among the relevant factors, cross-linguistic influence is the most prominent one.

Appendices

A.1. Effects of language group, knowledge of lexical gender and L2 fluency

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: AOI ~ Language_group + Gender_condition + Amharic_fluency_centered +
Gender_knowledge_centered + Amharic_fluency_centered*Gender_condition + Gender_knowledge_centered *
Gender_condition + Amharic_fluency_centered*Gender_knowledge_centered + (1 + Language_group |
participant_id) + (1 + Gender_condition | target_item)
   Data: statzdat
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  683393   683594  -341679   683357   507027

Scaled residuals:
   Min     1Q Median     3Q    Max
-1.908 -0.961  0.664  0.937  2.131

Random effects:
 Groups         Name                            Variance Std.Dev. Corr
 participant_id (Intercept)                     0.00654  0.0808
                Language_group[L2Amharic]       0.10103  0.3179   -0.04
                Language_group[L2Western]       0.33138  0.5757    0.29 -0.34
 target_item    (Intercept)                     0.07231  0.2689
                Gender_condition [Different]    0.12022  0.3467   -0.15
Number of obs: 507045, groups:  participant_id, 170; target_item, 16

Fixed effects:
                                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                0.03314    0.06683    0.50   0.6200
Language_group[L2Amharic]                                 -0.21208    0.07391   -2.87   0.0041 **
Language_group [L1Western]                                -0.13092    0.05700   -2.30   0.0216 *
Gender_condition Different]                                0.01435    0.08051    0.18   0.8586
Amharic_fluency_centered                                   0.02801    0.01311    2.14   0.0326 *
Gender_knowledge_centered                                  0.05509    0.01690    3.26   0.0011 **
Gender_condition[Different]:Amharic_fluency_centered       0.03211    0.00359    8.93   <2e-16 ***
Gender_condition[Different]:Gender_knowledge_centered      0.02989    0.00271   11.04   <2e-16 ***
Amharic_fluency_centered:Gender_knowledge_centered         0.00690    0.00534    1.29   0.1964
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Lngg_1 Lngg_2 Gnd_G_ Amhr__ Gndr__ G_G_:A G_G_:G
Langug_grp1 -0.131
Langug_grp2  0.051 -0.040
Gndr_cndtG_ -0.137 -0.018  0.023
Amhrc_flnc_ -0.032  0.078  0.102  0.003
Gndr_knwld_  0.097 -0.817 -0.104  0.013 -0.278
Gndr_G_:A__  0.002 -0.001  0.002  0.000  0.002 -0.001
Gndr_G_:G__ -0.001  0.002 -0.001 -0.002 -0.004 -0.001 -0.398
Amhrc__:G__ -0.097 -0.083  0.007  0.002  0.188  0.069 -0.002  0.000
```

A.2. The Effect of gender congruency

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: AOI ~ Language_group + Gender_condition + gender_congruency + gender_congruency * Language_group +
gender_congruency * Gender_condition + (1 + Language_group | participant_id) + (1 + Gender_condition | target_item)
   Data: statzdat_1
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  493423   493552  -246699   493399   365808

Scaled residuals:
   Min     1Q Median     3Q    Max
-1.911 -0.959  0.652  0.940  2.237

Random effects:
 Groups         Name                             Variance Std.Dev. Corr
 participant_id (Intercept)                      0.0221   0.149
                Language_group                   0.3143   0.561    0.50
 target_item    (Intercept)                      0.0463   0.215
                Gender_condition                 0.1170   0.342    -0.09
Number of obs: 365820, groups:  participant_id, 116; target_item, 16

Fixed effects:
                                                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                        -0.0134     0.0591   -0.23   0.8203
Language_group[Eastern]                                            -0.1003     0.0578   -1.74   0.0825 .
Gender_condition [Different]                                        0.0340     0.0796    0.43   0.6696
gender_congruency [Congruent]                                       0.3210     0.0935    3.43   0.0006 ***
Language_group [Eastern]:gender_congruency[Congruent]               0.0467     0.0144    3.24   0.0012 **
Gender_condition [Different]:gender_congruency[Congruent]          -0.0214     0.1554   -0.14   0.8906
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Lng_L_ Gnd_G_ gndr__ L_L_:_
Lngg_grpLn_  0.184
Gndr_cndtG_ -0.083 -0.008
gndr_cngrn_  0.019  0.003 -0.003
Lngg_gL_:__ -0.005 -0.008 -0.002  0.017
Gndr_cG_:__ -0.040 -0.015  0.009 -0.049 -0.003
```

Data Availability Statement

All stimulus materials, code and data associated with the study (including supplementary materials) are available at https://osf.io/k6q53/?view_only=a64c3fc16bf442ada4893b257a1a4d13.

Ethics and consent

The experiment received approval of the Norwegian Center for Research Data, Ref. 897325, on 03 October 2022.

Acknowledgements

This project has been supported by AcqVA Aurora Research Center at UiT The Arctic University of Norway, under grant number 2062165.

Competing interests

The authors have no competing interests to declare.

Author contributions

Tekabe Legesse Feleke: conceptualization, data collection, methodology, statistical analyses, writing

Terje Lohndal: conceptualization, fund acquisition, methodology, review and editing

ORCiD IDs

Tekabe Legesse Feleke: 0000-0002-5627-2043

Terje Lohndal: 0000-0002-8514-1499

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. DOI: https://doi.org/10.1016/S0010-0277(99)00059-1

Andersson, A., Sayehli, S., & Gullberg, M. (2019). Language background affects online word order processing in a second language but not offline. *Bilingualism: Language and Cognition*, *22*(4), 802–825. DOI: https://doi.org/10.1017/S1366728918000573

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. DOI: https://doi.org/10.3758/s13428-019-01237-x

Aumeistere, A., Bultena, S., & Brouwer, S. (2022). Wisdom comes with age. The role of grammatical gender in predictive processing in Russian children and adults. *Applied Psycholinguistics*, *43*(4), 867–887. DOI: https://doi.org/10.1017/S0142716422000170

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Bañón, J. A., & Martin, C. (2021). The role of cross-linguistic differences in second language anticipatory processing: An event-related potentials study. *Neuropsychologia*, *155*(1), 1–18. DOI: https://doi.org/10.1016/j.neuropsychologia.2021.107797

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed effects models using lme 4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. DOI: https://doi.org/10.1016/j.tics.2022.09.015

Blumenfeld, H. K., Bobb, S. C., & Marian, V. (2016). The role of language proficiency, cognate status and word frequency in the assessment of Spanish–English bilinguals' verbal fluency. *International*

Journal of Speech-Language Pathology, *18*(2), 190–201. DOI: https://doi.org/10.3109/17549507.2015.1081288

Bobb, S. C., Kroll, J. F., & Jackson, C. N. (2015). Lexical constraints in second language learning: Evidence on grammatical gender in German. *Bilingualism: Language and Cognition*, *18*(3), 502–523. DOI: https://doi.org/10.1017/S1366728914000534

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*(3), 299–314. DOI: https://doi.org/10.1017/S1366728907003082

Bosch, J., & 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.

Bott, N. T., Lange, A., Rentz, D., Buffalo, E., Clopton, P., & Zola, S. (2017). Web camera-based eye tracking to assess visual memory on a visual paired comparison task. *Frontiers in Neuroscience*, *11*, 370. DOI: https://doi.org/10.3389/fnins.2017.00370

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 607–624. DOI: https://doi.org/10.3758/s13415-015-0340-0

Casillas, M., & Frank, M. C. (2013). The development of predictive processes in children's discourse understanding. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 299–304). Cognitive Society.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1029–1040. DOI: https://doi.org/10.1037/a0015901

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

Clamons, C. (1992). *Gender in Oromo*. PhD thesis, University of Minnesota, Minnesota, USA.

Clamons, C. (1993). Gender assignment in Oromo. In E. Mushira & I. Gregory (Eds.), *Principles and Prediction: The analysis of natural language* (pp. 269–286). Amsterdam: John Benjamins Publishing. DOI: https://doi.org/10.1075/cilt.98.22cla

Corbett, G. 1991. *Gender.* Cambridge: CUP. DOI: https://doi.org/10.1017/CBO9781139166119

Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*(1), 42–71. DOI: https://doi.org/10.1111/lang.12321

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480. DOI: https://doi.org/10.1006/jmla.1999.2688

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, *34*(2), 183–193. DOI: https://doi.org/10.1016/j.system.2006.01.005

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930. DOI: https://doi.org/10.1080/23273798.2019.1591469

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881–900.

Dunn, L., and Dunn, D. (2009). *The British picture vocabulary scale*, 3rd Edn. Brentford: GL Assessment.

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. DOI: https://doi.org/10.1017/S0272263112000915

Dyson, B. P. (2016). Variation, individual differences and second language processing: A Processability Theory study. *Linguistic Approaches to Bilingualism*, *6*(4), 341–395. DOI: https://doi.org/10.1075/lab.14007.dys

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, *27*(1), 59–82. DOI: https://doi.org/10.1177/0267658310386647

Feleke, T. L., Gooskens, C., & Rabanus, S. (2020). Mapping the dimensions of linguistic distance: A study on South Ethiosemitic languages. *Lingua*, *243*, 1–31. DOI: https://doi.org/10.1016/j.lingua.2020.102893

Feleke, T. L. (2021). Ethiosemitic languages: Classifications and classification determinants. *Ampersand*, *8*, 1–15. DOI: https://doi.org/10.1016/j.amper.2021.100074

Feleke, T. L. (2023). Issues in classifying and mapping the Semitic languages of Ethiopia. In S. D. Brunn & R. Kehrein (Eds.), *Language, society and the state in a changing world* (pp. 33–85). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-031-18146-7_2

Feleke, T. L., & Lohndal, T. (2023). Gender variation across the Oromo dialects: A corpus-based study. *Studia Linguistica*, *77*(3), 453–495. DOI: https://doi.org/10.1111/stul.12213

Feleke, T. L. (2024a). Dialect separation and cross-dialectal influences: A study on Oromo grammatical gender systems. *Linguistics*. DOI: https://doi.org/10.1515/ling-2022-0119

Feleke, T. L. (2024b). Dialect recognition via lexical processing: Is it a viable litmus test? *Languages*, *9*(6), 1–26. DOI: https://doi.org/10.3390/languages9060186

Foucart, A. (2021). Language prediction in second language. In E. Kan & T. Güter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 91–114). John Benjamins. DOI: https://doi.org/10.1075/bpa.12.05fou

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379–399. DOI: https://doi.org/10.1017/S136672891000012X

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), 1–10. DOI: https://doi.org/10.1037/a0036756

Fowler, J., & Jackson, C. N. (2017). Facilitating morphosyntactic and semantic prediction among second language speakers of German. *Journal of Cognitive Psychology*, *29*(8), 883–901. DOI: https://doi.org/10.1080/20445911.2017.1353517

Garrido-Pozú, J. J. (2022). Predictive processing of grammatical gender: Using gender cues to facilitate processing in Spanish. *Lingua*, *278*, 1–19. DOI: https://doi.org/10.1016/j.lingua.2022.103416

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. DOI: https://doi.org/10.1016/j.jml.2007.07.001

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411–1423. DOI: https://doi.org/10.1037/0278-7393.27.6.1411

Grüter, T., & Kaan, E. (2021). Prediction in Second Language Processing and Learning. Prediction in Second Language Processing and Learning. In E. Kan & T. Güter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 1–24). John Benjamins.

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. DOI: https://doi.org/10.1080/23273798.2019.1648840

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. DOI: https://doi.org/10.1177/0267658312437990

Grüter, T., & Rohde, H. (2013). *L2 processing is affected by RAGE: Evidence from reference resolution.* Talk presented at 12th conference on Generative Approaches to Second Language Acquisition (GASLA). University of Florida, FL.

Halberstadt, L., Valdés Kroff, J. R., & Dussias, P. E. (2018). Grammatical gender processing in L2 speakers of Spanish: The role of cognate status and gender transparency. *Journal of Second Language Studies*, *1*(1), 5–30. DOI: https://doi.org/10.1075/jsls.17023.hal

Hawkins, R. (2009). Statistical learning and innate knowledge in the development of second language proficiency: Evidence from the acquisition of gender concord. In A. G. Benati (ed.), *Issues in second language proficiency* (pp. 63–78). London: Continuum International Publishing.

Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, *118*(4), 595–612. DOI: https://doi.org/10.1016/j.lingua.2007.01.009

Hetzron, R. (1972). *Ethiopian Semitic: Studies in classification* (No. 1). Manchester: Manchester University Press.

Hockett, Charles F. 1958. *A course in modern linguistics.* New York: Macmillan.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. DOI: https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. In *BUCLD 36: Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 226–245). Cascadilla Press.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. DOI: https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. DOI: https://doi.org/10.1515/iral-2015-0014

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. DOI: https://doi.org/10.1177/0267658315624960

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5-27.

Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, *8*, 235–256. DOI: https://doi.org/10.1146/annurev-linguistics-030821-054113

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. DOI: https://doi.org/10.1017/S0272263116000437

Howard, Martin, 2011. Input perspectives on the role of learning context in second language acquisition. *International Review of Applied Linguistics*, *49*(2), 71–82. DOI: https://doi.org/10.1515/iral.2011.004

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. DOI: https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movement similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. DOI: https://doi.org/10.1017/S1366728917000050

Ito, A., & Knoeferle, P. (2023). Analyzing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. DOI: https://doi.org/10.3758/s13428-022-01969-3

Jakubowicz, C., Roulet, L., 2004. Narrow syntax or interface deficit? Gender agreement in French SLI. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The Role of Formal Features in Second Language Acquisition* (pp. 184–225). Lawrence Erlbaum Associates, Hillsdale. DOI: https://doi.org/10.4324/9781315085340-7

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

Kaan, E., Dallas, A. C., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In J.-W. Zwart, & M. de Vries (Eds.), *Structure preserved. Festschrift in the honor of Jan Koster* (pp. 207–213). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.164.23kaa

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, *19*(1), 1–18. DOI: https://doi.org/10.1017/S1366728914000844

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670. DOI: https://doi.org/10.1111/j.1749-818X.2008.00072.x

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. DOI: https://doi.org/10.1016/S0749-596X(03)00023-8

Kilgarriff, A., Marcowitz, F., Smith, S., & Thomas, J. (2015). Corpora and language learning with the Sketch Engine and SKELL. *Revue Française de Linguistique Appliquée*, *20*(1), 61–80. DOI: https://doi.org/10.3917/rfla.201.0061

Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language: A visual-word eye tracking study. *Studies in Second Language Acquisition*, *43*(1), 133–154. DOI: https://doi.org/10.1017/S0272263120000443

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, *5*(1), 115–132. DOI: https://doi.org/10.22599/jesla.79

Kramer, R. (2015). *The morphosyntax of gender* (Vol. 58). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199679935.001.0001

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. DOI: https://doi.org/10.1080/23273798.2015.1102299

Kutlu, E., & Hayes-Harb, R. (2023). Towards a just and equitable applied psycholinguistics. *Applied Psycholinguistics*, *44*(3), 293–300. DOI: https://doi.org/10.1017/S0142716423000280

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Lago, S., Mosca, M., & Stutter Garcia, A. (2021). The role of cross-linguistic influence in multilingual processing: Lexicon versus syntax. *Language Learning*, *71*(S1), 163–192. DOI: https://doi.org/10.1111/lang.12412

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition*, *26*(1), 21–45. DOI: https://doi.org/10.1080/10489223.2017.1391815

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. DOI: https://doi.org/10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. DOI: https://doi.org/10.1016/j.jml.2010.07.003

Lozano-Argüelles, C., Sagarra, N., & Casillas, J. V. (2020). Slowly but surely: Interpreting facilitates L2 morphological anticipation based on suprasegmental and segmental information. *Bilingualism: Language and Cognition*, *23*(4), 752–762. DOI: https://doi.org/10.1017/S1366728919000634

Luna, D., & Peracchio, L. A. (2002). "Where there is a will…": Motivation as a moderator of language processing by bilingual consumers. *Psychology & Marketing*, *19*(7-8), 573–593. DOI: https://doi.org/10.1002/mar.10026

Lundquist, B., Rodina, Y., Sekerina, I. A., & Westergaard, M. (2016). Gender change in Norwegian dialects: Comprehension is affected before production. *Linguistics Vanguard*, *2*(s1), 1–15. DOI: https://doi.org/10.1515/lingvan-2016-0026

Lundquist, B., & Vangsnes, Ø. A. (2018). Language separation in bidialectal speakers: Evidence from eye tracking. *Frontiers in Psychology*, *9*, 1–18. DOI: https://doi.org/10.3389/fpsyg.2018.01394

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. DOI: https://doi.org/10.1016/j.jneumeth.2007.03.024

Meir, N., Parshina, O., & Sekerina, I. A. (2020). The interaction of morphological cues in bilingual sentence processing: An eye-tracking study. In *Proceedings of the 44th annual Boston University Conference on Language Development* (pp. 376-389). Sommerville, MA: Cascadilla Press.

Meyer, M., Lamers, D., Kayhan, E., Hunnius, S., & Oostenveld, R. (2021). Enhancing reproducibility in developmental EEG research: BIDS, cluster-based permutation tests, and effect sizes. *Developmental Cognitive Neuroscience*, *52*, 1–10. DOI: https://doi.org/10.1016/j.dcn.2021.101036

Meyer, R. (2006). Amharic as lingua franca in Ethiopia. *Lissan: Journal of African Languages and Linguistics*, *20*(1/2), 117–132.

Mitsugi, S. (2020). 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. DOI: https://doi.org/10.1515/iral-2017-0118

Morales, L., Paolieri, D., Dussias, P. E., Kroff, J. R. V., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism: Language and Cognition*, *19*(2), 294–310. DOI: https://doi.org/10.1017/S1366728915000176

Niemi, J., & Laine, M. (1989). The English language bias in neurolinguistics: New languages give new perspectives. *Aphasiology*, *3*(2), 155–159. DOI: https://doi.org/10.1080/02687038908248984

Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *Quarterly Journal of Experimental Psychology*, *63*(8), 1631–1645. DOI: https://doi.org/10.1080/17470210903511210

Paolieri, D., Demestre, J., Guasch, M., Bajo, T., & Ferré, P. (2020). The gender congruency effect in Catalan–Spanish bilinguals: Behavioral and electrophysiological evidence. *Bilingualism: Language and Cognition*, *23*(5), 1045–1055. DOI: https://doi.org/10.1017/S1366728920000073

Paolieri, D., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2019). Gender congruency effects in Russian–Spanish and Italian–Spanish bilinguals: The role of language proximity and concreteness of words. *Bilingualism: Language and Cognition*, *22*(1), 112–129. DOI: https://doi.org/10.1017/S1366728917000591

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*(2), 103–133. DOI: https://doi.org/10.1191/026765800677556046

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. *Behav Res*. DOI: https://doi.org/10.3758/s13428-023-02176-4

R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430. DOI: https://doi.org/10.1177/0267658308090186

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism*, *10*(3), 257–272. DOI: https://doi.org/10.1017/S1366728907003069

Scheutz, M. J., & Eberhard, K. M. (2004). Effects of morphosyntactic gender features in bilingual language processing. *Cognitive Science*, *28*(4), 559–588. DOI: https://doi.org/10.1207/s15516709cog2804_3

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. DOI: https://doi.org/10.1017/S1366728922000736

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 79–95. DOI: https://doi.org/10.1515/iral-2016-9991

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. DOI: https://doi.org/10.1017/S1366728920000607

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 133–150. DOI: https://doi.org/10.1515/iral-2016-9994

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the visual world paradigm. *Glossa Psycholinguistics*, *1*(1), 1–37. DOI: https://doi.org/10.5070/G6011131

Weber, K., Lau, E. F., Stillerman, B., & Kuperberg, G. R. (2016). The yin and the yang of prediction: An fMRI study of semantic predictive processing. *PLoS One*, *11*(3), 0148637. DOI: https://doi.org/10.1371/journal.pone.0148637

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. DOI: https://doi.org/10.1017/S1930297500008512

Yimam, B. (1988). Focus in Oromo. *Studies in African Linguistics*, *19*(3), 365–384. DOI: https://doi.org/10.32473/sal.v19i3.107460