Bilingualism: Language and Cognition

cambridge.org/bil

Subject–verb agreement by Turkish speakers of English: Differential cue-weighting in L1 and L2 and distinct mechanisms for comprehension and production

Nazik Dinçtopal Deniz[1], Özge Bakay[2] and Cem Murat Deniz[3]

[1]Boğazici University, Department of Foreign Language Education, Türkiye; [2]University of Massachusetts Amherst, Department of Linguistics, USA and [3]New York University Grossman School of Medicine, Department of Radiology, USA

Research Article

Cite this article: Dinçtopal Deniz, N., Bakay, Ö. and Deniz, C.M. (2026). Subject–verb agreement by Turkish speakers of English: Differential cue-weighting in L1 and L2 and distinct mechanisms for comprehension and production. *Bilingualism: Language and Cognition* 1–12. https://doi.org/10.1017/S1366728926101497

Received: 23 July 2025 UTC
Revised: 15 May 2026 UTC
Accepted: 28 May 2026 UTC

Keywords:
second language processing; cue-retrieval; cue-weighting; syntax; semantics

Corresponding author:
Nazik Dinçtopal Deniz;
Email: nazik.dinctopal@bogazici.edu.tr

This research article was awarded Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Employing a self-paced reading experiment, a read-aloud production task and a pen-and-paper questionnaire, we examined if Turkish (L2) and native (L1) speakers of English rely on similar or different mechanisms in comprehending and producing English subject–verb number agreement. The results showed that the syntactic distance of an attractor to the head noun affected the L1 speakers' parsing decisions, while the L2 speakers were influenced by its linear proximity to the verb. The results also showed *singular attraction* in comprehension for both groups and (a tendency for) *plural attraction* in production. We argue that L1 and L2 speakers differ in the sources of attraction and in their cue weighting/encoding while processing and producing S-V agreement. Whereas syntactic proximity is more crucial in the L1 speakers' agreement computation, L2 speakers weigh linear distance cue more heavily. We also argue that S-V agreement computation is governed through different mechanisms in comprehension and production.

1. Introduction

Subject–verb number agreement (henceforth, S-V agreement) occurs when a verb and its subject agree in their number feature. S-V agreement errors occur when the verb fails to agree with its subject in number as in *The key to the cabinets were lost* (Bock & Miller, 1991, p. 56). Several first language (L1) studies have reported S-V agreement errors in production (e.g., Bock & Miller, 1991; Vigliocco et al., 1995) and insensitivity to such errors in comprehension (e.g., Pearlmutter et al., 1999; Wagers et al., 2009). They are reported to be even more common in the second language (L2; e.g., Jackson et al., 2018).

This paper examines S-V agreement in the L2 and investigates if the human sentence processing mechanism (henceforth, the parser) operates in an essentially similar fashion across the languages an individual speaks (Foote, 2011; Lago & Felser, 2018) or differently in the L1 and L2 due to representational (Clahsen & Felser, 2006) or parsing mechanism variances (Clahsen & Felser, 2018; Cunnings, 2017a). L2 speakers' (L2ers) use of syntactic cues during processing has been a major issue of debate: while some argue that syntax is underused in the L2 (Clahsen & Felser, 2006, 2018) or it is weighted less heavily than other (linguistic) cues (Cunnings, 2017a; Deniz, 2022), others claim that, like native speakers, L2ers weight syntactic cues more heavily than nonlinguistic information (e.g., Lago & Felser, 2018). We aim to contribute to this debate by investigating how L2ers use (non)linguistic cues in their sentence processing.

We also investigate, through the phenomenon of *(reversed) mismatch asymmetry* (Bock & Miller, 1991; Pearlmutter, 2000), if the computation of S-V agreement in L2 relies on similar mechanisms in the comprehension and production modalities. *Mismatch asymmetry* or *plural attraction* is the observation that agreement errors are more likely to occur when the attractor noun is plural than when it is singular (e.g., Bock & Miller, 1991; Pearlmutter, 2000; Vigliocco et al., 1995). This is often attributed to the markedness of plural nouns, resulting in an increased likelihood of interference while computing the number feature of the verb. However, some studies reported a *reversed mismatch asymmetry* or *singular attraction* in comprehension (Pearlmutter, 2000; Deniz et al., 2023), suggesting that agreement could be computed through different operations in comprehension and production (Deniz et al., 2023; Gillespie & Pearlmutter, 2011; Tanner et al., 2014a; cf., Badecker & Kuminiak, 2007; Franck et al., 2015; Pearlmutter et al., 1999; Wagers et al., 2009). To our knowledge, S-V agreement computation has not been compared across comprehension and production in the L2 although there is work showing L1 and L2 differences in production (Konopka et al., 2018) as well as comprehension (see above).

We tested L1ers and Turkish L2ers of English in their processing and production of S-V agreement in sentences with complex subject noun phrases (NPs) such as *the pond(s) near the*

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

trail(s) for the horse(s) and examined the role of syntactic embeddedness and linear distance in cue interference in agreement computation. In such sentences, the processing/production of agreement can be influenced by the closely preceding local noun, *horse(s)*, whose number feature may be easier to recall than that of the head noun, *pond(s)* (Bock & Miller, 1991; Quirk et al., 1972). It is also possible that the number features of the noun that is syntactically closer to the head, *trail(s)*, can affect agreement more than a syntactically distant noun, *horse(s)*, through what is called *feature percolation* as its features can more easily transmit to the head (e.g., Eberhard et al., 2005; Franck et al., 2002; Nicol et al., 1997; Pearlmutter, 2000).

S-V agreement has also been examined in relation to cue retrievals from memory (e.g., Wagers et al., 2009). The *cue-based memory retrieval model* maintains that linguistic dependencies are formed through an associative, content-addressable cue-based memory retrieval process (Lewis & Vasishth, 2005). It predicts for S-V agreement that upon processing the verb, retrieval cues such as [±plural] are generated and checked against the features of the subject. The target item with matching cues is then retrieved, but this process can be subject to similarity-based interference from partially matching cues of intervening items. Interference is predicted to be *inhibitory* (i.e., increase processing difficulty) for grammatical sentences because the intervening noun(s)' features may interfere with retrieving the target from memory (van Dyke, 2007). It is predicted to be *facilitatory* (i.e., ease comprehension) for ungrammatical sentences because then an attractor noun with a mismatching number feature may be erroneously retrieved instead of the target.[1] Some argue that (e.g., Wagers et al., 2009) the cue-based retrieval mechanism is employed only during reanalysis in comprehension of ungrammatical sentences because, there, the verb's actual features may conflict with its predicted features and the parser can deploy cue-based retrievals to "recheck" the features during the first pass.

These approaches to S-V agreement computation have also been extended to L2 research. Studies have examined the role of linear distance and structural embeddedness within the context of gender agreement (e.g., Keating, 2009) under the *Shallow Structure Hypothesis* (SSH, Clahsen & Felser, 2006) and syntactic/linear cue-weighting under the cue-based memory retrieval model (e.g., Cunnings, 2017a; Serin Demirler & Deniz, 2019).

The SSH maintains that L2 sentence processing is mainly shallow as L2 speakers cannot compute complex syntax due to lack of syntactic detail in their representations (Clahsen & Felser, 2006). Although its revised version (Clahsen & Felser, 2018) is less restrictive in its predictions for syntactic representations in L2 and attributes the use of linguistic information to weighted constraints, the model still predicts underuse of syntactic information in L2 sentence processing. This results in increased reliance on nonstructural, semantic or pragmatic cues (e.g., Keating, 2009; but cf., Hopp, 2010). The SSH does not make specific predictions for the L2 processing/production of S-V agreement. But it is possible to predict under an SSH view that L2 speakers would not show sensitivity to syntactic embeddedness of attractor nouns. This could result in linear proximity to the verb to be more influential than syntactic proximity to the head noun or no reliable difference between the two.

Cunnings (2017a) argues that instead of referring to across-the-board sensitivity or insensitivity to syntactic information (or capacity differences), L2 processing can be explained with reference to cue-based memory retrievals. Accordingly, L2 processing involves syntactically detailed parses but is more susceptible to interference than L1 processing due, partially, to differential memory retrieval implementations. L1 speakers rely primarily on syntactic cues in resolving dependencies, but L2 speakers may underweight them due to their more abstract nature (Cunnings, 2017b). For L2 computation of agreement, Cunnings (2017a) argues that as the distance between the subject head and the verb increases, the number of attractor nouns may increase, and the activation level of the cues associated with the head may decay. This may cause the cues from the linearly closer noun to have a higher activation and stronger impact on agreement interpretation.

We aim to contribute to these arguments with data from Turkish speakers of English. We also examine if they can be extended to production in a parallel fashion.

1.1. The production and comprehension of S-V agreement

Originally proposed for S-V agreement production, the *Marking and Morphing* (MM) model (Eberhard et al., 2005) attributes agreement errors to faulty or ambiguous mental representation of the subject's number feature. Singular and plural values for count nouns are 0 and 1, and the head and attractor nouns' number features both contribute to the number representation of a complex subject. Plural number is grammatically and morphologically marked, while singular is an unmarked/default value; hence, only plural nouns affect the head noun's number feature. The probability of making an agreement error in production is predicted to increase in presence of a plural distractor in subject NP, and as such, the model predicts plural (but not singular) attraction. Syntactically embedded attractor nouns also have a decreased weight in the summation of number, affecting agreement to a lesser extent than less embedded attractors.

Pearlmutter et al. (1999) and Staub (2009) extend the MM model to S-V agreement comprehension. Pearlmutter et al. (1999) argue that the comprehension of agreement, like its production, involves a *head-overwriting* mechanism in which the head NP's number feature is overwritten by the local NP's feature as the subject's number feature is *computed on the fly*. Staub (2009) similarly argues that when the number feature of a complex NP becomes more ambiguous (i.e., in the presence of number-mismatching attractor nouns), decision times slowdown in proportion to the strength of attraction.

Badecker and Kuminiak (2007) and Wagers et al. (2009) also argue that similar mechanisms govern the comprehension and production of agreement, but they maintain that it is due to cue-based memory retrievals. As in comprehension (see above), the production of the verb's number feature requires the retrieval of the subject's number, which may be susceptible to attraction from the intervening nouns between the head noun and the verb.

Nicol et al. (1997) and Tanner et al. (2014a), however, argue that comprehension and production of S-V agreement are different. Nicol et al. argue that a *feature percolation* approach can explain its production but not comprehension because comprehension is incremental, and there is no input for the verb as the NPs are processed. That is, "the nature of computation of [S-V] agreement in production is "forward-specifying," while in comprehension, it is "backward-checking"" (p. 586).

Tanner et al. (2014a) argue that although agreement production can be explained with the MM model, agreement comprehension

[1]Some predict such illusions only in ungrammatical sentences (Wagers et al., 2009), but they have also been observed in grammatical sentences (Franck et al., 2015; Nicol et al., 1997; Pearlmutter, 2000; also see Hammerly et al., 2019 and Jäger et al., 2017 for a review).

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

must be explained through cue-based memory retrievals because Tanner et al.'s ERP and judgment data showed differential attraction in grammatical and ungrammatical sentences. They observed reduced P600 responses, indicating syntactic reanalysis, to plural attractors for ungrammatical but not grammatical verbs although there was some interference in judgment data. Tanner et al. (pp. 206–207) took these to indicate that "initial parsing of verb number relies on a combination of predictive mechanisms in tandem with cue-based retrieval mechanisms, which are susceptible to similarity-based interference, while later processes are more sensitive to the global representation of subject NP number, consistent with the MM model," and such an approach can indeed "reconcile conflicting findings which have shown no interference in grammatical sentences using on-line measures at the verb (Wagers et al., 2009), but reliable interference in off-line, global measures of sentence comprehension (Nicol et al., 1997)."

Deniz et al. (2023) examined production-comprehension differences through *(reversed) mismatch asymmetry*. They investigated if it was a consequence of different comprehension and production processes or due to design differences in previous comprehension and production studies. Using the same materials in production and comprehension experiments and controlling for working memory (WM) demands in production, they showed plural attraction in production and singular attraction in comprehension. They concluded that the mismatch asymmetry in production and the reversed mismatch asymmetry in comprehension were genuine effects attributable to the mechanisms employed in production and comprehension of S-V agreement but not to the methodological discrepancies in previous work.

Although several studies investigated agreement computation in production or comprehension, limited work examined it in comprehension and production in the same study (cf., Deniz et al., 2023; Staub, 2009). The results of such work are mixed, but the emerging findings suggest that plural attraction in production and singular attraction in comprehension may be due to the differential operations of the parser across the two modalities. As far as we are concerned, production/comprehension asymmetry has not been examined in the L2. We contextualize this relatively new question in a study that also tests the more widely discussed questions on the operations of the L2 parser (e.g., Clahsen & Felser, 2018; Cunnings, 2017a).

1.2. Subject–verb agreement in the second language

S-V agreement has been examined from the perspectives of heuristic or combinatorial parsing in the L2 (Tanner et al., 2014b), competence and performance deficit (Foote, 2011; Jiang, 2004), syntactic/semantic number agreement (e.g., Jackson et al., 2018) and differential cue-weighting in the L2 (Armstrong et al., 2018; Lago & Felser, 2018; Serin Demirler & Deniz, 2019). Age of arrival, proficiency and WM were also investigated. For conciseness, we will only provide an overview of work that is closely related to the questions addressed in the present study (i.e., excluding work such as Jackson et al., 2018).

Tanner et al. (2014b) examined S-V agreement in English by Spanish learners of English in an ERP experiment. The data for (un) grammatical stimuli such as *The winner of the big trophy has/have proud parents* showed that a participant's dominant ERP responses (N400 or P600) varied as a function of certain individual differences. Although proficiency was overall positively correlated with sensitivity to agreement violations, the participants with earlier AoA and higher motivation were more likely to show P600 responses to agreement violations and those with later AoA and lower motivation were more likely to show N400 responses. Tanner et al. argue that experience with L2 can produce within-learner shifts from a more (semantic/lexical) memory-based processing routine, reflected as N400 response, to more combinatorial processing, reflected as P600 response.

L2ers' difficulty with S-V agreement has also been examined in relation to competence/performance deficit, a context of debate parallel to the SSH (Clahsen & Felser, 2006). The competence deficit view attributes difficulty with S-V agreement morphology to its incomplete acquisition and deficiencies at the representation level (Johnson & Newport, 1989). The performance deficit view attributes them to problems at the performance level, to processes such as accessing, retrieving, or controlling information that has been internalized (Sorace, 1985).

Jiang (2004) tested Chinese L2ers of English for their sensitivity to number agreement in (un)grammatical sentences in this context. The L2ers in the study did not show sensitivity to number mismatches in self-paced reading experiments, but they did so in an untimed written task. Jiang concluded that the results indicated deficiency at the competence level because if they were due to performance deficiency, then, sensitivity would be shown in both forms of comprehension. The performance on the written task was taken to show explicit knowledge of S-V agreement, not integrated at the competence level.

Foote (2011) tested early and late L1 English-L2 Spanish bilinguals whose L1 and L2 were comparable in morphological marking of plurality but not gender to examine if implicit knowledge could be selective based on the L1 transferability of the structure. Like native speakers, both bilingual groups were sensitive to both types of agreement, but this was more pronounced when the head noun was linearly closer to the verb than when it was distant. The results suggested integrated knowledge of agreement regardless of age of acquisition and agreement type as well as effects of linear distance between the head and the verb.

Sagarra and Herschensohn (2010) examined L2 number and gender agreement in the context of representational accessibility or deficit of uninterpretable (u) features in the L2 which are absent in the L1. Universal Grammar (UG) to L2 would predict availability of *u* features that check the interpretable gender [+/− feminine] and number [+/− plural] through Agree operations (Chomsky, 2000), whereas deficit approaches would not when the L2 is acquired after the critical period (Franceschina, 2001). The results of a self-paced reading experiment and a grammaticality judgment task showed that intermediate (but not beginner) L1 English-L2 Spanish learners were sensitive to both gender and number agreement and sensitivity to gender agreement was more evident with high WM capacity. Intermediate learners' sensitivity to semantic (gender) and syntactic (number) agreement violations was taken to reject a strict shallow parsing view in the L2 (Clahsen & Felser, 2006), and gender agreement was interpreted to be cognitively more taxing.

Lago and Felser (2018) investigated if linguistic (syntactic structure) and nonlinguistic (linear distance) constraints are weighted differentially in the L2. L1 German and L1 Russian-L2 German speakers chose a singular or plural verb, in two experiments, for preambles like *the smell of the stable(s) of/and the farmer(s)*. Response times by both groups of participants were affected by the number feature of the NP2, which was more evident for the embedded (*of-of*) structures than the coordinated (*of-and*) structures. The results were taken to support similar S-V agreement processing in the L1 and L2 and prioritization of linguistic cues over linear distance.

Armstrong et al. (2018) examined, in an ERP study, if quantificational cues to number would affect S-V agreement processing by

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

Mandarin speakers. Their test sentences included quantified/unquantified subjects such as *Most/The cookies taste/tastes*… Although Chinese lacks overt number agreement marking, it marks plurality on NPs through quantification. Armstrong et al. examined if the additional cue (*most*) to number would improve Mandarin learners' sensitivity to S-V agreement in English. The acceptability judgment data showed no difference between quantified and unquantified sentences. The ERP data showed sensitivity (a P600 effect) to agreement violations by both groups. Although native speakers showed increased sensitivity to agreement violations in quantified sentences, Mandarin speakers' sensitivity was reduced. The authors argued that when an additional cue to agreement was present, the L2 group ignored the morpho-syntactic cue to agreement and followed an L1-like strategy. When the quantificational cue was absent, they relied on the only source of agreement, i.e., morpho-syntax.

Serin Demirler and Deniz (2019) examined lexical and morpho-syntactic cues in S-V agreement in English with L1 Turkish-L2 English speakers. Experimental sentences included complex subject NPs with three nouns with(out) the singular quantifier *one*, for example, *the daughters of the/one author(s) of the/one book(s)*, testing the role of syntactic/linear distance in L2 processing of S-V agreement and the predictions of cue-retrieval from memory. Their eye-tracking data showed that L2ers were sensitive to linear distance; sensitivity to syntactic distance was evident only when the syntactically closer noun's number feature was marked by lexical information as in *the daughters of one author of the books.* Serin Demirler and Deniz concluded that L2ers can make syntactically detailed computations like L1ers, but they do so through lexical cues.

The present study also examines Turkish and English as a language pair (see below for motivations) but examines syntactic and linearity cues as well as cue-weighting across comprehension and production modalities.

2. The present study

Turkish and English are different in terms of their number marking and morphological richness. Although otherwise morphologically poor, English overtly marks subject–verb number agreement. Although morphologically rich, Turkish does not always overtly mark plurality on verbs, but does so on nouns (e.g., *kadın(−lar) gel-di "woman(-PL) come-PAST"*). The verbal suffix for the third person singular is null in Turkish, and the third person plural marking on the verb depends on the subject's being overt or null (i.e., *pro*) and its animacy, definiteness and specificity. For all other persons, the verb is morphologically marked (Göksel & Kerslake, 2005).

Previous studies have reported that in strict word-order languages with poor-inflection and overt subjects (e.g., English), agreement is computed through a feature-copying mechanism (e.g., Nicol et al., 2001) as it would rely more on syntactic agreement. In morphologically rich languages with null subjects (e.g., Spanish), agreement would be realized through semantic feature-unification between the subject and the verb (Franck et al., 2002; Nicol et al., 2001). Inclusion of Turkish and English as a language pair provides an interesting test case because although Turkish, a *pro*-drop language, is morphologically rich, it is not necessarily so for third-person S-V agreement marking (i.e., our test conditions). Although English is morphologically poor, it marks S-V agreement in more contexts for third person than Turkish. Given this, it is not clear if Turkish speakers would entertain a feature-copying mechanism in their L2 English (given that third-person S-V agreement can be ambiguously marked in Turkish) or a feature-unification mechanism (given the morphological richness in their L1 and an otherwise strong agreement for all other persons). They can be less sensitive than English speakers in S-V agreement marking because of restricted S-V agreement marking for 3rd person in their L1. Alternatively, they can show sensitivity to 3rd person agreement because of the rich inflectional system in their L1.

Two groups participated in the study: L1 Turkish-L2 English speakers with advanced proficiency in English and English native speakers as control. Each group took part in a word-by-word self-paced reading experiment, a read-aloud production experiment and a pen-and-paper sentence completion task in the order given here with breaks in between. The participants also took part in an operational span task to test their WM capacity (Sanchez et al., 2010), but since it was not predictive of the data in any of the experiments, presumably due to ceiling effects, especially for the L2 group (M_{L2_WM} = 43), its results are not reported.

2.1. Participants

Fifty-three L1-Turkish speakers with advanced proficiency in English (20 females, M_{Age} = 30) and 39 L1-English speakers (15 females, M_{Age} = 31.5) participated in all the experiments. One L2er had limited daily L2 use as she resided in Turkey and was excluded from analyses. All other participants (except for three L2ers from Toronto) lived in New York City at the time of data collection.[2]

The participants' (language) demographics are summarized in Table 1.[3]

The participants received $10 for compensation. All reported (corrected-to-)normal vision.

2.2. Experiment 1: Self-paced reading

Experiment 1 used a self-paced reading (SPR) task to examine if L1ers and Turkish L2ers of English followed similar strategies in processing English S-V agreement.

2.2.1. Materials

The materials in Experiment 1 (and in Experiments 2, 3) were adapted from Franck et al.'s (2002) production study in English. They included complex subject NPs with three nouns as in (1). All the conditions included grammatical sentences, that is, the verbs and subjects agreed in number (like in Franck et al., 2015; Hammerly et al., 2019; Serin Demirler & Deniz, 2019). There were eight conditions, manipulating the number feature (singular vs. plural) of the head noun and feature (mis)match between the head and the attractors (match vs. syntactic, linear or double mismatch). The head was either singular (S) as in (1a–d) or plural (P) as in (1e–f). The number feature of the intermediate and local nouns either matched that of the head as in (1a, e) or mismatched it as in (1b–d) and (1f–h). The N2 is syntactically closer to the head noun and N3 is linearly closer to the verb.

(1) a. SSS: The e-mail from the secretary of the manager was confidential.
b. SSP: The e-mail from the secretary of the managers was confidential.
c. SPS: The e-mail from the secretaries of the manager was confidential.

[2]The native speaker data are published elsewhere (Deniz, et al., 2023) and were used as control in the present study.

[3]Reproduced from Table 3, (Deniz, 2022, p. 719).

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

Table 1. (Language) demographics for English as L1 and L2 groups

		POB	Mean age	Gender	Highest degree
Demographics	L1	USA (*N* = 38) Canada (*N* = 1) S. Korea (*N* = 1)	31.5	*N* of females = 15	High school (*N* = 5) College (*N* = 17) Master's (*N* = 16) PhD (*N* = 2)
	L2	Turkey (*N* = 50)	30	*N* of females = 20	College (*N* = 26) Master's (*N* = 16) PhD (*N* = 8)

	Mean AOE	Mean AOA	Mean LOS	School language	SPR	Score in TOEFL	Weekly Turkish use (in hours)	Weekly English use (in hours)
Language demographics for L2	11.4	24.7	5.3	Elementary school: Turkish (*N* = 50) Secondary school: Turkish (*N* = 33) English (*N* = 15) German (*N* = 1) French (*N* = 1) College: Turkish (*N* = 5) English (*N* = 45) Graduate school: Turkish (*N* = 2) French (*N* = 1) English (*N* = 41)	Advanced (*N* = 50)	IBT > 90 CBT > 225 PBT > 554	Home *M* = 19.2 Social *M* = 6.5 Work *M* = 0.69	Home *M* = 3.45 Social *M* = 6.4 Work *M* = 24.11

Note: POB = place of birth; AOE = age of exposure to English (in years); AOA = age of arrival to an English-speaking country (in years); LOS = length of stay in an English-speaking country (in years); SRP = self-reported English proficiency.

d. SPP: The e-mail from the secretaries of the managers was confidential.
e. PPP: The e-mails from the secretaries of the managers were confidential.
f. PPS: The e-mails from the secretaries of the manager were confidential.
g. PSP: The e-mails from the secretary of the managers were confidential.
h. PSS: The e-mails from the secretary of the manager were confidential.

The conditions with number match between the head and attractor nouns (SSS, PPP) were predicted to be processed the fastest (Eberhard et al., 2005; cf., Wagers et al., 2009). For the mismatch conditions, there is converging evidence for L1 for interference from the syntactically closer noun to the head (e.g., Eberhard et al., 2005; Pearlmutter et al., 1999). For L2 processing of (S-V) agreement, studies have shown either linear (e.g., Foote, 2011; Keating, 2009; Serin Demirler & Deniz, 2019) or syntactic distance effects (Lago & Felser, 2018).

Under an SSH view, L2ers are predicted to underuse syntactic information (Clahsen & Felser, 2006, 2018). Similarly, L2ers are predicted to give more weight to linearity cues (Cunnings, 2017a). Given activation-based decay in cue retrievals from memory, the feature mismatch between N3 and the head noun as in (1b,f) can delay the processing of the verb more than the mismatch between N2 and the head noun as in (1c, g). But if the L2ers show sensitivity to N2's number, that would indicate use of syntactic cues in the L2 like L1 (Lago & Felser, 2018).

The number feature of the head noun was manipulated to test the (reversed) mismatch asymmetry for S-V agreement in the L2. Mismatch asymmetry (plural attraction) is widely reported for L1 in production (e.g., Bock & Miller, 1991). In comprehension, there is emerging evidence for it to be reversed (i.e., singular attraction; Deniz et al., 2023; Pearlmutter, 2000). The present study tested if the relatively less reported *singular attraction* in comprehension is also generalizable to L2 processing of S-V agreement. If so, this would reveal as sensitivity to the conditions with plural heads and singular attractors as in (1f–h) but not for singular heads and plural attractors as in (1c–d).

There were 32 experimental sentence sets distributed across eight lists counterbalancing for head noun number and feature (mis)match between the head noun and attractors. Each list contained 32 experimental sentences, intermingled with 64 filler sentences involving structural dependencies or ambiguous constructions, six practice and six implicit warm-up items, totaling up to 108 sentences. A *yes/no* comprehension question followed each sentence.

2.2.2. Procedure

Experiment 1 (and Experiment 2) was conducted using DMDX, version 4.0.4.8 (Forster & Forster, 2003). Sentences were presented word-by-word in a non-cumulative, SPR paradigm. The task lasted 20–25 minutes.

2.2.3. Data analysis

The data were analyzed using the *R* statistical computing software (R Core Team, 2019) with linear mixed effects models (Baayen et al., 2008) using the *lmer* function of the *lme4* package (Bates et al., 2015). Model building proceeded incrementally: we first fitted baseline models with single fixed-effect predictors and then added additional predictors and their interactions. Random-effects structures were evaluated in parallel; all models included random intercepts only, as including random slopes did not improve model

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

fit, χ^2's ≤ 5.95, *p*'s ≥ .202. Each successive model was compared to the simpler nested model using likelihood-ratio tests. The *p*-values were calculated using Satterthwaite approximations for degrees of freedom (Luke, 2017) in the *lmerTest* package (Kuznetsova et al., 2017).

All participants had ≥75% accuracy (*M* = 91.4%), confirming comprehension. Trials with incorrect responses (4.4%) were eliminated before the analyses. The analyses were run on RTs of the critical region (the 9th word, the main verb) and its spillover region (the 10th word), which were log-transformed before the analyses (Ratcliff, 1993). For ease of interpretation, we report the back-transformed (exponentiated) model estimates. Fixed effects were group (L1, L2), the head noun number (singular, plural) and feature (mis)match between the head noun and attractors (match, N2, N3 and double mismatch); random effects were subjects and items.

The best-fitting models were inspected for residual distribution (Baayen & Milin, 2010) and, through *influence.ME*, for influential data (Nieuwenhuis et al., 2012). The final reported models have standardized residuals below/above 2 standard deviations (in the main analyses for both groups and subanalyses for the L1ers) or 3 standard deviations (in the subanalyses for the L2ers) removed (Baayen & Milin, 2010) and influential data (Nieuwenhuis et al., 2012), if any, excluded, corresponding to 2.3%–7.4% of data. (See the analysis code for details.)

2.2.4. Results

In both regions, the L2ers had longer RTs than L1ers, critical region: β = 80.11, *SE* = 34.29, *t* = 2.52, *p* = .013; spillover region: β = 130.79, *SE* = 62.73, *t* = 2.92, *p* = .024. The results are reported separately for each region below.

Critical region. Figure 1 shows the data for the critical region.

In the critical region, there was a marginal interaction between group, head noun and feature match, β = 91.15, *SE* = 52.61, *t* = 1.89, *p* = .058. The model including an interaction of the three predictors explained the data better than the simpler models with an interaction of head noun and feature match, $\chi^2(8) = 17.29$, *p* = .027, and group and feature match, $\chi^2(8) = 18.73$, *p* = .016. We thus examined the L1 and L2 data separately.

The L1 data showed an interaction of plural-head and feature mismatch of N2, β = 37.27, *SE* = 19.69, *t* = 1.97, *p* = .049, and the model with the interaction of head noun and feature match explained the data marginally better than the simpler model with feature match, $\chi^2(4) = 8.64$, *p* = .071. Thus, the data were examined separately for each head noun condition. The match (SSS and PPP) conditions were taken as the baseline in these analyses. For singular heads, there was no difference between the SSS and any of the mismatch conditions, β's ≤ 12.81, *SE*'s ≤ 13.65, *t*'s ≤ .97, *p*'s ≥ .33. For plural heads, the PSP condition took longer to read than the PPP condition, β = 27.69, *SE* = 14.24, *t* = 2.00, *p* = .046; there were no significant differences between the PPP and other mismatch conditions, β's ≤ 15.5, *SE*'s ≤ 14.10, *t*'s ≤ 1.12, *p*'s ≥ .264.

In the L2 data, there was a significant interaction between plural-head and feature mismatch of N3, β = 92.22, *SE* = 37.83, *t* = 2.64, *p* = .008. The models with the two predictors interacting explained the data better than each simple model with feature match or head noun as the predictor, χ^2's(4/6) ≥ 10.82, *p*'s ≤ .034. The data were thus inspected for each head noun condition. For singular heads, there was no significant difference between the SSS and any of the mismatch conditions, β's ≤ 26.47, *SE*'s ≤ 18.54, *t*'s ≤ 1.46, *p*'s ≥ .143. For plural heads, the PPS took longer to read than the PPP condition, β = 43.32, *SE* = 20.33, *t* = 2.22, *p* = .027; there was no difference between PPP and other mismatch conditions, β's ≤ 17.31, *SE*'s ≤ 19.29, *t*'s ≤ .91, *p*'s ≥ .361.

Spillover region. Figure 2 shows the data for the spillover region.

There was a main effect of group, where the L2ers were slower than L1ers, β = 148.76, *SE* = 65.84, *t* = 2.52, *p* = .013, but there was no interaction between any predictors, β's ≤ 19.56, *SE*'s ≤ 21.83, *t*'s ≤ .88, *p*'s ≥ .378. Since the model with an interaction of group and head noun explained the data better than the simple model with only head noun or group as a predictor, $\chi^2(2) = 7.23$, *p* = .027, each group's data were analyzed separately. The L1 data showed no main effect or interaction between the predictors, β's ≤ 14.29, *SE*'s ≤ 17.39, *t*'s ≤ 0.83, *p*'s ≥ .405. The L2 data showed a marginal

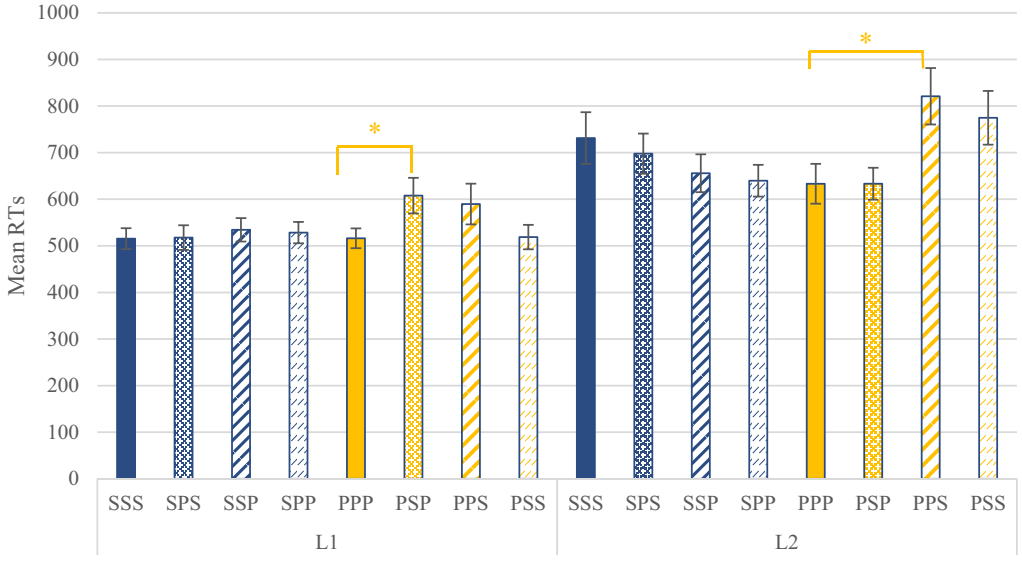


Figure 1. Mean RTs, with standard errors, for the critical region, by group and condition. "*" denotes significance at the alpha level .05 or smaller.

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

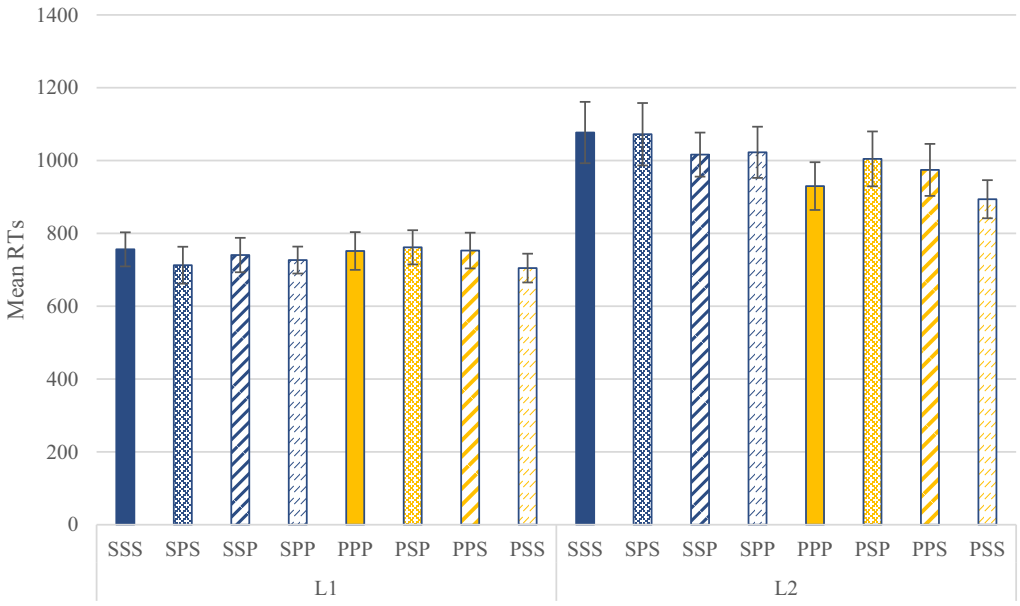


Figure 2. Mean RTs, with standard errors, for the spillover region, by group and condition.

effect of head noun where plural-head conditions were read faster than singular-head conditions, β = 29.02, SE = 15.82, t = 1.8, p = .073. There was no other main effect or interaction, β's ≤ 24.97, SE's ≤ 23.49, t's ≤ 1.08, p's ≥ .28.

2.2.5. Discussion

The results showed a syntactic distance effect for the L1ers and a linear distance effect for the L2ers in the plural-head singular-attractor conditions. The L2ers' sensitivity to linear distance, as opposed to the L1ers' sensitivity to syntactic distance, can either indicate that they underused syntactic information (Clahsen & Felser, 2006, 2018) or were influenced by recency and weighted the linear distance cue more heavily than the syntactic cue for the verb's number feature (Cunnings, 2017a). We will address these possibilities in detail in General Discussion.

The mismatch effects were observed only in plural head conditions, extending the observation of *singular attraction* in L1 comprehension (Deniz et al., 2023; Pearlmutter, 2000) to L2. Pearlmutter argued, for his findings, that when there are multiple intervening nouns (between the verb and the subject head) whose number features must all be computed, the markedness of the plural head weakens. This increases its vulnerability to attraction. The production data will reveal if this is a true *reversed mismatch asymmetry*.

2.3. Experiment 2: Production

Previous research showed plural attraction in S-V agreement production (e.g., Bock & Miller, 1991) but singular attraction for its comprehension (e.g., Pearlmutter, 2000). The SPR experiment in this study confirmed singular attraction in comprehension for both groups. To examine if the usual plural attraction is observed in production, and if it can be extended to L2, Experiment 2 included a production task.

2.3.1. Materials

The experimental materials were the same as those in Experiment 1 except that the copula verb was replaced by a length-neutral underscore.

It was predicted that S-V agreement production would be the most accurate in the baseline (SSS, PPP) conditions. For the mismatch conditions, given the results of Experiment 1, it is reasonable to predict a syntactic distance (i.e., N2 mismatch) effect for the L1ers and a linear distance (N3 mismatch) effect for the L2ers, with the reservation that production and comprehension behavior can differ.

If the widely reported plural attraction for production in L1 extends to L2, the L2ers (and the L1ers) will make more errors in singular-head plural-distractor conditions.

The number of experimental, filler, practice, warm-up sentences and reading lists was the same as in Experiment 1.

2.3.2. Procedure

The participants were asked to read the sentences out loud as soon as they appeared on the screen (without preview) and complete those that had a missing component (e.g., the copula verb) as they read them. They moved to the next sentence with a key-press.

2.3.3. Data analysis

Data from two L2 participants were missing in Experiment 2. In some proportion of the sentences (L1: 1.12%; L2: 2.94%), the participants re-read the sentence and changed their insertions to a(n) (un)grammatical form of the verb. Sentence completions were thus categorized as *first* and *revised reading*, and the data were analyzed separately for each pass.

The data were analyzed with logistic regression mixed effects models using the *glmer* function in the *lme4* package (Bates et al., 2015). The fixed-effects structure was identical to that used in Experiment 1. Subject was included as a random effect; models that additionally specified items as random effects did not converge and resulted in singular fits and were not further considered. As the response variable was binary, the models were only inspected for influential data (Nieuwenhuis et al., 2012). (See the Supplementary Analysis code for details.)

2.3.4. Results

Accuracy in sentence completions was high in both singular-head (first reading: M_{L1} = 85.90%, M_{L2} = 90.38%; revised reading:

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

Table 2. Mean percent accuracies and standard errors (in parentheses) of L1 and L2 speakers in the first and revised readings in the production task

Conditions	First readings		Revised readings	
	L1	L2	L1	L2
SSS	96.2 (1.54)	99 (0.71)	95.5 (1.66)	99 (0.71)
SSP	91 (2.30)	89.5 (2.17)	91 (2.3)	92 (1.92)
SPS	84.6 (2.9)	91.5 (1.98)	84.6 (2.9)	92.5 (1.87)
SPP	71.8 (3.61)	81.5 (2.75)	73.1 (3.56)	85.5 (2.5)
PPP	85.3 (2.85)	93.5 (1.75)	85.9 (2.8)	94 (1.68)
PPS	82.1 (3.08)	83 (2.66)	83.3 (2.99)	88 (2.30)
PSP	80.8 (3.17)	89.5 (2.17)	81.4 (3.12)	94 (1.68)
PSS	78.2 (3.32)	80.5 (2.81)	80.1 (3.21)	84.5 (2.57)

M_{L1} = 86.06%, M_{L2} = 92.25%) and plural-head conditions (first reading: M_{L1} = 81.57%, M_{L2} = 86.63%; revised reading: M_{L1} = 82.69%, M_{L2} = 90.13%). The L1 group's accuracy did not change after revisions, $\beta = .06$, $SE = .12$, $z = .51$, $p = .609$, while the L2 group's accuracy improved, $\beta = .33$, $SE = .12$, $z = 2.70$, $p = .007$. Table 2 shows mean percent accuracy in both readings.

The accuracy results for the first and revised readings were, in general, similar. We thus report the results for the first reading as they would better reflect initial analyses.

The model including an interaction of the three predictors showed a significant main effect for group, where the L2ers had better accuracy than L1ers, $\beta = 1.14$, $SE = .47$, $z = 2.4$, $p = .016$, $OR = 3.13$; a main effect of head noun, with higher accuracy for singular heads, $\beta = 2.04$, $SE = .55$, $z = 3.67$, $p < .001$; $OR = 7.69$, a marginally significant interaction between group and double mismatch, $\beta = -.95$, $SE = .50$, $z = -1.88$, $p = .059$; $OR = .359$, and an interaction between head noun and feature match, $\beta = -2.48$, $SE = .63$, $z = -3.94$, $p < .001$; $OR = .73$. This model also explained the data better than the simpler models with an interaction of group and head noun, $\chi^2(12) = 127.85$, $p < .001$, and group and feature match, $\chi^2(8) = 39.95$, $p < .001$. We thus examined the L1 and L2 data separately.

In the L1 data, there were significant interactions between head noun number and feature mismatch of N2, $\beta = -1.76$, $SE = .70$, $z = -2.49$, $p = .013$; $OR = .17$, and between head noun number and double mismatch, $\beta = -2.21$, $SE = .69$, $z = -3.20$, $p = .001$; $OR = .11$; there was a main effect of head noun number, with higher accuracy for singular heads, $\beta = 1.69$, $SE = .61$, $z = 2.79$, $p = .005$; $OR = 5.42$; and an effect of double mismatch, $\beta = -.73$, $SE = .38$, $z = -1.93$, $p = .053$; $OR = .408$. The models with the two predictors interacting explained the data better than each simple model with feature match or head noun as the predictor, χ^2's(4) ≥ 26.98, p's ≤ .001. The data were thus examined separately for singular- and plural-heads. For singular heads, SPS and SPP conditions had lower accuracy than the baseline SSS: SPS, $\beta = -2.18$, $SE = .61$, $z = -3.57$, $p < .001$; $OR = .11$; SPP, $\beta = -3.06$, $SE = .61$, $z = -5.05$, $p < .001$; $OR = .05$. There was no number mismatch effect for plural heads, β's ≤ −.2, SE's ≤ .39, z's ≤ −0.53, p's ≥ .102.

In the L2 data, there were significant interactions between head noun number and feature mismatch of N2, $\beta = -1.66$, $SE = .85$, $z = -1.94$, $p = .052$; $OR = 0.19$, and between head noun number and double mismatch, $\beta = -1.67$, $SE = .82$, $z = -2.04$, $p = .041$; $OR = .19$; there were main effects of head noun number, $\beta = 1.62$, $SE = .77$, $z = 2.09$, $p = .036$; $OR = 5.05$, feature mismatch of N3, $\beta = -1.32$, $SE = .389$ $z = -3.39$, $p < .001$; $OR = .27$, and double mismatch, $\beta = -1.54$, $SE = .38$, $z = -4.02$, $p < .001$, $OR = .22$, with higher accuracies in singular head and baseline conditions. The models with the two predictors interacting explained the data better than each simple model with feature match or head noun as the predictor, χ^2's(4) ≥ 13.71, p's ≤ .008. The data were split by head noun number. For singular head nouns, all the mismatch conditions had lower accuracy than the baseline, β's ≤ −2.32, SE's ≤ .77, z's ≤ −3.02, p's ≤ .002, OR ≤ .1. For plural heads, the PPS and PSS conditions had lower accuracy than the baseline: PPS, $\beta = -1.32$, $SE = .38$, $z = -3.45$, $p < .001$; $OR = .27$; PSS, $\beta = -1.54$, $SE = .38$, $z = -4.06$, $p < .001$; $OR = .22$.

2.3.5. Discussion

The production data showed, for singular head nouns, an effect of syntactic distance and double mismatch for the L1ers, whereas all mismatch conditions had lower accuracy for the L2ers. There was no effect of number mismatch for plural heads for the L1ers. The L2ers' accuracy was lower for PPS and PSS, favoring a linear mismatch effect.

The results confirm plural attraction in production for L1ers, which was mainly triggered by syntactic distance. Although both singular and plural attractors affected S-V agreement for the L2ers, plural attractors did so in all mismatch conditions, but singular attraction was observed for NPs linearly closer to the verb. These results suggest that production is more effortful than comprehension in general and more so for L2ers. Although there is a tendency toward it, we cannot claim that reversed mismatch asymmetry truly holds for L2. (See General Discussion for further details.)

For the L1ers, the syntactic proximity of the intervening NPs to the head affected S-V agreement errors. (We assume that the effect observed in double mismatch conditions was triggered by that of NP2.) For the L2ers, syntactic effects appeared only in singular-head conditions. For plural heads, L2ers' decisions were affected more by linear distance. (We assume that the effect observed in double mismatch conditions was triggered by that of NP3.) We entertain, in interim, that these differences can be attributable to the markedness and encoding of the head noun such that the more marked form's encoding requires more cognitive resources, making linear distance effects more prevalent for plural-head conditions for the L2ers. When the head noun is singular, the L2ers behave similarly to the L1ers presumably because their cognitive resources are relatively less challenged in such conditions. The overall reduced accuracy in plural-head conditions also supports this observation.

2.4. Pen-and-paper questionnaire

The last experiment, a pen-and-paper questionnaire, was conducted to examine the L2ers' knowledge of English S-V agreement rules. But the data were also analyzed to examine the L1 and L2ers' off-line sensitivity to attraction in S-V agreement.

2.4.1. Materials

The materials were the same as in Experiment 2. The task was to choose the singular or plural form of the verb, given as options below the sentence, to complete the sentence.

The number of the experimental sentences and lists was the same as in previous experiments. There were 32 filler sentences including ambiguous constructions. There were no practice or warm-up sentences. Each list had 64 sentences.

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

Table 3. Means and standard errors (in parentheses) of L1 and L2 speakers' accuracies in the pen-and-paper task

Conditions	L1	L2	Conditions	L1	L2
SSS	96.8 (1.41)	99.5 (0.53)	PPP	94.8 (1.77)	99.5 (0.53)
SSP	96.8 (1.41)	100 (0)	PPS	94.2 (1.87)	97.9 (1.06)
SPS	89.1 (2.51)	97.9 (1.06)	PSP	96.1 (1.54)	98.9 (0.76)
SPP	92.9 (2.06)	98.4 (0.92)	PSS	94.9 (1.77)	98.4 (0.92)

2.4.2. Data analysis

Data from five L2 participants were missing in Experiment 3. The analyses were run on accuracy in sentence completions. The model building steps and fixed and random effects were the same as in previous experiments.

2.4.3. Results

Accuracy was high for both groups ($M_{overall}$ = 96.8%; singular head noun: M_{L1} = 93.9%, M_{L2} = 98.9%; plural head noun: M_{L1} = 95%, M_{L2} = 98.7%). Table 3 summarizes the data.

The analyses showed that the L1ers had lower accuracy in the SPS condition than the baseline SSS, $\beta = -2.14$, $SE = .74$, $z = -2.91$, $p = .003$; $OR = 0.12$. There was no effect of number mismatch for any of the other comparisons for either group, β's ≤ 1.42, SE's ≤ 1.13, z's ≤ 1.25, p's $\geq .21$.

2.4.4. Discussion

The L1 data showed syntactic distance effects, confirming the results in Experiments 1 and 2, and for only singular-head conditions, as in Experiment 2, confirming plural attraction in offline measures.

The high accuracy rates (≥98.8) by the L2ers show that they knew S-V agreement rules in English, confirming that the effects observed in Experiments 1 and 2 were genuine processing effects, comparable to those by the L1ers. Unlike the L1ers, the L2ers' accuracy was not affected by the mismatching number features of the intervening nouns. This is likely because when there is no time pressure, highly proficient L2ers with substantial experience with their L2 are better able to control their L2 and resist interference from attractor nouns (Lee & Phillips, 2023). It has also been reported that language behavior in untimed L2 pen-and-paper tasks can be affected by prescriptive/declarative knowledge (Deniz, 2022).

3. General discussion

Table 4 summarizes the results.

The results showed, for the L1ers, singular attraction in the SPR experiment and plural attraction in the production and pen-and-paper tasks. Syntactic proximity was a determining factor in the L1ers' processing and production behavior. The L2ers also showed singular attraction in the SPR task, but their decisions were affected by the attractor's linear proximity to the verb. Although the production experiment showed both singular and plural attraction and both linear and syntactic proximity effects, plural attraction and linear proximity were more prominent. Following Tanner et al. (2014a) and Deniz et al. (2023), we argue that the data indicate different mechanisms for S-V agreement across the production and comprehension modalities, not only for L1 but also for L2. We also argue that L1 and L2ers differ in the sources of attraction and in their cue weighting/encoding in processing and producing S-V agreement.

Table 4. The conditions showing agreement attraction in the SPR, production and pen-and-paper tasks for L1 and L2 speakers

	SPR	Production	Pen-and-paper
L1	PSP	SPS, SPP	SPS
L2	PPS	SPS, SSP, SPP, PPS, PSS	–

Let us first examine the production and comprehension differences. It has been argued for L1 (Nicol et al., 1997; Tanner et al., 2014a) that agreement is computed through different mechanisms in production and comprehension. We tested if this was also true for L2ers, through *(reversed) mismatch asymmetry* which refers to the widely reported *plural attraction* in production (e.g., Bock & Miller, 1991) to be reversed in comprehension and observed as *singular attraction* (Deniz et al., 2023; Pearlmutter, 2000). Like L1ers, the L2ers in the present study showed *singular attraction* in comprehension. Although their production data showed number attraction for both singular- and plural-head conditions, attraction was evident in all mismatch conditions in singular-head conditions but was triggered by the linearly closer attractors in plural-head conditions. This can be interpreted as a tendency for *plural attraction* in production. Both the L1 and L2 data showed attraction in more conditions in production than comprehension.

Following Tanner et al. (2014a), Nicol et al. (1997) and Deniz et al. (2023), we argue that in the L2, like L1, speakers employ different mechanisms for producing and comprehending S-V agreement. Tanner et al. maintain that S-V agreement comprehension can be explained through cue-based memory retrievals, but its production can be explained under the MM model (Eberhard et al., 2005). They argue that the mechanisms responsible for attraction in comprehension are subsets of those operative in production, and S-V agreement differences between comprehension and production are attributable to the differences between comprehension and production in general. That is, the intended message in production is known to the speaker in advance, including information about the number features of the NPs to be uttered. This results in "more avenues for interference in subject–verb agreement production than in subject–verb agreement comprehension" (Tanner et al., 2014a, p. 211). We synthesize Tanner et al.'s arguments with Gillespie and Pearlmutter's (2011) who attribute agreement errors in production to advance planning and with Pearlmutter's (2000) who attribute singular attraction in comprehension to cue strength.

Gillespie and Pearlmutter argue for production that "[i]f the scope of advance planning at a given level is large, multiple items are likely to be simultaneously available, which increases the chance of interference and certain speech errors (Garrett, 1975)" (p. 1615). That is, nouns planned overlappingly with the head noun are more likely to interfere with verb's agreement marking. In the case of plural attraction in production (as opposed to singular attraction in comprehension), it is possible that planning for intervening NPs that have a marked [+plural] feature requires more cognitive resources due to enhanced memory encoding. Hence, their planning would be more likely to overlap with that of the head noun. This is also in line with MM's predictions for plural attraction in production (Eberhard et al., 2005). It appears that the L1ers planned N1 and N2 together (N2 attraction in the L1 data), but the L2ers most likely planned all three NPs together since their data showed both N2 and N3 attraction. This would be supported by the L2ers' high WM scores (M_{L2_WM} = 43).

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press

Comprehension, however, is incremental, "restricting the window of temporal co-activation of nouns" (Tanner et al., 2014a; p. 211). That is, the input is processed as it is received, with multiple levels of prediction for the upcoming items (e.g., Levy, 2008). For singular attraction in comprehension, we concur with Pearlmutter (2000) and maintain that the plural feature of the head NP in existence of singular attractors weakens in a relatively long complex subject with two intervening NPs. Thus, although processing is incremental and speakers are subject to cue-retrievals from memory in their comprehension of S-V agreement, production involves planning for multiple units in advance. Thus, not only the type (singular vs. plural) but also the extent (number of conditions) of attraction is different for comprehension and production.

The increased proneness to error in production is presumably related to the computational burdens of planning and producing an utterance compared to comprehension (MacDonald, 2013). Considering the reduced saliency of singular heads, it is likely that any distracting plural NP attracted the verb's agreement. For plural heads though, the linearly closer singular NPs attracted the verb's agreement in the L2 group. Given our participants' high accuracy in the pen-and-paper task, we do not think that the source of L2ers' increased proneness to error in S-V agreement is attributable to their knowledge of L2. They were highly advanced in their L2 and frequently used English daily (see Table 1). Their accuracy was also significantly higher than the L1ers. We thus think that they were like L1ers, even more vigorous, in their command of English S-V agreement rules. The increased likelihood of errors by the L2ers in production is more likely to be due to performance requirements such as limited attentional resources as some of the resources need to be allocated to controlling activation levels of their L1 and L2 (Green, 2000), retrieval of lexical elements from the relevant language (de Bot, 2000) and attending to competition between L1 and L2 items for selection (e.g., Poulisse & Bongaerts, 1994).

We now turn to the sources of attraction in the L1 and L2 processing of S-V agreement. The SPR data show a clear distinction for the L1 and L2ers. Although the L1ers were mainly affected by the attractor's syntactic proximity to the subject head (attraction in the PSP condition), the L2ers were affected by its linear distance to the verb (attraction in the PPS condition). Both the L1 and L2ers were affected by the syntactic and linear proximity of the attractors in their production of S-V agreement, but the L2ers were influenced by linear proximity in more conditions than by syntactic proximity. These data suggest either that (i) the L2ers underused syntactic information as predicted by the SSH (Clahsen & Felser, 2018) or (ii) they were more susceptible to interference in memory retrieval/gave heavier weight to linear proximity than syntactic information (Cunnings, 2017a). We believe the latter approach explains the comprehension data better especially when the differences between comprehension and production modalities are taken into consideration. The cue-based memory retrieval approach to L2 processing does not predict L2ers to underuse syntactic information, but it predicts differential weightings of cues in the L1 and L2 due to increased susceptibility to interference in memory retrievals in the L2 (Cunnings, 2017a). Cunnings (2017a) argues for linearity effects in L2 agreement that "slower reading speed in the L2 may lead to enhanced effects of activation-based decay, which could lead to delayed or less robust retrieval of target items, and thus less robust violation effects as dependency length increases" (p. 667). This would be supported by Kaan et al.'s (2015) data which showed similar parsing patterns for L1 and L2ers when they were matched in reading speed. Following Kaan et al., we calculated each participant's average reading speed across all nonfinal words in correctly-answered filler items (about 513 data points per participant). Our L2ers' RTs (M_{L2_RT} = 721.54) were overall slower than those of the L1ers (M_{L1_RT} = 551.29), t = 3.81, p < .001, in line with Cunnings' arguments for increased susceptibility to interference and enhanced activation-based decay. It is also possible that the L2 group weighted the linear proximity cue more heavily than the syntactic cue to agreement.

Grüter et al. (2020) argue that language learners are optimally adaptive and L2 processing, in general, is more effortful (Sorace, 2011). If linear proximity was taken to be less effortful and more effective for agreement dependency in the L2, that cue could have been weighted more heavily than the syntactic cue. This would be in line with Deniz's (2022) argument that differential cue-weighing in the L2 may be due to L2 sentence processing being less modular than L1 sentence processing especially when the L2 is acquired after certain cognitive developments have taken place (which is the case with our participants).

Note that our results contradict with Lago and Felser's (2018) findings on number agreement but are in line with other research on number/gender agreement that showed effects of linearity in the L2 (Foote, 2011; Keating, 2009). Thus, although our findings concur with the growing body of L2 research showing differential cue weighting/retrieval interference in the L2 (e.g., Deniz, 2022; Grüter et al., 2020; Serin Demirler & Deniz, 2019; Tanner et al., 2014b), this conclusion has the reservation that more work is needed for conclusive remarks.

The language pairing in the present study also informs on L1 effects in L2 processing. Recall that Turkish does not overtly distinguish between third-person singular and plural marking on the verb, but its inflectional morphology is otherwise rich. The Turkish participants were sensitive to S-V agreement in English, suggesting that the lack of overt S-V agreement for the given context did not hinder their sensitivity to S-V agreement in English. This was, perhaps, due to the otherwise rich inflectional system in Turkish. Our data cannot speak to whether this was due to feature unification from their L1 or to their ability to feature-copy in the L2. Future work can investigate that in a design that better controls for factors contributing to these processes.

It is also noteworthy that the L2ers were more accurate in their second responses compared to their first responses in the production task. This was not the case for the L1ers. This suggests that L2ers monitor their production more than L1ers, and despite their advanced proficiency, their language acquisition has not ceased.

Finally, for comprehension, the cue-based retrieval model (Lewis & Vasishth, 2005) does not always predict attraction in grammatical sentences (Wagers et al., 2009). But our data, for both groups, showed attraction effects in S-V agreement in grammatical sentences, which is in line with a large body of work (e.g., Franck et al., 2015, Hammerly et al., 2019; Pearlmutter, 2000). This would be predicted in the cue-based memory retrieval account (Lewis & Vasishth, 2005) if mis-retrievals of feature-matching attractors are assumed (Jäger et al., 2015).

Replication package. The analysis codes, reports with model outputs and study materials, are available at https://osf.io/2e9az/overview?view_only=419ab0715b2f444aaca7819fb0beea2f.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728926101497.

Acknowledgments. We would like to thank Didem Kurt for her help with annotations on the production data. This work was supported by Boğaziçi University Scientific Research Projects Grant [#13781] to the first author. Partial financial support was received from the City University of New York, Graduate Center, Student Research fund.

Competing interests. The authors declare none.

Data availability statement. Because the study's IRB approval and the informed consent forms at the time of data collection ensured the participants that the data would be accessible only to the research team, the data are not provided.

References

Armstrong, A., **Bulkes, N.**, & **Tanner, D.** (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers: An ERP study. *Studies in Second Language Acquisition*, **40**(4), 731–754. https://doi.org/10.1017/S0272263118000013

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. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & **Milin, P.** (2010). Analyzing reaction times. *International Journal of Psychological Research*, **3**(2), 12–28. https://doi.org/10.21500/20112084.807

Badecker, W., & **Kuminiak, F.** (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, **56**(1), 65–85. https://doi.org/10.1016/j.jml.2006.08.004

Bates, D., **Maechler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bock, K., & **Miller, C. A.** (1991). Broken agreement. *Cognitive Psychology*, **23**(1), 45–93. https://doi.org/10.1016/0010-0285(91)90003-7

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by Dtep: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–155). MIT Press.

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

Clahsen, H., & **Felser, C.** (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, **40**(3), 693–706. https://doi.org/10.1017/S0272263117000250

Cunnings, I. (2017a). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, **20**(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I. (2017b). Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition*, **20**(4), 712–721. https://doi.org/10.1017/S1366728916001243

De Bot, K. (2000). Simultaneous interpreting as language production. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives* (pp. 65–88). John Benjamins. https://doi.org/10.1075/btl.40.06bot

Deniz, N. D. (2022). Processing syntactic and semantic information in the L2: Evidence for differential cue-weighting in the L1 and L2. *Bilingualism: Language and Cognition*, **25**(5), 713–725. https://doi.org/10.1017/S1366728921001140

Deniz, N. D., **Bakay, Ö.**, & **Kurt, D. B.** (2023). (reversed) mismatch asymmetry in English subject-verb agreement. *Dilbilim Araştırmaları Dergisi*, **34**(1), 1–26. https://doi.org/10.18492/dad.1181172

Eberhard, K. M., **Cutting, J. C.**, & **Bock, K.** (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, **112**(3), 531–559. https://doi.org/10.1037/0033-295X.112.3.531

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, **32**(1), 187–220. https://doi.org/10.1017/S0142716410000342

Forster, K. I., & **Forster, J. C.** (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, **35**(1), 116–124. https://doi.org/10.3758/BF03195503

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, **17**(3), 213–247. https://doi.org/10.1177/026765830101700301

Franck, J., **Colonna, S.**, & **Rizzi, L.** (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology*, **6**, 349. https://doi.org/10.3389/fpsyg.2015.00349

Franck, J., **Vigliocco, G.**, & **Nicol, J.** (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, **17**(4), 371–404. https://doi.org/10.1080/01690960143000254

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. **9**, p. 133–177). New York: Academic Press.

Gillespie, M., & **Pearlmutter, N. J.** (2011). Hierarchy and scope of planning in subject-verb agreement production. *Cognition*, **118**(3), 377–397. https://doi.org/10.1016/j.cognition.2010.10.008

Göksel, A., & **Kerslake, C.** (2005). *Turkish: A comprehensive grammar*. Routledge.

Green, D. W. (2000). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. In L. Wei (Ed.), *The bilingualism reader* (pp. 371–383). New York.

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

Hammerly, C., **Staub, A.**, & **Dillon, B.** (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, **110**, 70–104. https://doi.org/10.1016/j.cogpsych.2019.01.001

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

Jackson, C. N., **Mormer, E.**, & **Brehm, L.** (2018). The production of subject-verb agreement among Swedish and Chinese second language speakers of English. *Studies in Second Language Acquisition*, **40**(4), 907–921. https://doi.org/10.1017/S0272263118000025

Jäger, L. A., **Engelmann, F.**, & **Vasishth, S.** (2015). Retrieval interference in reflexive processing: Experimental evidence from mandarin, and computational modeling. *Frontiers in Psychology*, **6**, 125783. https://doi.org/10.3389/fpsyg.2015.00617

Jäger, L. A., **Engelmann, F.**, & **Vasishth, S.** (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, **94**, 316–339. https://doi.org/10.1016/j.jml.2017.01.004.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, **25**(4), 603–634. https://doi.org/10.1017/S0142716404001298

Johnson, J. S., & **Newport, E. L.** (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, **21**(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Kaan, E., **Ballantyne, J. C.**, & **Wijnen, F.** (2015). Effects of reading speed on second-language sentence processing. *Applied Psycholinguistics*, **36**(4), 799–830. https://doi.org/10.1017/S0142716413000519

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, **59**(3), 503–535. https://doi.org/10.1111/j.1467-9922.2009.00516.x

Konopka, A. E., **Meyer, A.**, & **Forest, T. A.** (2018). Planning to speak in L1 and L2. *Cognitive Psychology*, **102**, 72–104. https://doi.org/10.1016/j.cogpsych.2017.12.003

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. https://doi.org/10.18637/jss.v082.i13

Lago, S., & **Felser, C.** (2018). Agreement attraction in native and nonnative speakers of German. *Applied Psycholinguistics*, **39**(3), 619–647. https://doi.org/10.1017/S0142716417000601

Lee, E. K. R., & **Phillips, C.** (2023). Why non-native speakers sometimes outperform native speakers in agreement processing. *Bilingualism: Language and Cognition*, **26**(1), 152–164. https://doi.org/10.1017/S1366728922000414

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, **106**(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, R. L., & **Vasishth, S.** (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, **29**(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, **49**(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, **4**, 226. https://doi.org/10.3389/fpsyg.2013.00226

Nicol, J. L., **Forster, K. I.**, & **Veres, C.** (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, **36**(4), 569–587. https://doi.org/10.1006/jmla.1996.2497

Nicol, J. L., **Teller, M.**, & **Greth, D.** (2001). Production of verb agreement in monolingual, bilingual and second language speakers. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 117–133). Blackwell Publishers.

Nieuwenhuis, R., **Grotenhuis, M. T.**, & **Pelzer, B.** (2012). Influence.ME: Tools for detecting influential data in mixed effects models. *The R Journal*, **4**(2), 38–47. https://doi.org/10.32614/RJ-2012-011

Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research*, **29**(1), 89–98. https://doi.org/10.1023/A:1005128624716

Pearlmutter, N. J., **Garnsey, S. M.**, & **Bock, K.** (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, **41**(3), 427–456. https://doi.org/10.1006/jmla.1999.2653

Poulisse, N., & **Bongaerts, T.** (1994). First language use in second language production. *Applied Linguistics*, **15**(1), 36–57. https://doi.org/10.1093/applin/15.1.36

Quirk, R., **Greenbaum, S.**, **Leech, G. N.**, & **Svartvik, J.** (1972). *A grammar of contemporary English (vol. 1985)*. Longman.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Sagarra, N., & **Herschensohn, J.** (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, **120**(8), 2022–2039. https://doi.org/10.1016/j.lingua.2010.02.004

Sanchez, C. A., **Wiley, J.**, **Miura, T. K.**, **Colflesh, G. J.**, **Ricks, T. R.**, **Jensen, M. S.**, & **Conway, A. R.** (2010). Assessing working memory capacity in a non-native language. *Learning and Individual Differences*, **20**(5), 488–493. https://doi.org/10.1016/j.lindif.2010.04.001

Serin Demirler, H., & **Deniz, N. D.** (2019, August 28–21). *Linear and syntactic distance in S-V agreement in the L2* [poster session]. *The 29th conference of the European second language association (EuroSLA 29)*, Lund University, Lund, Sweden. https://www.humlab.lu.se/fileadmin/user_upload/humlab/EuroSLA_Book_of_Abstracts.pdf

Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition-poor environments. *Applied Linguistics*, **6**(3), 239–254. https://doi.org/10.1093/applin/6.3.239

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, **1**(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, **60**(2), 308–327. https://doi.org/10.1016/j.jml.2008.11.002

Tanner, D., **Inoue, K.**, & **Osterhout, L.** (2014b). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, **17**(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., **Nicol, J.**, & **Brehm, L.** (2014a). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, **76**, 195–215. https://doi.org/10.1016/j.jml.2014.07.003

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**(2), 407–430.

Vigliocco, G., **Butterworth, B.**, & **Semenza, C.** (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, **34**(2), 186–215. https://doi.org/10.1006/jmla.1995.1009

Wagers, M. W., **Lau, E. F.**, & **Phillips, C.** (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, **61**(2), 206–237. https://doi.org/10.1016/j.jml.2009.04.002

https://doi.org/10.1017/S1366728926101497 Published online by Cambridge University Press