Journal of Memory and Language 111 (2020) 104083

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Effects of chronological age on native and nonnative sentence processing: Evidence from subject-verb agreement in German

Jana Reifegerste[a,b,*], Rebecca Jarvis[a,c], Claudia Felser[a]

[a] Potsdam Research Institute for Multilingualism, University of Potsdam, Potsdam, Germany
[b] Department of Neuroscience, Georgetown University, Washington, DC, USA
[c] Department of Linguistics, Harvard University, Cambridge, MA, USA

ARTICLE INFO

Keywords:
Sentence processing
Subject-verb agreement
Attraction errors
Second-language processing
Aging

ABSTRACT

While much attention has been devoted to the cognition of aging multilingual individuals, little is known about how age affects their grammatical processing. We assessed subject-verb number-agreement processing in sixty native (L1) and sixty non-native (L2) speakers of German (age: 18–84) using a binary-choice sentence-completion task, along with various individual-differences tests. Our results revealed differential effects of age on L1 and L2 speakers' accuracy and reaction times (RTs). L1 speakers' RTs increased with age, and they became more susceptible to attraction errors. In contrast, L2 speakers' RTs decreased, once age-related slowing was controlled for, and their overall accuracy increased. We interpret this as resulting from increased L2 exposure. Moreover, L2 speakers' accuracy/RT patterns were more strongly affected by cognitive variables (working memory, interference control) than L1 speakers'. Our findings show that as regards bilinguals' grammatical processing ability, aging is associated with both gains (in experience) and losses (in cognitive abilities).

Introduction

The past ten years have seen a steep rise in the amount of research in the field of bilingualism and aging. This is unsurprising given that we are living in a society that is becoming both older (Rechel et al., 2013) and increasingly multilingual (Baker & Prys Jones, 1998). However, the vast majority of studies focusing on aging and bilingualism have examined the state of cognition in older bilinguals, and whether or not speaking more than one language confers benefits on cognitive functioning in aging, such as improved executive control and attention (Bak, Vega-Mendoza, & Sorace, 2014; Vega-Mendoza, West, Sorace, & Bak, 2015) or a delayed onset of dementia (Bialystok, Craik, & Freedman, 2007; Woumans et al., 2015). Relatively little attention has been paid to older bilinguals' processing of language, and we still know very little about how aging affects bilinguals' language skills.

Here we examine native (L1) and non-native (L2) sentence-level processing cross-sectionally across the lifespan using a binary-choice sentence-completion task. In the following, we will first provide an overview of existing research on language processing in older bilinguals and then turn to the grammatical phenomenon under investigation – agreement attraction – and a more in-depth discussion of previous research on this widely-studied phenomenon, before laying out the design of the present study and our hypotheses.

Language in older bilinguals

Most studies to date on language processing in older bilinguals have focused on the lexicon. From this literature, it appears that aging has similar effects on L1 and L2 speakers' lexical processing as assessed by tasks such as verbal fluency, lexical decision, or naming from pictures or definitions (Bialystok, Craik, & Luk, 2008; Gollan, Montoya, Cera, & Sandoval, 2008; Johns, Sheppard, Jones, & Taler, 2016). As regards grammar, we are aware of only a small number of studies examining morphological processing in older bilinguals. These studies found little to no effect of age on combinatorial operations involved in the processing of regular inflections, which are often viewed as the result of grammatical rule application: Although reaction times (RTs) and accuracy rates increased with increasing age for both L1 and L2 speakers, the respective morphological priming patterns for L1 and L2 speakers remained the same across adulthood (Clahsen & Reifegerste, 2017; Reifegerste, Elin, & Clahsen, 2019). Interestingly, we are aware of one study suggesting that grammatical processing in older L2 speakers may require greater executive control than L1 processing. Using fMRI, Prehn, Taud, Reifegerste, Clahsen, and Flöel (2018) found greater

* Corresponding author at: University of Potsdam, Potsdam Research Institute for Multilingualism, Haus 2, Campus Golm, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany.
E-mail address: jana.reifegerste@gmail.com (J. Reifegerste).

https://doi.org/10.1016/j.jml.2019.104083
Received 8 February 2019; Received in revised form 1 December 2019; Accepted 5 December 2019

activation of the bilateral medial superior frontal gyrus when older L2 speakers responded to incorrectly (vs. correctly) inflected German participle forms, as compared to age-matched L1 speakers. This brain area has been associated with executive functioning in general (du Boisgueheneuc et al., 2006; Schel et al., 2014), and language control in particular (Abutalebi & Green, 2016; Green & Abutalebi, 2013), suggesting that grammatical processing in a late-learned L2 might require greater involvement of cognitive resources compared to L1 processing.

Gaps in previous studies have precluded a deeper understanding of grammatical processing in aging L2 speakers. First, while there are now a handful of studies on L2 morphological processing in aging, we are not aware of any published studies on L2 sentence-level processing in aging – even though previous studies with L1 speakers have indicated that the developmental trajectories of regular morphological and syntactic processing may differ. While regular morphology appears to show very few effects of age, sentence-level processing appears to be more sensitive to effects of aging (Kemper, 1986, 1988; Kemtes & Kemper, 1997; Kynette & Kemper, 1986; Light & Capps, 1986; Obler, Fein, Nicholas, & Albert, 1991; Reifegerste & Felser, 2017; Yoon et al., 2015). These different degrees of vulnerability during L1 sentence-level processing in older adults suggest that findings from morphology may not necessarily extend to syntax.

Second, most of the aforementioned studies that investigated morphological processing in older L2 speakers did not assess the involvement of age-sensitive cognitive resources, such as executive functioning. The brain-imaging study by Prehn et al. (2018), which found a correlation between letter fluency and amount of activation in superior frontal gyrus across language groups, did not have a young control group. Moreover, while there is currently no published evidence that age-sensitive cognitive resources such as working memory or inhibitory control are involved in L1 or L2 morphological processing (cf. Dronjic, 2013; Reifegerste, 2014), there is increasing evidence for the involvement of executive control in syntactic processing, both in aging and in bilinguals (Novick, Trueswell, & Thompson-Schill, 1998; Reifegerste, Hauer, & Felser, 2017; Veenstra, Antoniou, Katsos, & Kissine, 2017; but see Waters & Caplan, 2001). These findings render sentence-level processing a particularly good test case for investigating the involvement of cognitive abilities in native and non-native grammatical processing across the lifespan, including the question of whether declines in such abilities affect L1 and L2 speakers in similar ways.

In the present study, we aim to address these gaps by examining sentence processing in native and non-native speakers of German across the adult lifespan, while also measuring executive control and other cognitive functions within-subjects.

Agreement attraction

The grammatical phenomenon of interest in the present study is subject-verb agreement, that is, the morphosyntactic feature match between two structurally dependent sentence constituents. This type of agreement occurs in the majority of the world's languages (Mallinson & Blake, 1981); it is acquired relatively early during L1 acquisition (around age 2 in German; Poeppel & Wexler, 1993); and its computation in an L1 is relatively error-free, given the large number of occurrences (Bock, 2004). Nonetheless, even native speakers make agreement errors from time to time, among which so-called agreement attraction errors are particularly common.

Agreement attraction was first described by Bock and Miller (1991), who found that participants produced a disproportionally high number of incorrect plural verb forms when producing a sentence in which the singular subject noun phrase (NP) was modified by a following plural NP (e.g., **The key to the cabinets are*…). Similarly, in comprehension paradigms (e.g., self-paced reading, eye-tracking), participants have been found to be more likely to fail to notice a mismatch in features between the subject NP and the verb when the two were separated by an intervening NP that matched the verb in number (Pearlmutter, Garnsey, & Bock, 1999). Agreement attraction has since been observed in many languages (Arabic: Tucker, Idrissi, & Almeida, 2015; Dutch: Hartsuiker, Antón-Méndez, & Van Zee, 2001; Hartsuiker, Schriefers, Bock, & Kikstra, 2003; French: Vigliocco, Hartsuiker, Jarema, & Kolk, 1996; Hebrew: Deutsch & Dank, 2009; Italian: Vigliocco, Butterworth, & Semenza, 1995; Russian: Lorimor, Bock, Zalkind, Sheyman, & Beard, 2008; Serbian: Ristić, Molinaro, & Mancini, 2016; Slovak: Badecker & Kuminiak, 2007; Spanish: Vigliocco, Butterworth, & Garrett, 1996; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Turkish: Lago et al., 2019) – including German, the language under study here (Hartsuiker et al., 2003; Häussler, 2012). These studies have helped to further characterize and refine the circumstances under which agreement attraction typically occurs.

Two broad classes of accounts have been posited to explain these phenomena: misrepresentation accounts and cue-based retrieval accounts. Early accounts have attributed agreement attraction to the misrepresentation of the subject head. The earliest of these misrepresentation models postulated that agreement attraction comes about as a consequence of hierarchical feature movement or “percolation” (Franck, Vigliocco, & Nicol, 2002; Nicol, 1995; Nicol, Forster, & Veres, 1997; Vigliocco & Nicol, 1998). Under such accounts, morphosyntactic feature information (e.g., number) percolates upwards from the attractor to the subject NP and may “overwrite” the number feature of the subject head (Pearlmutter et al., 1999). Findings of nonlinear attraction (e.g., Wagers, Lau, & Phillips, 2009), however, are difficult to explain under a percolation account. Indeed, percolation accounts have by now been all but abandoned, but they gave rise to another type of misrepresentation model, the Marking and Morphing model (Bock, 2004; Eberhard, Cutting, & Bock, 2005). Under this account, the number information of the subject NP (or any NP, for that matter) can range from unambiguously singular to unambiguously plural. The presence of a plural attractor inside a singular subject NP can shift the value for the subject NP from unambiguously singular to more ambiguous by spreading activation, which may then in turn lead to choosing the wrong verb form.

The other family of accounts – cue-based retrieval accounts – assumes that both the subject head and the attractor NP are simultaneously activated during sentence planning and comprehension. Attraction errors then emerge as a consequence of faulty retrieval of the correct NP and its number feature from memory when the verb is processed or selected for production (Badecker & Kuminiak, 2007; Slevc & Martin, 2016; Wagers et al., 2009). Within this view, we can distinguish between direct-access models (McElree, 2000; McElree, Foraker, & Dyer, 2003) and activation-based models of sentence processing (Lewis & Vasishth, 2005; see also Tucker et al., 2015, for discussion). While both types of models are based on the assumption that agreement attraction is the consequence of retrieving the wrong NP from memory, they differ in the assumed consequences that the strength of the representation in memory has on the retrieval process. Whereas in direct-access models retrieval speed is always the same but the accuracy of retrieval can differ, in activation-based models both speed and accuracy of the retrieval process may vary.

Singular-plural asymmetry

One of the key findings from several agreement attraction studies is that attraction errors are much more common in sentences with singular subject heads and plural attractors (**The key to the cabinets are*) than in sentences with plural subject heads and singular attractors (**The keys to the cabinet is*) (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Bock, Nicol, & Cutting, 1999; Eberhard, 1997; Haskell & MacDonald, 2005; Schlueter, 2017; Veenstra, Acheson, & Meyer, 2014; Vigliocco et al., 1995). Misrepresentation-based accounts and cue-based retrieval accounts offer slightly different explanations for this widespread phenomenon, though both types of accounts put the locus of the effect at the representation of the NPs' number feature. According to percolation accounts, plural NPs carry an explicit morphosyntactic feature [+PL], whereas there is no corresponding explicit [+SG] feature, and singular NPs are considered the default simply by virtue of the absence of a [+PL] feature. Consequently, number features can only percolate from a plural attractor NP, but not from a singular one (e.g., Eberhard, 1997; Eberhard et al., 2005). See Badecker and Kuminiak (2007) for a discussion on how the Marking and Morphing model accounts for the asymmetry in a very similar way. Cue-based retrieval accounts claim that the reason for the asymmetry is that plural NPs are more salient in memory, and are thus more likely to be (erroneously) retrieved in cases of a feature mismatch than NPs with no such feature (Phillips, 2013; Staub, 2009). Such "pop out" effects are also commonly found in nonlinguistic psychological experiments such as visual search (Treisman & Gelade, 1980).

Consequently, one may predict that the asymmetry effect may decrease as a function of memory capacity, with people with lower memory abilities showing a greater asymmetry than people with higher memory abilities.

Attraction errors in older L1 speakers

Reifegerste et al. (2017) found that during sentence comprehension (timed grammaticality judgment and self-paced reading), older speakers were more susceptible to attraction errors, particularly if they had low working memory (WM) capacities, compared to younger speakers. This finding is in line with previous studies reporting age effects on sentence-level processing (Kemper, 1988; Kemtes & Kemper, 1997; Kynette & Kemper, 1986; Light & Capps, 1986; Obler et al., 1991; Peelle, Troiani, Wingfield, & Grossman, 2010; Reifegerste & Felser, 2017; Waters & Caplan, 2001; Yoon et al., 2015; but see Altmann & Kemper, 2006; Davidson, Zacks, & Ferreira, 2003). In some of these cases, the observed declines were argued to be tied to age-related decreases in WM abilities as well, though this has rarely been tested explicitly, and indeed some studies were not able to tie declines in sentence processing to WM abilities (Caplan & Waters, 2005; Waters & Caplan, 2001). Note that the above-mentioned study by Reifegerste et al. (2017) investigated agreement processing in comprehension whereas the present study taps production, and that attraction effects in comprehension may be different from those found in production (Acuña-Fariña, 2012; Tanner, Nicol, & Brehm, 2014; see also Schlueter, 2017).

Agreement attraction in L2 processing

L2 speakers show general difficulty with agreement computation as evidenced by higher error rates, longer RTs, differences in eye-movement patterns, and/or different neural responses, in both comprehension and production (Chen, Shu, Liu, Zhao, & Li, 2007; Grüter, Lew-Williams, & Fernald, 2012; Keating, 2009; Lardiere, 1998; Sato & Felser, 2010; Shibuya & Wakabayashi, 2008; VanPatten, Keating, & Leeser, 2012; but see Armstrong, Bulkes, & Tanner, 2016; Foote, 2011), sometimes as a function of proficiency (Hopp, 2006; Keating, 2009; Sagarra & Herschensohn, 2010). The picture of attraction effects in L2 speakers is somewhat mixed, with a few studies reporting similar attraction effects between L1 and L2 speakers (Lago & Felser, 2018; Tanner, Nicol, Herschensohn, & Osterhout, 2012), one study reporting delayed attraction effects for L2 speakers (Lim & Christianson, 2015), and a third group of studies finding no attraction effects in L2 speakers in speeded judgment or comprehension tasks (Jiang, 2004; Schlueter, 2017).

In the present study we investigate agreement attraction with a sentence completion paradigm that is generally considered to tap into production processes (see below for further details). To the best of our knowledge, only one study has thus far compared L1 and L2 adult speakers in production. Lago and Felser (2018), who employed the same experimental paradigm as we used in the present study, found no overall differences between native and non-native speakers' susceptibility to attraction. Unlike the present study, Lago and Felser (2018) focused on effects of NP-verb distance and did not assess whether L1 and L2 speakers differed with regard to the singular-plural asymmetry. Another study by Veenstra et al. (2017) investigated agreement errors in production in monolingual and bilingual children. This study found a greater incidence of attraction errors for singular subjects, and – similar to the findings by Lago and Felser (2018) – no difference in susceptibility to attraction errors between the two groups. Veenstra et al. (2017) also reported significant effects of verbal and nonverbal WM on agreement errors, with higher WM scores being associated with fewer errors in both monolingual and bilingual children. This suggests the possibility that aging L2 speakers may also be more susceptible to attraction errors, given their decreasing cognitive control abilities, which would be in line with the findings of Reifegerste et al. (2017) for aging L1 speakers.

The present study

We examined native and non-native agreement processing in German across the adult lifespan using a binary-choice sentence-completion paradigm, similar to the one first used by Staub (2009). In this paradigm, participants read a sentence preamble (e.g., *The key to the cabinets…*) and then choose the correct singular or plural verb form (e.g., *was* vs. *were*). This task has since been used with both L1 and L2 speakers (see, e.g., Lago & Felser, 2018; Ristić et al., 2016; Smith, Franck, & Tabor, 2018; Veenstra, Acheson, Bock, & Meyer, 2014). While this paradigm has thus far not been employed in studies with older adults, it may in fact prove to be very useful for this particular group of participants. In more traditional sentence-production experiments, in which participants may, for example, be asked to describe a picture, RTs and accuracy rates may be strongly affected by age-related increases in word-finding difficulties (Au et al., 1995; Connor, Spiro, Obler, & Albert, 2004; Newman & German, 2005; Randolph, Lansing, Ivnik, Cullum, & Hermann, 1999). This potential confound is not a concern when using Staub's (2009) paradigm.

Our study employs a SUBJECT NUMBER (Singular/Plural; within-subjects) × MATCH (Match/Mismatch; within-subjects) × AGE (continuous; between-subjects) × LANGUAGE GROUP (L1/L2; between-subjects) design. We further assessed several cognitive variables (working memory, interference control, processing speed) to examine their effects on agreement processing and to control for potential L1/L2 differences in these variables. With this study, we sought to answer the following research questions. First, what is the pattern of native and non-native agreement processing across the lifespan? Second, does end-state L2 processing resemble that of L1 speakers? That is, do L2 speakers, after decades of exposure, show a performance pattern similar to that of L1 speakers (e.g., with regard to asymmetrical singular-plural effects)? Third, to what extent do cognitive abilities modulate native and non-native agreement computation across the lifespan?

Regarding the first research question, we expect decreases in accuracy and increases in RTs for aging L1 speakers, in line with the majority of previous work on L1 sentence-level processing (incl. agreement processing), which reports age-related declines in performance. Given the dearth of research on L2 sentence processing across the lifespan, our hypotheses are speculative. While L2 speakers may well experience the same age-related declines as L1 speakers, it is also possible that their performance actually improves with increasing age due to greater experience with and exposure to the L2, assuming similar ages of acquisition and arrival across the age range.

As regards the second research question, various outcomes are possible, given the complexity of our design, in which we are assessing both the role of language status (L1 vs. L2) and the role of age, which may or may not affect the two groups of speakers differently. On the one hand, at young adult age, we may not find any L1/L2 differences, in line with the findings by Lago and Felser (2018) and Tanner et al. (2012). This lack of a difference may either persist across the lifespan, or differences may start to emerge in older participants. This latter possibility (and indeed all outcomes in which the size of the L1/L2 difference increases with increasing age) might be predicted by models of L2 processing which posit that native and non-native processing are inherently similar, but the latter is more cognitively demanding (Cunnings, 2017; Foote, 2011; Hopp, 2006, 2010; McDonald & Roussel, 2010; McDonald, 2006; Sagarra & Herschensohn, 2010). Under such a view, the younger L2 speakers in our study might be cognitively "fit" enough to compensate for the cognitive demands that non-native sentence processing imposes, while age-related declines in cognitive resources (e.g., working memory, interference control, processing speed) lead to the emergence of L1/L2 differences at older age.

On the other hand, we might observe L1/L2 differences at younger age – for example, while L1 speakers are expected to show attraction effects, these effects might be smaller for the L2 group, look somewhat different (e.g., with regard to the singular-plural asymmetry), or might be completely absent. Whatever the nature of these L1/L2 differences at young age, they might either be the same across the age range, or be larger or smaller in older adults. Findings in line with the first of these possibilities have been reported by Reifegerste et al. (2019), who found that differences in the way younger L1 and L2 speakers process regular inflections and derivations persist into old age. The second possibility – L1/L2 differences increasing with age – would be in line with the notion that L2 processing is more cognitively demanding than L1 processing, with L2 speakers being disproportionally affected by age-related declines of cognitive resources. Lastly, it is also possible that L1/L2 differences that exist at a younger age may decrease (or even disappear) with increasing age. This view is in line with usage- and exposure-based models of L2 processing, according to which those aspects of language processing which differ in L2 compared to L1 processing may become more native-like, given the right amount and type of exposure. For example, the declarative/procedural (DP) model (Ullman, 2001, 2005, 2016) postulates that in L1 speakers, lexical processing relies on declarative memory, while grammatical processing is governed by procedural memory. Initially, L2 grammatical processing depends more heavily on declarative memory, but evidence suggests that longer exposure leads to more native-like processing (i.e., a dependence of L2 speakers' grammatical processing on procedural memory) (Hamrick, Lum, & Ullman, 2018; Ullman & Lovelett, 2018).

Lastly, regarding the third research question, collecting several individual-differences measures allows us to statistically account for the influence of age-sensitive cognitive factors on both L1 and L2 processing. If these factors mediate the relationship between aging and sentence-processing performance, it is quite possible that their inclusion may offset the effects of chronological age if their variance explains the variance in our data better than the factor age does.

Method

Participants

We recruited sixty German native speakers (L1 group; age range: 18–84 years) and sixty English native speakers, with German as their late-learned non-native language (L2 group; age-range: 21–81 years). All participants gave written informed consent, reported normal hearing and (corrected-to-) normal vision, and no neurological or language-related impairments. None of the participants had started learning a second language before the age of 11. All participants were living in Germany (greater Berlin area) at the time of testing, and none of the L1 speakers had spent more than one year in a non-German-speaking country. Handedness was assessed using the Edinburgh Handedness Inventory (Oldfield, 1971). See Table 1 for demographic information, as well as German language abilities ('Goethe Score'), cognitive health ('Mini-Mental State'), self-reported overall daily use of German ('Average overall daily use of German'), total number of foreign languages learned over the course of one's life ('Number of L2s learned'), and several cognitive measures that were assessed during the same testing session (see below for details). Note that Table 1 splits up the participants by decade to illustrate the degree to which participants differed on certain demographic variables and cognitive measures across the lifespan. Our main analyses, however, treat age as a continuous variable.

The L2 participants' German skills were examined using the Goethe Placement Test, a 30-item multiple-choice cloze test (www.goethe.de/cgi-bin/einstufungstest/einstufungstest.pl) assessing vocabulary and grammar knowledge in German. The average score for each decade (see Table 1) corresponds to the C1 skill level, labeled "effective operational proficiency" in the Common European Framework for Languages (CEFR; Verhelst, Van Avermaet, Takala, Figueras, & North, 2009).[1] All participants scored at least 17 points on the Goethe test, corresponding to B2 skill level ('intermediate proficiency').[1]

All L1 and L2 speakers were also tested with the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), which was administered in the respective native language. While there was a significant age-related decline in MMSE scores for the L1 speakers and a marginally significant decline for the L2 speakers, no participant of either group scored less than 27 points (out of 30), suggesting that our participants were not affected by severe age-related pathological cognitive impairment (e.g., Alzheimer's dementia).

Agreement-processing experiment

Materials

We constructed 48 experimental and 84 filler sentences. The experimental sentences contained a complex subject NP comprising a noun phrase (the subject head) modified by another NP in genitive case (the "attractor"). Four experimental conditions were created by manipulating the grammatical number of both the subject head and the attractor NP as shown in Table 2. See Appendix A for the experimental sentences.

All plural nouns were clearly marked for plurality by an overt plural suffix and/or by umlauting (*Anwalt* 'lawyer' → *Anwälte* 'lawyers') so as to make noun plurality maximally salient. Singular NPs were always introduced by an unambiguously singular article. For all trials, participants selected between the third-person singular present-tense form and the third-person plural present-tense form of the verb *haben* 'to have': *hat* 'has' and *haben* 'have'. These forms were chosen over the corresponding forms of 'to be' because the latter have a much greater form-frequency imbalance between singular and plural form (*sein*'to be': 7453 occurrences per million vs. 2961 per million; *haben* 'to have': 3356 per million vs. 2517 per million).[2]

The filler sentences were designed so as to discourage participants from developing a strategy in which they memorized the number feature of the first noun of the sentence to inform their decision. To this end, we included sentences containing coordinated subjects (*Die Oma und der Opa [HAT/HABEN] einen großen Garten.* 'The grandmother and the grandfather [HAS/HAVE] a big garden.') as well as sentences in which the critical verb occurred in an embedded clause which itself had a plural subject (*Der Verkäufer denkt, die Kunden [HAT/HABEN] zu hohe Ansprüche*. 'The salesman thinks the customers [HAS/HAVE] too high expectations.').

All sentences were nine words in length. Half of the sentences required a singular verb form, and half required a plural verb form. Four

[1] While the American Council on the Teaching of Foreign Languages (ACTFL) has published one-directional conversion tables to align ACTFL scores with CEFR scores (American Council on the Teaching of Foreign Languages, 2015), no such conversions are available in the reverse direction (converting CEFR to ACTFL ratings). However, ACTFL ratings of "Advanced High"/"Superior" correspond to C1 level (the average performance level of our participants), and ACTFL ratings of "Advanced Mid" correspond to B2 level (the lowest performance level of our participants).

[2] All frequencies reported in this paper are based on the Mannheim corpus as reported in CELEX (Baayen, Piepenbrock, & Gulikers, 1995).

Table 1
Demographic information on participants, broken down by language group and age decade. Note that the oldest decade labeled "70s+" contains the data from three L1 speakers and one L2 speaker above the age of 80. Standard deviations are provided in parentheses.

		20s	30s	40s	50s	60s	70s+	Correlation with age
L1	n	10	10	10	10	10	10	
	Sex	5 F, 5 M	8 F, 2 M	9 F, 1 M	7 F, 3 M	7 F, 3 M	8 F, 2 M	
	Handedness	7 R, 3 L	10 R	9 R, 1 L	10 R	10 R	9 R, 1 L	
	Age	24.1	32.9	45.5	55.1	65.0	77.2	
		(3.2)	(3.0)	(3.2)	(2.7)	(1.9)	(4.3)	
	Education	15.1	16.6	15.1	15.6	14.1	16.2	$r < 0.01$,
		(1.9)	(2.3)	(1.9)	(2.2)	(2.1)	(2.0)	*ns*
	Mini-Mental State	29.7	29.7	29.7	29.1	28.4	28.4	$r = -0.57$,
		(0.5)	(0.7)	(0.5)	(0.7)	(1.1)	(1.2)	$p < .001$
	Working Memory	5.9	5.5	5.0	5.0	4.6	4.4	$r = -0.53$,
		(0.7)	(0.8)	(1.9)	(0.6)	(1.0)	(0.5)	$p < .001$
	Flanker Congruency Cost	45.8	44.3	49.5	51.4	56.4	58.6	$r = 0.29$,
		(14.0)	(16.5)	(15.0)	(19.1)	(29.7)	(27.2)	$p = .025$
	Processing Speed	406	422	478	475	549	565	$r = 0.80$,
		(15)	(29)	(48)	(31)	(44)	(72)	$p < .001$
L2	n	10	12	8	10	10	10	
	Sex	6 F, 4 M	8 F, 4 M	4 F, 4 M	6 F, 4 M	7 F, 3 M	8 F, 2 M	
	Handedness	9 R, 1 L	11 R, 1 L	7 R, 1 L	10 R	9 R, 1 L	9 R, 1 L	
	Age	25.0	33.3	46.5	53.9	64.2	72.7	
		(2.8)	(2.8)	(3.3)	(2.2)	(2.7)	(3.1)	
	Education	15.4	16.3	16.5	15.1	15.7	15.9	$r < 0.01$,
		(1.6)	(0.9)	(2.7)	(1.4)	(1.2)	(2.3)	*ns*
	Goethe Score	24.2	25.5	22.4	24.2	25.9	24.8	$r = 0.10$,
		(2.3)	(3.6)	(3.8)	(4.2)	(3.9)	(4.2)	*ns*
	Mini-Mental State	29.5	29.6	29.0	29.8	29.1	28.8	$r = -0.22$,
		(0.7)	(0.7)	(1.6)	(0.4)	(1.1)	(1.2)	$p = .09$
	Working Memory	6.2	5.8	6.6	5.5	5.2	4.6	$r = -0.54$,
		(0.7)	(0.7)	(0.8)	(0.8)	(0.7)	(0.8)	$p < .001$
	Flanker Congruency Cost	44.3	48.2	44.8	41.3	36.5	45.9	$r = -0.05$,
		(15.9)	(19.9)	(24.7)	(26.8)	(20.9)	(17.3)	*ns*
	Processing Speed	389	434	438	484	544	549	$r = 0.76$,
		(19)	(32)	(32)	(45)	(94)	(48)	$p < .001$
	AoA German	14.9	16.9	23.5	20.9	21.1	22.3	$r = 0.23$,
		(4.5)	(6.3)	(12.5)	(7.3)	(12.6)	(9.4)	$p = .075$
	Age of Arrival in Germany	19.3	25.3	28.8	28.3	30.7	31.5	$r = 0.36$,
		(5.8)	(4.5)	(11.8)	(7.7)	(13.6)	(13.3)	$p = .004$
	Length of Residence	5.7	8.0	17.7	25.9	33.7	41.2	$r = 0.81$,
		(4.9)	(4.2)	(9.6)	(9.1)	(15.1)	(14.4)	$p < .001$
	Average overall daily use of German	38.5%	30.1%	39.7%	45.0%	37.1%	46.4%	$r = 0.15$,
		(16.7%)	(15.2%)	(15.1%)	(8.7%)	(12.9%)	(10.1%)	*ns*
	Number of L2s learned	2.0	2.6	2.3	2.2	2.6	2.8	$r = 0.02$,
		(0.9)	(1.7)	(1.1)	(1.9)	(1.1)	(1.5)	*ns*

Table 2
Overview of the stimulus materials. Participants saw the preamble before selecting the correct verb form, which was followed by the continuation of the sentence.

Condition	Example
Singular-Match	*Der Brief des diplomatischen Anwalts [HAT/HABEN] zu viele Fehler.*
Singular-Mismatch	*Der Brief der diplomatischen Anwälte [HAT/HABEN] zu viele Fehler.*
Plural-Match	*Die Briefe der diplomatischen Anwälte [HAT/HABEN] zu viele Fehler.*
Plural-Mismatch	*Die Briefe des diplomatischen Anwalts [HAT/HABEN] zu viele Fehler.*
Translation:	'The letter/letters from the diplomatic lawyer/lawyers [HAS/HAVE] too many mistakes.'

experimental lists were created based on a Latin-Square design, so that a given sentence was presented to each participant in only one of the four experimental conditions laid out in Table 2. The lists were presented in reversed order to half of the participants (within each age decade and within each language group). The items were pseudorandomized, such that (i) no more than two target (vs. filler) sentences occurred in a row, (ii) no more than three trials in a row required a left- (or right-)button press (if responded to correctly), and (iii) no two consecutive sentences were of the same experimental condition (Van Casteren & Davis, 2007).

Apparatus and procedure

The sentence-processing experiment was programmed and presented on a laptop computer with a 15-inch screen using the DMDX software package (Forster & Forster, 2003). The sentence preambles were presented word by word at a rate of 500 ms per word (directly after one another; i.e., with an ISI of 0), with words appearing in the center of the screen in Arial font size 36 in black letters against a white background.

Participants were instructed to read the words appearing on the screen silently. Each sentence started with a sequence of five fixation crosses "+ + + + +". After the 5-word preamble, participants were prompted to choose the correct verb form, either singular *HAT* or plural *HABEN* by pressing the corresponding button on a gamepad, with the participant's dominant hand controlling the singular response button. Unlike the other words (which appeared in the center of the screen), *HAT* and *HABEN* appeared in all uppercase letters on the left and the right edge of the screen, respectively, to remind participants which button corresponds to which form. After the participant made their choice (or, if the participant did not respond, after a timeout of 5000 ms), the sentence continued with three additional words.[3] There was no feedback on accuracy.

The experiment was preceded by eight practice sentences. The sentence-completion task lasted between 15 and 20 min. Participants were tested individually in a quiet room. Before the experiment, they filled out a short biographic questionnaire. The sentence-completion task was followed by the individual differences tests described below.

Individual differences tests

In order to examine the influence of cognitive abilities on sentence-processing abilities in our participants, we assessed their Working Memory, Interference Control, and Processing Speed. These individual differences tests were run using the open-source Psychology Experiment Building Language (PEBL) test battery (Mueller & Piper, 2014; Mueller, 2012).

Working Memory was assessed as spatial span using the Corsi block-tapping task backwards (Corsi, 1972).[4] Participants saw nine blue squares against a black background on the computer screen. The squares lit up one after another in a certain sequence, which the participant was asked to reproduce afterwards in reverse order using the computer mouse. The length of the sequence to memorize increased by one if participants got at least one of two trials per sequence length correct. As can be seen in Table 1, both L1 speakers and L2 speakers showed a significant age-related decline in Working Memory scores.

Interference Control was assessed using the Eriksen Flanker task (Eriksen & Eriksen, 1974). Participants saw arrows on the screen and were asked to indicate in which direction the center arrow faced (left/right) via button press on the computer keyboard. The center arrow was flanked either by arrows pointing in the same direction (congruent condition: →→→→→), by arrows pointing in the opposite direction (incongruent condition: ←←→←←), or by dashes (neutral condition: – – → – –). The difference between RTs in the incongruent condition and the congruent condition constitutes a participant's Congruency Cost, a proxy for their Inhibitory Control skill (with smaller Congruency Costs corresponding to greater Inhibitory Control). While Congruency Costs significantly increased with increasing age for L1 speakers, there was no such effect for L2 speakers (see Table 1).

For Processing Speed, we took a participant's average RT in the neutral condition of the Flanker task. Correct performance in this condition involves the perception of a visual stimulus, a binary-decision component, and the motor execution of pressing a button. As such, the speed of performance in this condition captures the speed of non-linguistic processing (perception, decision making, motor execution) involved in our linguistic task. Processing Speed decreased significantly with increasing age in both the L1 and the L2 group, as evidenced by increasing average RTs (see Table 1).

Analyses

The dependent measures were accuracy rate and RTs (for correct responses) in the sentence-completion task. We calculated mixed-effects logistic regression models (binomial family) for the accuracy rates and linear mixed-effects regression models for log-transformed RTs (natural log) and log-transformed inverse efficiency scores (IES; calculated as RT/Accuracy per participant per condition (Townsend & Ashby, 1978, 1983); natural-log-transformed for analyses), using the languageR package (Baayen, 2013) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). The following fixed factors of interest were included: LANGUAGE GROUP (2 levels: L1, L2), SUBJECT NUMBER (2 levels: singular, plural), MATCH (2 levels: match, mismatch), AGE (continuous), EDUCATION (continuous, in years), WORKING MEMORY (continuous), INTERFERENCE CONTROL (continuous), PROCESSING SPEED (continuous), as well as, for L2 speakers only, GOETHE SCORE (continuous) and AOA GERMAN (continuous).[5] In order to control for the influence of differences in frequency between singular (M_{Freq} for subjects = 69.75, M_{Freq} for attractors = 12.94) and plural (M_{Freq} for subjects = 31.74, M_{Freq} for attractors = 33.50) nouns, we further added SUBJECT FREQUENCY and ATTRACTOR FREQUENCY (both log-transformed using the natural log) as covariates. Lastly, TRIAL NUMBER (position of trial within the experiment) and PREVIOUS TRIAL RT (the RT of the trial prior to the one in question) as well as their interaction were added, both to remove residual auto-correlation and to control for trial-level task effects (Baayen & Milin, 2010). As SUBJECT FREQUENCY, ATTRACTOR FREQUENCY, TRIAL NUMBER, and PREVIOUS TRIAL RT were not manipulated, effects involving these control variables will be reported in the results tables (regardless of their significance) but will not be discussed.[6] All continuous predictors were mean-centered; all categorical predictors were assigned sum-coded contrasts (e.g., −0.5 and 0.5) (Barr, Levy, Scheepers, & Tily, 2013).

Random factors were participants and items for accuracy and RT analyses, and participants for inverse efficiency score analyses. Following Barr et al. (2013), we started with a maximal random-effects structure and simplified the model in cases of convergence failure. This led to the inclusion of SUBJECT NUMBER and MATCH as well as their interaction as by-participant random intercepts. For continuous outcome variables (RTs and inverse efficiency scores), for which *p*-values are not automatically computed in R, *p*-values were obtained from *t*-tests with the number of degrees of freedom calculated as the difference between the number of data points minus the number of fixed effect estimates (Baayen, Davidson, & Bates, 2008). See table notes for information on the degrees of freedom for the respective models.

Results

We excluded trials with RTs shorter or longer than 3 SDs from the

[3] We added a continuation after the forced-choice task because pilot tests indicated that especially older participants were confused and dissatisfied when the sentences were left "unfinished" after participants had made their decision of what they considered to be the correct verb form.

[4] We also collected data from our participants using the forwards version of the Corsi block-tapping task. However, these values were collinear with the backwards data (r = .59). For this reason, we decided to disregard the forwards version data, in order to retain the potential explanatory power of the backwards-version data.

[5] Note that in our sample, chronological AGE and LENGTH OF RESIDENCE are highly collinear (r = .82); while there is a considerable range in LENGTH OF RESIDENCE for older participants, our theoretical focus on the processing of a late-acquired L2 restricted the LENGTH OF RESIDENCE for younger participants (see Table 1). For this reason, LENGTH OF RESIDENCE was not entered into the final model, so as not to mask potential effects of chronological age, which are the focus of the present paper. (See Wurm & Fisicaro, 2014, for discussion on how to treat collinear predictors in regression models.) Exploratory analyses revealed that the inclusion of LENGTH OF RESIDENCE does not improve the fit of the models (accuracy: $\chi^2(1) = 0.1554$, p = .694).

[6] In response to a comment by a reviewer, we calculated the mixed-effects regression models also without the non-manipulated covariates (i.e., without SUBJECT FREQUENCY, ATTRACTOR FREQUENCY, TRIAL NUMBER, PREVIOUS TRIAL RT, GOETHE SCORE, or AGE OF ACQUISITION). This yielded virtually no difference in the pattern of results, with the exception of a main effect of AGE on the accuracy results of the L2 speakers (which was significant in the original analysis and marginal in the analysis without covariates) and a main effect of AGE on the RT results of the L1 speakers (which was non-significant in the original analysis and marginal in the analysis without covariates). Model comparisons using a likelihood ratio test yielded greater goodness of fit (lower AIC) for the more complex models, with the exception of L1 speakers' accuracy rates (for which the more complex model and the model without covariates did not significantly differ in goodness of fit; p = .937).

Table 3
Descriptive mean accuracy and RT data (SDs in parentheses), broken down by language group, age decade, and condition.

			20s	30s	40s	50s	60s	70s+	Average
L1	Accuracy	Singular-Match	0.983 (0.129)	0.983 (0.129)	0.992 (0.092)	0.992 (0.092)	0.950 (0.219)	1.000 (0.000)	0.983 (0.128)
		Singular-Mismatch	0.965 (0.185)	0.975 (0.157)	0.974 (0.159)	0.940 (0.238)	0.948 (0.222)	0.813 (0.292)	0.937 (0.244)
		Plural-Match	0.983 (0.130)	0.975 (0.157)	0.974 (0.159)	0.966 (0.181)	0.948 (0.222)	0.966 (0.183)	0.969 (0.174)
		Plural-Mismatch	0.914 (0.282)	0.957 (0.203)	0.983 (0.129)	0.939 (0.240)	0.914 (0.184)	0.965 (0.282)	0.946 (0.227)
		Average	0.961 (0.193)	0.973 (0.163)	0.981 (0.137)	0.960 (0.197)	0.970 (0.171)	0.907 (0.190)	0.959 (0.199)
	RTs	Singular-Match	664 (276)	829 327)	924 (392)	779 (278)	940 (483)	933 (340)	845 (369)
		Singular-Mismatch	784 (357)	943 (547)	1073 (486)	942 (393)	1053 (510)	1070 (368)	975 (459)
		Plural-Match	768 (395)	844 (393)	875 (417)	791 (348)	870 (389)	970 (443)	852 (402)
		Plural-Mismatch	744 (426)	930 (477)	972 (473)	853 (434)	931 (395)	1112 (554)	926 (475)
		Average	739 (368)	886 (445)	961 (448)	840 (370)	944 (448)	1019 (437)	898 (430)
L2	Accuracy	Singular-Match	0.966 (0.181)	0.930 (0.257)	1.000 (0.000)	0.992 (0.092)	0.966 (0.183)	1.000 (0.000)	0.973 (0.162)
		Singular-Mismatch	0.914 (0.282)	0.935 (0.248)	0.955 (0.208)	0.905 (0.294)	0.904 (0.297)	0.973 (0.161)	0.930 (0.255)
		Plural-Match	0.864 (0.344)	0.879 (0.328)	0.875 (0.333)	0.893 (0.311)	0.939 (0.241)	0.871 (0.337)	0.887 (0.317)
		Plural-Mismatch	0.763 (0.427)	0.891 (0.313)	0.800 (0.402)	0.765 (0.426)	0.819 (0.387)	0.855 (0.354)	0.818 (0.386)
		Average	0.877 (0.329)	0.908 (0.289)	0.908 (0.289)	0.890 (0.314)	0.907 (0.291)	0.925 (0.264)	0.902 (0.297)
	RTs	Singular-Match	847 (648)	864 (579)	775 (436)	750 (416)	1048 (582)	763 (467)	854 (455)
		Singular-Mismatch	797 (529)	884 (541)	1006 (786)	743 (405)	1043 (584)	750 (401)	977 (540)
		Plural-Match	845 (532)	849 (378)	833 (520)	782 (424)	1041 (479)	781 (352)	947 (542)
		Plural-Mismatch	889 (489)	1046 (581)	968 (608)	895 (487)	1187 (597)	867 (402)	867 (555)
		Average	847 (551)	911 (529)	892 (599)	795 (438)	1079 (562)	792 (407)	887 (524)

mean on a per-participant basis as well as timeouts, resulting in 2.4% data loss for the L1 speakers and 3.8% data loss for the L2 speakers. See Table 3 for an overview of the descriptive accuracy and RT data.

L1 group

Accuracy rates

See Table 4 for the mixed-effects model fit to the accuracy data from the L1 group, Fig. 1 for an illustration of these data, and Table 3 for the raw descriptive data.

L1 speakers showed a significant effect of match, with greater accuracy for sentence preambles in which the attractor matched in number with the subject (97.6%) compared to sentence preambles in which the two did not match (94.1%), as well as a main effect of subject number, with greater accuracy for sentence preambles with singular subjects (96.0%) than for preambles with plural subjects (95.7%). Importantly, both of these effects were qualified by an interaction between subject number and match: The effect of match was present only for preambles with singular subject heads and plural attractors (Singular-Match vs. Singular-Mismatch; $b = -2.1010$, $SE = 0.4589$, $z = -4.58$, $p < .001$), but not for preambles with plural subject heads and singular attractors (Plural-Match vs. Plural-Mismatch; $b = -0.4490$, $SE = 0.3608$, $z = -1.25$, $p = .213$).

A significant three-way interaction between subject number, match, and age revealed that the asymmetrical nature of the attraction effect was modulated by chronological age. In order to assess the nature of this effect, we binned participants into a group of "younger adults" (ages 18–39, $n = 20$), "middle-aged adults" (ages 40–59, $n = 20$), and "older adults" (ages 60–84, $n = 20$); see Table B1 in the Appendix for the descriptive data. While neither the group of younger adults nor the group of middle-aged adults showed an interaction between subject number and match (younger: $b = -0.0559$, $SE = 1.0830$, $z = -0.05$, $p = .959$; middle-aged: $b = -0.7572$, $SE = 1.0910$, $z = -0.69$, $p = .488$), the interaction between subject number and match was significant for the older adults ($b = -6.0310$, $SE = 1.3480$, $z = -4.47$, $p < .001$). The younger and the middle-aged group, on the other hand, showed only main effects of match (younger: $b = -1.0830$, $SE = 0.4708$, $z = -1.99$, $p = .046$; middle-aged: $b = -2.0660$, $SE = 0.5486$, $z = -2.69$, $p = .007$). In other words, whereas younger speakers showed a symmetrical attraction effect, with increasing age the attraction effect became increasingly asymmetrical.

There were no other effects of or interactions with any of the factors of interest.

Reaction times

All RT analyses were performed over correct responses only. See Table 5 for the mixed-effects model fit to the RT data from the L1 group, Fig. 2 for an illustration of these data, and Table 3 for the raw descriptive data.

L1 speakers' RTs showed significant main effects of match (longer RTs when the subject and the attractor did not match in number compared to when they did; 950 ms vs. 849 ms) and of subject number (longer RTs for singular compared to plural subjects; 907 ms vs. 888 ms). The effects of match and subject number were qualified by an

Table 4
The best-fit model for the L1 accuracy data.

Random effects:	Name	Variance	SD	Correlation		
participant	Intercept	1.5147	1.2307			
	Subject Number	4.1632	2.0404	0.33		
	Match	0.1890	0.4347	0.41	0.21	
	Subject Number : Match	1.6394	1.2804	−0.56	−0.90	−0.60
item	Intercept	1.5147	1.2307			

Fixed effects:	b	SE	z-value	p-value
Intercept	4.5670	0.2449	18.65	<.001
Subject Number	0.9051	0.4268	2.12	.034
Match	−1.3850	0.2818	−4.91	<.001
Age	0.0017	0.0130	0.13	.898
Trial Number	−0.0002	0.0024	−0.09	.925
Previous Trial RT	0.0001	0.0006	0.13	.899
Subject Frequency	0.0002	0.0018	0.12	.906
Attractor Frequency	0.0031	0.0033	0.94	.346
Subject Number : Match	−1.5800	0.5864	−2.70	.007
Subject Number : Age	0.0027	0.0225	0.12	.906
Match : Age	−0.0180	0.0153	−1.18	.239
Trial Number : Previous Trial RT	<0.0001	<0.0001	−0.39	.699
Subject Number : Match : Age	−0.0715	0.0313	−2.29	.022

Table 5
The best-fit model for the L1 RT data.

Random effects:	Name	Variance	SD	Correlations		
participant	Intercept	0.0562	0.2371			
	Subject Number	0.0079	0.0889	−0.42		
	Match	0.0053	0.0731	0.52	0.55	
	Subject Number : Match	0.0040	0.0633	−0.07	−0.17	−0.22
item	Intercept	0.0021	0.0460			
Residual		0.1106	0.3326			

Fixed effects:	b	SE	t-value	p-value
Intercept	6.7100	0.0345	194.30	<.001
Subject Number	0.0366	0.0179	2.04	.041
Match	0.1038	0.0169	6.14	<.001
Age	0.0031	0.0022	1.42	.156
Trial Number	−0.0009	0.0001	−6.47	<.001
Previous Trial RT	0.0003	0.0001	4.19	<.001
Subject Frequency	0.0002	0.0001	1.93	.054
Attractor Frequency	−0.0002	0.0002	−1.01	.313
Interference Control	0.0047	0.0014	3.37	<.001
Match : Age	0.0022	0.0010	2.14	.032
Subject Number : Match	0.0870	0.0281	3.09	.002
Trial Number : Previous Trial RT	<0.0001	<0.0001	3.66	<.001

Note: *P*-values are calculated with 2683 degrees of freedom (see Methods).

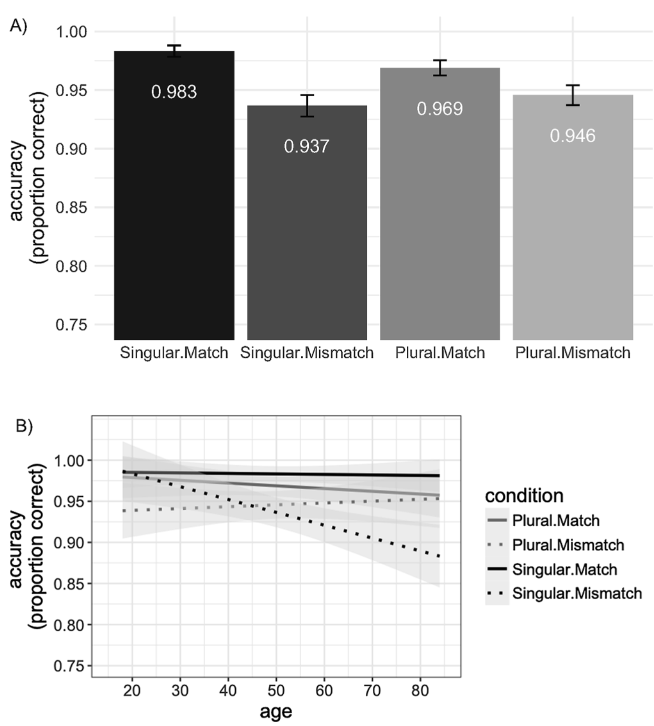

Fig. 1. Accuracy rates across all L1 participants (A) and as a function of age in years (B). Note that all figures are based on raw untransformed data, rather than logits (used in the accuracy analyses), log-transformed RTs (used in the RT analyses), or back-transformed data based on regression model estimates. Thus, the figures may diverge from the model output in some cases. In all figures, error bars and error bands correspond to standard errors of the mean.

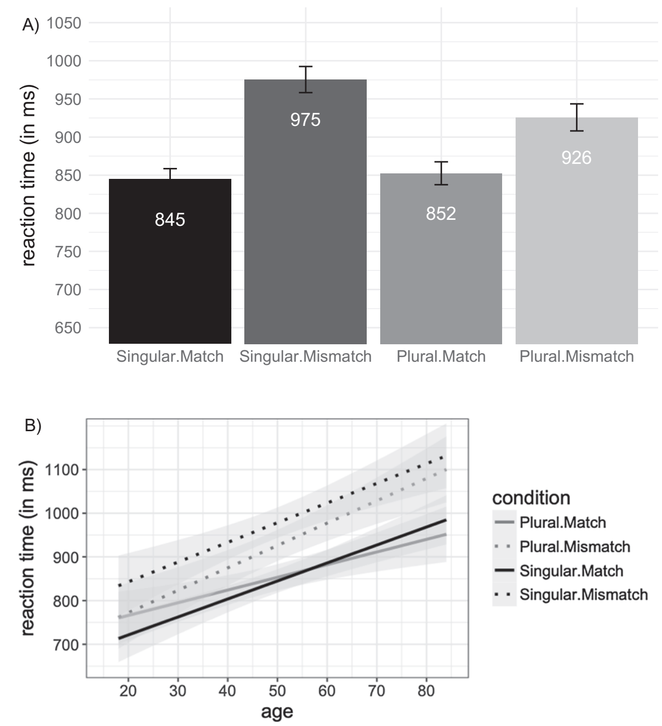

Fig. 2. RTs (based on the raw untransformed data) across all L1 participants (A) and as a function of age in years (B).

interaction between these two factors; while preambles with singular subject heads and plural attractors as well as preambles with plural subject heads and singular attractors both yielded attraction effects, the effect was much stronger for Singular-Mismatch sentences (b = 0.1524, SE = 0.0235, t = 6.50, p < .001) than for Plural-Mismatch sentences (b = 0.0589, SE = 0.0243, t = 2.43, p = .015). The size of the attraction effect (across SUBJECT NUMBER) increased with increasing age, as evidenced by a significant interaction between MATCH and AGE; while AGE did not appear to affect RTs for preambles with matching NPs (b = 0.0020, SE = 0.0021, t = 0.92, p = .358), there was a marginal effect of AGE on RTs for preambles with mismatching NPs (b = 0.0043, SE = 0.0025, t = 1.75, p = .080). Lastly, we found evidence for effects of general cognitive abilities on L1 speakers' RTs, with shorter RTs being associated with better INTERFERENCE CONTROL (i.e., smaller congruency costs).

There were no other effects of or interactions with any of the factors of interest.

Table 6
The best-fit model for the L1 inverse efficiency score data. Note that since efficiency scores are computed on a per-participant basis, only participant was included as a random effect.

Random effects:	Name	Variance	SD	Correlations	
participant	Intercept	0.0908	0.3014		
	Subject Number	0.0410	0.2025	0.05	
	Match	0.0151	0.1228	0.69	0.62
Residual		0.0197	0.1403		

Fixed effects:	*b*	*SE*	*t*-value	*p*-value
Intercept	6.8261	0.0410	166.61	<.001
Subject Number	0.0306	0.0318	0.96	.338
Match	0.1473	0.0246	5.98	<.001
Age	0.0079	0.0023	3.48	<.001
Interference Control	0.0055	0.0020	2.74	.007
Subject Number : Match	0.1107	0.0362	3.06	.002
Match : Age	0.0027	0.0014	1.99	.048

Note: *P*-values are calculated with 233 degrees of freedom (see Methods).

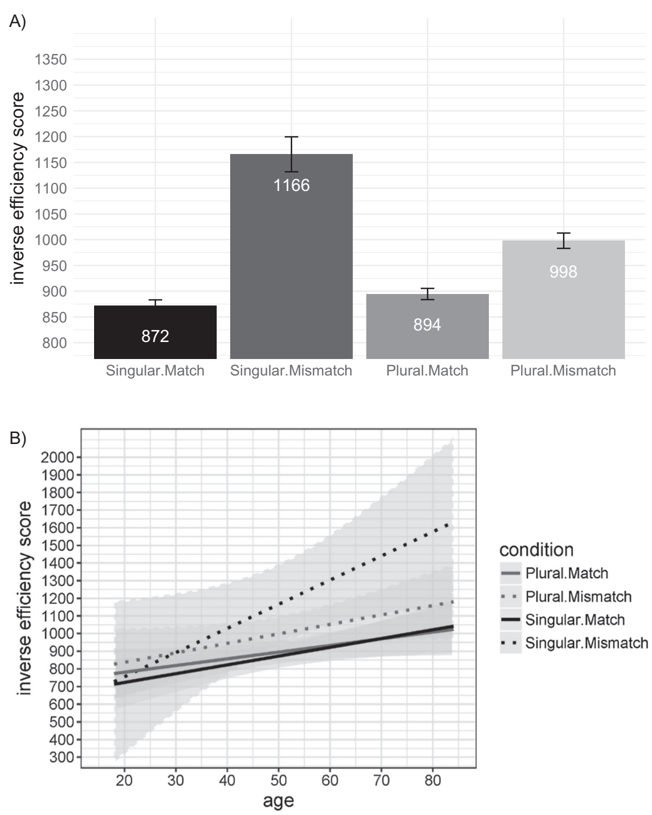

Fig. 3. Inverse efficiency scores (based on the raw untransformed data) across all L1 participants (A) and as a function of age in years (B).

Efficiency scores

In order to account for potential differences in speed-accuracy trade-off between groups of participants (e.g., between younger and older participants [see, e.g., Brébion, 2001; Forstmann et al., 2011; Rabbitt, 1979; Starns & Ratcliff, 2010], or between L1 and L2 speakers), we calculated inverse efficiency scores (IES; Townsend & Ashby, 1978, 1983) per participant per condition (IES = RT/accuracy). Such inverse efficiency scores are sometimes thought of as "RTs corrected for accuracy" – that is, smaller inverse efficiency scores correspond to faster reactions. See Table 6 for the mixed-effects model fit to the inverse efficiency score data from the L1 group and Fig. 3 for an illustration of these data.

We found significant main effects of MATCH (with lower inverse efficiency scores for matching sentences vs. mismatching sentences; 883 ms vs. 1082 ms), of AGE (with larger inverse efficiency scores for older participants), and of INTERFERENCE CONTROL (with lower susceptibility for interference associated with smaller inverse efficiency scores). A significant interaction between SUBJECT NUMBER and MATCH indicated an asymmetrical attraction effect; while both singular and plural subject NPs yielded attraction effects when followed by mismatching attractor NPs, the effect of MATCH was stronger for singular subject NPs ($b = 0.2079$, $SE = 0.0323$, $t = 6.43$, $p < .001$) than for plural subject NPs ($b = 0.0938$, $SE = 0.0295$, $t = 3.18$, $p = .002$). An interaction between MATCH and AGE indicated that the effect of MATCH increased with increasing AGE.

Table 7
The best-fit model for the L2 accuracy data.

Random effects:	Name	Variance	SD	Correlation		
participant	Intercept	0.9218	0.9601			
	Subject Number	2.2296	1.4932	0.10		
	Match	1.4371	1.1988	0.22	0.15	
	Subject Number : Match	0.6939	0.833	−0.16	−0.49	0.74
item	Intercept	0.2791	0.5283			

Fixed effects:	*b*	*SE*	*z*-value	*p*-value
Intercept	3.8110	0.2362	16.13	<.001
Subject Number	1.5830	0.2906	5.45	<.001
Match	−1.0820	0.2538	−4.26	<.001
Age	0.0388	0.0133	2.91	.004
Trial Number	0.0076	0.0020	3.84	<.001
Previous Trial RT	−0.0010	0.0003	−3.38	.001
Goethe Score	0.1601	0.0547	2.93	.003
AoA German	−0.0244	0.0220	−1.11	.268
WM	0.3340	0.2069	1.61	.107
Subject Frequency	0.0018	0.0015	1.17	.240
Attractor Frequency	0.0039	0.0026	1.50	.135
Trial Number : Age	0.0002	0.0001	1.83	.067
Match : WM	0.4786	0.2258	2.12	.034
Trial Number : Previous Trial RT	<0.0001	<0.0001	−0.33	.742

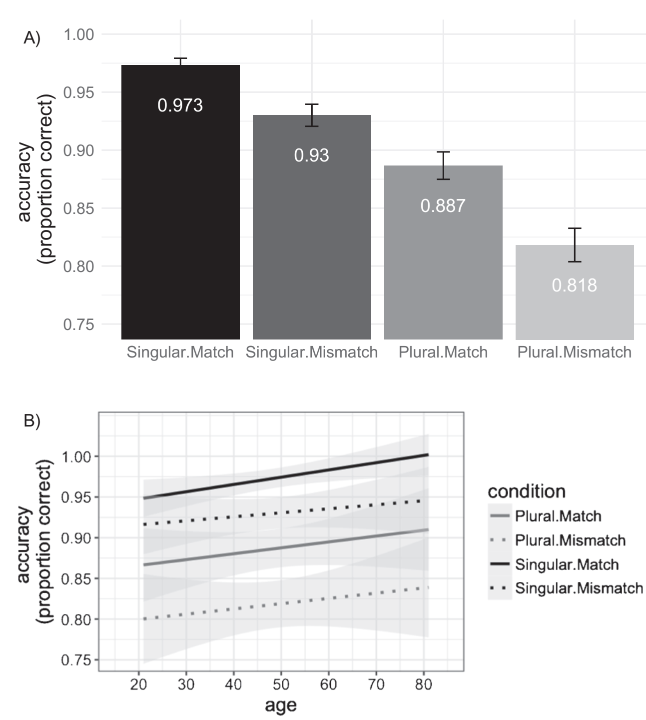

Fig. 4. Accuracy rates (based on the raw untransformed data) across all L2 participants (A) and as a function of age in years (B).

There were no other effects of or interactions with any of the factors of interest.

L2 group

Accuracy rates

See Table 7 for the mixed-effects model fit to the accuracy data from the L2 group, Fig. 4 for an illustration of these data, and Table 3 for the raw descriptive data.

L2 speakers' accuracy rates showed significant main effects of AGE (greater accuracy with increasing age), of MATCH (greater accuracy for preambles in which the subject and the attractor matched in number compared to preambles in which they did not match; 93.1% vs. 87.4%), and of SUBJECT NUMBER (greater accuracy for preambles with singular subjects compared to those with plural subjects; 95.2% vs. 85.2%). There was no interaction between MATCH and SUBJECT NUMBER ($b = -0.3152$, $SE = 0.3898$, $z = -0.81$, $p = .419$); that is, L2 speakers showed a symmetrical attraction effect with similar attraction from plural attractors as from singular attractors. We further found a main effect of GOETHE SCORE (greater accuracy for speakers with greater German skills).

A significant interaction between MATCH and WM indicated that the size of the attraction effect was modulated by WM, with higher WM scores being associated with a decreased attraction effect. While accuracy rates for matching NPs (i.e., Singular-Match and Plural-Match) were not affected by WM ($b = 0.0005$, $SE = 0.2584$, $z = 0.01$, $p = .998$), higher WM scores led to greater accuracy with mismatching NPs (i.e., Singular-Mismatch and Plural-Mismatch; $b = 0.5838$, $SE = 0.2386$, $z = 2.45$, $p = .014$).

There were no other effects of or interactions with any of the factors of interest.

Reaction times

See Table 8 for the mixed-effects model fit to the RT data from the L2 group, Fig. 5 for an illustration of these data, and Table 3 for the raw descriptive data.

L2 speakers' RTs showed main effects of AGE (shorter RTs for older speakers), SUBJECT NUMBER (longer RTs for sentence preambles with singular subjects than for those with plural subjects; 913 ms vs. 857 ms), and MATCH (longer RTs for mismatching compared to matching NPs; 925 ms vs. 851 ms). These main effects of SUBJECT NUMBER and MATCH were qualified by a significant interaction between these two factors, indicating an asymmetrical attraction effect; while singular subject heads and plural attractors yielded attraction effects ($b = 0.1502$, $SE = 0.0230$, $t = 6.54$, $p < .001$), this was not the case for sentence preambles with plural subject heads and singular attractors ($b = 0.0251$, $SE = 0.0264$, $t = 0.95$, $p = .342$).

Similar to L2 speakers' accuracy rates, the size of the attraction effect in RTs was modulated by one of the cognitive individual-differences measures that were collected. Greater congruency costs (i.e., poorer interference control) were associated with larger attraction effects in RTs, as evidenced by an interaction between MATCH and INTERFERENCE CONTROL: Greater congruency cost was associated with longer RTs for non-matching sentence preambles ($b = 0.0039$, $SE = 0.0019$, $t = 2.03$, $p = .042$), while there was no effect of INTERFERENCE CONTROL on RTs for matching ones ($b = 0.0001$, $SE = 0.0018$, $t = 0.04$, $p = .968$).

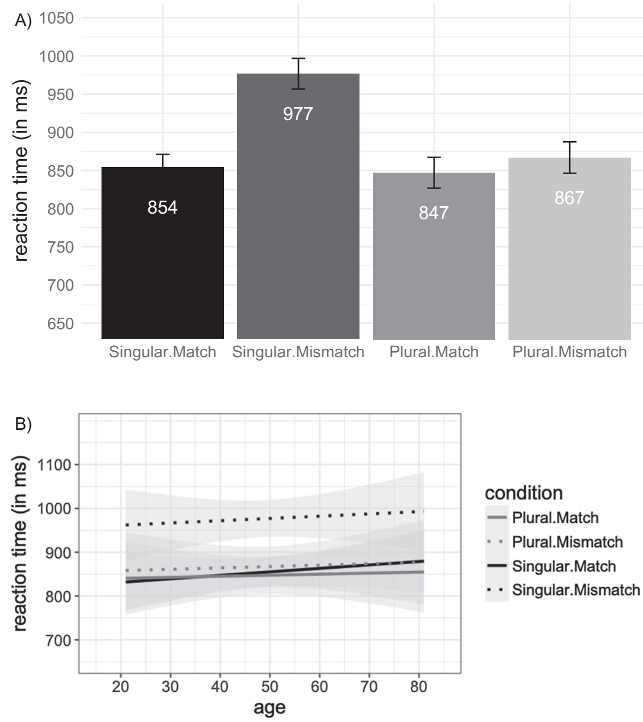

Fig. 5. RTs (based on the raw untransformed data) across all L2 participants (A) and as a function of age in years (B). Error bars and error bands correspond to standard errors of the mean.

Table 8
The best-fit model for the L2 RT data.

Random effects:	Name	Variance	SD	Correlations		
participant	Intercept	0.0742	0.2724			
	Subject Number	0.0191	0.1381	−0.45		
	Match	0.0008	0.0289	0.67	−0.96	
	Subject Number : Match	0.0208	0.1444	0.37	0.42	−0.24
item	Intercept	0.0015	0.0390			
Residual		0.1484	0.3852			

Fixed effects	*b*	*SE*	*t*-value	*p*-value
Intercept	6.6770	0.0423	157.86	<.001
Subject Number	0.0517	0.0241	2.14	.032
Match	0.0886	0.0163	5.45	<.001
Age	−0.0067	0.0033	−2.05	.040
AoA German	0.0036	0.0046	0.78	.435
Goethe Score	0.0026	0.0110	0.24	.810
Interference Control	−0.0010	0.0016	−0.61	.542
Processing Speed	0.0021	0.0007	2.83	.005
Trial Number	−0.0014	0.0002	−7.94	<.001
Previous Trial RT	0.0003	0.0001	4.79	<.001
Subject Frequency	0.0003	0.0001	2.43	.015
Attractor Frequency	−0.0005	0.0002	−2.42	.016
Subject Number : Match	0.1220	0.0373	3.27	.001
Match : Interference Control	0.0019	0.0008	2.45	.014
Trial Number : Previous Trial RT	<0.0001	<0.0001	2.83	.005

Note: *P*-values are calculated with 2490 degrees of freedom (see Methods).

Efficiency scores

See Table 9 for the mixed-effects model fit to the efficiency score data from the L2 group and Fig. 6 for an illustration of these data. Recall that inverse efficiency scores correspond to RTs controlling for accuracy (IES = RT/accuracy); thus, smaller inverse efficiency scores correspond to faster reactions, while larger inverse efficiency scores correspond to slower reactions.

There were main effects of SUBJECT NUMBER (with smaller inverse efficiency scores, i.e., faster reactions, for sentences with singular subject NPs vs. sentences with plural subject NPs; 985 ms vs. 1222 ms), MATCH (smaller inverse efficiency scores for matching vs. mismatching sentences; 1017 ms vs. 1189 ms), AGE (smaller inverse efficiency scores with increasing age), PROCESSING SPEED (smaller inverse efficiency scores with faster performance in the processing speed test), and AOA GERMAN (smaller inverse efficiency scores with earlier AoA of German). An

Table 9
The best-fit model for the L2 efficiency score data. Note that since efficiency scores are computed on a per-participant basis, only participant was included as a random effect.

Random effects:	Name	Variance	SD	Correlations	
participant	Intercept	0.1380	0.3714		
	Subject Number	0.1086	0.3296	−0.63	
	Match	0.0136	0.1165	0.57	−0.51
Residual		0.0274	0.1655		

Fixed effects:	*b*	*SE*	*t*-value	*p*-value
Intercept	6.8713	0.0576	119.27	<.001
Subject Number	−0.0991	0.0476	−2.08	.039
Match	0.1458	0.0279	5.22	<.001
Age	−0.0159	0.0043	−3.69	<.001
Processing Speed	0.0034	0.0008	3.96	<.001
WM	−0.0942	0.0582	−1.62	.107
Goethe Score	−0.0076	0.0140	−0.54	.590
AoA German	0.0146	0.0060	2.45	.015
Match : WM	−0.0860	0.0334	−2.58	.011

Note: *P*-values are calculated with 230 degrees of freedom (see Methods).

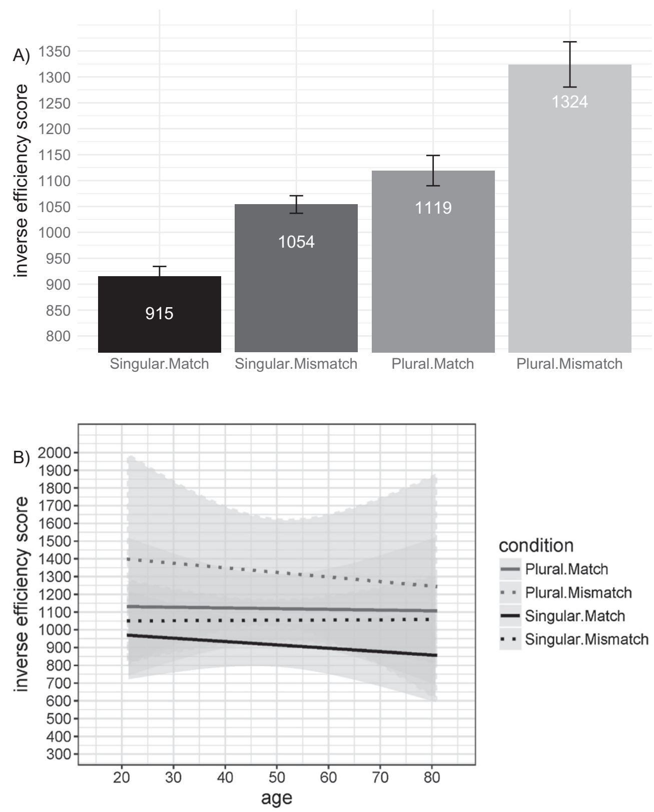

Fig. 6. Inverse efficiency scores (based on the raw untransformed data) across all L2 participants (A) and as a function of age in years (B).

interaction between MATCH and WM indicated a selective effect for WM; the effect of WM on inverse efficiency scores was smaller for matching ($b = -0.0323$, $SE = 0.0641$, $t = -0.50$, $p = .618$) than for mismatching sentences ($b = -0.1086$, $SE = 0.0773$, $t = -1.40$, $p = .164$), though neither of the effects reached significance, presumably due to the reduced power in the efficiency score analyses.

No other main effects or interactions reached significance.

Group comparisons

Finally, we fit two omnibus models across L1 and L2 speakers to the data to assess whether different performance patterns between the two groups also yielded higher-level interactions with the factor LANGUAGE GROUP. Given the high number of interactions, most of which mirror the same findings as were found for the individual groups, we focus on effects involving the factor LANGUAGE GROUP, but report all effects in the respective tables. Note that the L2-specific factors AOA GERMAN and GOETHE SCORE were not included in these analyses.

Table 10
The best-fit model for the accuracy data across L1 and L2 speakers.

Random effects:	Name	Variance	SD	Correlation		
participant	Intercept	1.4627	1.2094			
	Subject Number	2.9777	1.7256	0.17		
	Match	1.0744	1.0365	−0.10	−0.11	
	Subject Number : Match	1.2448	1.1157	−0.72	−0.69	0.52
item	Intercept	0.2013	0.4486			
	Language Group	0.2801	0.5293	0.01		

Fixed effects:	*b*	*SE*	*z*-value	*p*-value
Intercept	4.1780	0.1855	22.52	<.001
Subject Number	1.3590	0.2763	4.92	<.001
Match	−1.6260	0.2503	−6.50	<.001
Age	0.0267	0.0106	2.53	.012
Language Group	0.9369	0.3219	2.91	.004
WM	0.3468	0.1834	1.89	.059
Subject Frequency	0.0039	0.0021	1.81	.070
Attractor Frequency	0.0038	0.0016	2.44	.015
Trial Number	−0.0006	0.0003	−2.03	.042
Previous Trial RT	−1.6830	0.4132	−4.07	<.001
Subject Number : Match	4.1780	0.1855	22.52	<.001
Subject Number : Age	0.0093	0.0152	0.62	.539
Subject Number : Language Group	−1.3790	0.4462	−3.09	.002
Match : Age	−0.0176	0.0145	−1.21	.225
Match : Language Group	0.1059	0.4299	0.25	.805
WM : Match	0.0872	0.2422	0.36	.719
WM : Language Group	0.1302	0.3651	0.36	.721
Trial Number : Language Group	−0.0076	0.0031	−2.42	.015
Previous Trial RT : Language Group	0.0014	0.0006	2.16	.031
Trial Number : Previous Trial RT	<0.0001	<0.0001	−0.28	.777
Age : Trial Number	<0.0001	0.0001	−0.13	.901
Age : Language Group	−0.0359	0.0190	−1.89	.059
Subject Number : Match : Age	−0.0500	0.0225	−2.22	.026
Match : WM : Language Group	−0.8344	0.4033	−2.07	.039
Age : Language Group : Trial Number	−0.0004	0.0002	−2.34	.020

Accuracy rates

See Table 10 for the generalized linear mixed-effects model fit to the accuracy data from all participants.

The omnibus analyses on accuracy data yielded a main effect of LANGUAGE GROUP as well as several interactions with LANGUAGE GROUP, indicating performance differences (lower accuracy for the L2 group) as well as processing differences between the two groups. The effect of SUBJECT NUMBER (greater accuracy rates for sentence preambles with singular compared to plural subject heads; 95.6% vs. 90.5% across participant groups) was significantly greater for L2 speakers than L1 speakers, as evidenced by a two-way interaction between SUBJECT NUMBER and LANGUAGE GROUP. A three-way interaction between MATCH, WM, and LANGUAGE GROUP indicated that the effect of WM on the size of the attraction effect was different for L1 speakers (for whom WM did not affect accuracy rates) compared to L2 speakers (for whom higher WM scores yielded smaller attraction effects, see section 'Results, L2 speakers, Accuracy Rates'). No other interactions between LANGUAGE GROUP and effects of interest were significant.

Reaction times

See Table 11 for the linear mixed-effects model fit to the RTs from all participants.

As with the accuracy rates, several interactions between factors of

Table 11
The best-fit model for the RT data across L1 and L2 speakers.

Random effects:	Name	Variance	SD	Correlation		
participant	Intercept	0.0622	0.2493			
	Subject Number	0.0142	0.1192	−0.46		
	Match	0.0039	0.0626	0.44	−0.06	
	Subject Number : Match	0.0121	0.1101	0.27	0.24	−0.33
item	Intercept	0.0014	0.0368			
	Language Group	0.0022	0.0473	−0.05		
Residual		0.1283	0.3582			

Fixed effects:	*b*	*SE*	*t*-value	*p*-value
Intercept	6.7370	0.0316	213.08	<.001
Subject Number	0.0427	0.0151	2.83	.005
Match	0.0947	0.0118	8.04	<.001
Age	−0.0034	0.0021	−1.60	.110
Language Group	0.0127	0.0474	0.27	.787
Processing Speed	0.0016	0.0005	3.02	.003
WM	−0.0564	0.0272	−2.07	.039
Interference Control	0.0030	0.0011	2.80	.005
Subject Frequency	0.0002	0.0001	2.95	.003
Attractor Frequency	−0.0003	0.0001	−2.15	.032
Trial Number	−0.0011	0.0001	−10.25	<.001
Previous Trial RT	0.0003	<0.0001	7.18	<.001
Subject Number : Match	0.1041	0.0232	4.48	<.001
Match : Age	0.0007	0.0007	1.05	.294
Age : Language Group	0.0069	0.0026	2.69	.007
Match : Language Group	0.0126	0.0234	0.54	.589
Match : Interference Control	0.0006	0.0006	1.12	.263
Interference Control : Language Group	0.0039	0.0022	1.79	.074
Age : Processing Speed	<0.0001	<0.0001	−2.03	.042
Trial Number : Language Group	0.0005	0.0002	2.12	.034
Trial Number : Previous Trial RT	<0.0001	<0.0001	4.41	<.001
Match : Age : Language Group	0.0022	0.0013	1.68	.093
Match : Interference Control : Language Group	0.0024	0.0011	2.09	.037

Note: *P*-values are calculated with 5181 degrees of freedom (see Methods).

interest and language group indicate differences in processing between the two groups. The effect of chronological age on RTs was significantly different for the two groups, with L1 speakers showing no effect of age, while L2 speakers' RTs decreased with increasing age. A marginal interaction between match, age, and language group and a significant interaction between match, interference control, and language group (which qualified an interaction between interference control and language group) suggested that the size of the attraction effects displayed by the two groups were affected by different factors. While L1 speakers' attraction effects increased in size with increasing age, there was no such effect for L2 speakers; on the other hand, L2 speakers' attraction effects increased in size with decreasing interference control, while there was no such effect for L1 speakers.

Discussion

The present study investigated native and non-native speakers' processing of subject-verb number agreement across the lifespan. Both L1 speakers of German and L1 English speakers with German as a late-acquired L2 took part in a timed binary-choice sentence-completion task in which they read preambles containing complex subject NPs and were asked to choose the correct (singular or plural) verb form. Specifically, we examined (i) effects of age on native and non-native agreement processing, (ii) whether end-state L2 processing resembles that of L1 speakers, and (iii) how L1 and/or L2 agreement computation might be influenced by cognitive abilities or resource limitations. In the following we will discuss each of these questions in turn, followed by a brief discussion of the current study's limitations and our conclusions.

Effects of age on agreement processing

Native speakers

In their accuracy rates, the L1 speakers showed a reliable attraction effect, as was expected. Interestingly, the nature of this effect differed between participants of different ages: While the effect was symmetrical for younger and middle-aged speakers, older speakers' accuracy rates showed an asymmetrical attraction pattern, with significant attraction effects only for sentences with singular subject heads and plural attractors but not for plural subject heads with singular attractors.

An asymmetrical pattern has been found in numerous studies on agreement attraction with younger adults (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Bock et al., 1999; Eberhard, 1997; Haskell & MacDonald, 2005; Schlueter, 2017; Veenstra et al., 2014; Vigliocco et al., 1995). Misrepresentation accounts may account for this effect under the assumption that age-related declines in cognitive abilities might yield more ambiguous representations of singular subject NPs in the presence of a plural attractor due to the plural feature's activation spreading more easily to the subject if it is less stably represented. Cue-based retrieval accounts argue that the reason for asymmetry effects lies in the relatively greater salience of plural NPs compared to NPs without a plural feature, rendering the former more likely to be erroneously retrieved from memory upon encountering the verb. Thus, both misrepresentation accounts and cue-based retrieval accounts might predict for the asymmetry to become more apparent with increasing age, as aging is associated with a decline in cognitive capacities (including memory). Recall, however, that we did not find a direct relationship between the size of the attraction effect and any of our cognitive measures (WM, interference control, processing speed). This is particularly surprising since the same WM task was used in a study on agreement attraction by Veenstra et al. (2017), who found that performance in this task reliably predicted both monolingual and bilingual children's rate of attraction errors. One possibility is that there simply was too little variance in our participants' cognitive performance scores (see Table 2).[7]

The L1 speakers' overall RT pattern was in line with their overall accuracy pattern in that we found an asymmetrical attraction pattern, with greater attraction for singular sentences with plural attractors than vice versa; however, unlike for the accuracy rates, the nature of the attraction effect was not significantly modulated by age, but was instead asymmetrical across the age range examined here. As with accuracy rates, likely explanations for this asymmetrical pattern are that plural attractors are marked or more salient in memory compared to singular attractors, such that the successful suppression of a plural attractor takes more time than the suppression of a singular attractor. The size of the attraction effect (across sentences with either singular or plural nouns) increased with age, with older age being associated with longer RTs for mismatching but not for matching NPs. Again, this finding might be predicted by the assumption that the cognitive abilities necessary for creating stable representations of the subject NP and/or the maintenance and retrieval of sentence constituents in memory decline with age. However, we did not find the attraction effect in L1 speakers' RT data being modulated by any of the cognitive measures we collected.

It may seem surprising that we did not find a significant main effect of age on RTs, given well-established findings on general slowing across various cognitive and linguistic tasks (Madden, 1992; Myerson, Ferraro, Hale, & Lima, 1992; Ratcliff & McKoon, 2015; Salthouse, 1992, 1996). One likely reason for this is the inclusion of cognitive factors, such as interference control, which is correlated with age (see Table 1). The inclusion of such factors in the present study was warranted as we

[7] But see Caplan and Waters (2005) and Waters and Caplan (2001), for discussion on the extent to which global WM measures reliably predict efficiency in on-line syntactic operations.

sought to investigate their role in agreement processing, but they may have masked age effects in our statistical results. Indeed, an exploratory analysis without any such covariates revealed a significant effect of AGE on RTs ($b = 0.0058$, $SE = 0.0020$, $t = 2.83$, $p = .005$), supporting this explanation.

We are left with the question of why – in contrast to various previous production studies on agreement attraction – the younger L1 speakers' accuracy rates showed a symmetrical attraction effect. Two potential explanations come to mind. First, one difference between our study and previous production studies employing the same paradigm is the relatively long presentation time that was used in our study. While 500 ms constitute a presentation rate well-suited to accommodate older speakers' slower reading speeds (see, e.g., Reifegerste & Felser, 2017; Reifegerste et al., 2017), this presentation rate is well above the rate usually used with young adult L1 speakers (Bock, Carreiras, & Meseguer, 2012; Ristić et al., 2016; Smith et al., 2018; Staub, 2009; Tanner & Bulkes, 2015; Veenstra et al., 2014). This may have rendered the task comparatively easier for younger than for older speakers, freeing up processing resources for the former which then allowed them, for example, to rehearse the beginning of the sentence in WM, such that plural attractors were no more likely to yield errors than singular attractors. (Evidently, though, this did not lead to an overall smaller number of attraction errors for younger L1 speakers.)

As a second possibility, it may be useful to examine the accuracy rates in conjunction with the RTs, and to consider the issue of speed-accuracy tradeoff and age-related changes to response strategies. While the younger adults showed a symmetrical attraction effect in their accuracy rates, their RTs actually showed an asymmetrical attraction effect in the expected direction (interaction SUBJECT NUMBER × MATCH: $b = 0.1144$, $SE = 0.0351$, $t = 2.15$, $p = .032$). While older adults' accuracy rates showed evidence of an asymmetrical attraction effect, their RTs actually show a symmetrical attraction effect (interaction SUBJECT NUMBER × MATCH: $b = 0.0687$, $SE = 0.0473$, $t = 1.45$, $p = .147$); see also Fig. 2b (though note that for RTs, the interaction between SUBJECT NUMBER, MATCH, and AGE was not significant). In other words, it is possible that the "classic" asymmetrical attraction effect manifested itself differently in younger versus older participants. As the older L1 speakers presumably felt under greater time pressure than the younger ones, their accuracy pattern more closely reflects the pattern that was found in previous studies that used shorter presentation times. Indeed, when analyzing inverse efficiency scores, we found an asymmetrical attraction effect (greater attraction for preambles with plural attractors than for preambles with singular attractors). This effect did not appear to be modulated by age, supporting the notion that the attraction effect was indeed asymmetrical across the age range investigated here, but that it may have manifested itself differently for participants of different ages.

Non-native speakers

L2 speakers made more errors than L1 speakers overall, but more importantly, their accuracy rates also showed attraction effects, a finding in line with previous studies assessing attraction in young adult L2 speakers (Lago & Felser, 2018; Lim & Christianson, 2015; Schlueter, 2017; Tanner et al., 2012). However, unlike the asymmetrical effect we saw in the L1 group, the L2 group showed a highly significant symmetrical attraction effect, with similar numbers of attraction errors elicited by both mismatching conditions. To the best of our knowledge, this study is the first to examine the issue of a singular-plural attraction asymmetry in adult L2 speakers.

One possible reason for the observed symmetrical pattern might lie in L2 speakers' relatively reduced ability to quickly detect that a given noun is plural, which may result in erroneous or underspecified encodings. If a subject head's number feature was not encoded properly during processing, or if the subject head's number could not be retrieved from memory at the point at which the auxiliary had to be chosen, then we might expect (i) participants to show a response bias by defaulting to a singular auxiliary even in the presence of plural subject heads (e.g., because singulars represent the less marked form and perhaps also because the singular response button was controlled by the dominant hand; see also Hammerly, Staub, & Dillon, 2019, for evidence on how a response bias may affect the asymmetrical nature of the attraction effect), and (ii) auxiliary selection to be affected by the number feature of the more recently encountered attractor noun phrase. This is exactly the pattern that we observed in our L2 group: more incorrect singular verb choices than incorrect plural verb choices overall, and similar-sized attraction effects triggered by both singular and plural attractors.

If this reasoning is along the right lines, note that the unstable encoding of the subject head's number feature (or failure to retrieve it) is not specific to L2 speakers. Both participant groups made fewer errors with singular-headed subjects, although this tendency was much stronger for L2 (10 percentage points) than for L1 speakers (0.3 percentage points), as evidenced by an interaction between SUBJECT NUMBER and LANGUAGE GROUP. In other words, L2 speakers were more likely to default to a singular form than L1 speakers. Unlike what we saw in our older L1 participants, the attraction effects seen in our L2 data did not become asymmetrical with age; instead, the presence of either a singular or plural attractor elicited parallel numbers of errors in both younger and older L2 speakers.

A second interesting finding is that for L2 speakers, *age had an overall positive effect on accuracy rates*. That is, in contrast to the increase in attraction errors that we observed in our older L1 speakers, our L2 speakers actually became more accurate with age. This difference is unlikely to be attributable to age-related differences in L2 proficiency (which did not systematically differ with age) or to differences in the age-of-acquisition of German or age-of-arrival in Germany. Both of the latter actually increased with increasing chronological age across participants, and may thus have been expected to lead to worse performance with increasing age, if anything. Instead, we argue that the increase in exposure to German across our participants' lifespans (which for a participant in their 70s was eightfold the number of years compared to a participant in their 20s) led to an increase in accuracy.

In contrast to their accuracy pattern, L2 speakers' RTs showed an asymmetrical attraction pattern, with a larger slowdown triggered by mismatching NPs with plural attractors than by mismatching NPs with singular ones. This apparent discrepancy is most likely due to the fact that our RT analyses were based on correct trials only. That is, in cases in which L2 speakers were able to choose the correct verb form, they showed an asymmetrical RT pattern resembling that of L1 speakers.

Importantly, and in line with the positive effect of age on accuracy in L2, we found an inverse effect of AGE on RTs; that is, older L2 speakers actually performed faster at the task across all conditions than younger L2 speakers. This may appear counter-intuitive at first blush, since aging is strongly associated with slower processing in general. However, this finding may – at least in part – be a consequence of the inclusion of age-sensitive non-linguistic cognitive resources (PROCESSING SPEED and INTERFERENCE CONTROL), over and above which chronological age does not appear to be associated with significant RT increases. In an exploratory model without any covariates, there was no effect of AGE on RTs ($b = 0.0019$, $SE = 0.0026$, $t = 0.70$, $p = .484$). This indicates that even when L2 speakers' declining cognitive resources were not covaried out (and thus can be expected to affect their RT performance), older L2 speakers were *not* slower in responding than their younger counterparts. This was in contrast to older L1 speakers, who showed a marked age-related increase of several hundred milliseconds compared to the younger L1 speakers. Thus, it appears that increasing chronological age had different consequences for L1 speakers and L2 speakers, not only for accuracy (see above) but also for RTs. We explore this further in the next section.

L1/L2 differences in agreement processing

We observed a number of between-group differences, both in overall performance as well as differences in attraction patterns, which suggest L1/L2 processing differences. For one thing, L2 speakers made more mistakes at choosing the correct verb form overall, which is in line with previous work pointing to difficulties during non-native agreement computation. Second, L1 and L2 speakers differed in the nature of the attraction effect they displayed across the lifespan, with L1 speakers showing an increasingly asymmetrical attraction effect (more attraction errors triggered by plural than by singular attractors), while the effect was symmetrical for L2 speakers throughout. Third, we did not find any evidence that end-state L2 processing resembled that of L1 speakers. While numerical differences in overall accuracy rates between L1 and L2 speakers within the oldest age bracket (aged 60+, $n = 40$) were very small and indeed not significant anymore (91.6% vs. 93.9%; $b = 1.2080$, $SE = 0.8332$, $z = 1.45$, $p = .147$; see also Figs. 1b and 4b, and Table B1 in the Appendix), the two groups showed different accuracy patterns, with L1 speakers displaying an asymmetrical and L2 speakers displaying a symmetrical pattern (marginal interaction between SUBJECT NUMBER, MATCH, and LANGUAGE GROUP for the oldest group of participants: $b = -2.2750$, $SE = 1.3050$, $z = -1.74$, $p = .081$). That is, although L2 speakers' processing patterns did not become fully native-like, they did show significant age-related increases in response accuracy. Moreover, unlike what we observed in our L1 speakers, older L2 speakers did not show any significant increases in their RTs. Indeed, when cognitive factors such as PROCESSING SPEED and INTERFERENCE CONTROL were statistically controlled for, L2 speakers' RTs decreased with age. In other words, while several decades of exposure to an L2 did not "turn" L2 speakers into native speakers, the experience that comes from being immersed in a foreign language for longer did render their processing faster and more accurate.

As was laid out in the Introduction, most studies on cognitive or language processing in aging are concerned with the deterioration of these processes. The tacit assumption in the majority of those studies seems to be that if there are age differences in performance, they must reflect some sort of disadvantage for older participants. Only recently have studies begun to investigate the possibility that increasing age comes with not just declines in various cognitive abilities, but also leads to increases in experience and knowledge. Studies have pointed to age-related increases in expertise in various skills, such as playing chess or a musical instruments, where continuous and deliberate practice can offset age-related declines (see Krampe & Charness, 2006, for a review). But even non-specialized general cognitive abilities may show improvements with aging, though this area of research is still sparse. In language specifically, vocabulary size has been shown to increase with age (Bowles & Salthouse, 2008; Facal, Juncos-Rabadán, Rodríguez, & Pereiro, 2012; Verhaeghen, 2003). This increase in vocabulary is usually argued to be the reason for older native speakers' greater accuracy in lexical-decision tasks compared to younger ones (e.g., Cohen-Shikora & Balota, 2016; Madden, 1992; Ratcliff, Thapar, Gomez, & McKoon, 2004; Reifegerste et al., 2017). Moreover, several studies have reported that age-related increases in RTs for lexical tasks (e.g., lexical decisions) are smaller than those for non-lexical visuospatial tasks (e.g., visual search, line discrimination). In other words, older adults actually perform faster on language tasks than would be predicted based on their performance on visuospatial tasks (Lawrence, Myerson, & Hale, 1998; Lima, Hale, & Myerson, 1991). These findings point to the possibility that age-related exposure to and experience with a language may positively affect language processing.

We suggest that in the present study, prolonged L2 use and exposure across the lifespan led to improved agreement processing abilities in our older L2 participants, which offset any declines in accuracy rates and RTs they may have shown otherwise. L1 speakers, on the other hand, did not experience any such improvements, presumably because their performance was already at ceiling in their 20s and was thus more affected by the negative effects of aging on language processing.

Recall that prolonged L2 experience did not result in native-like processing, however, as might have been expected if increased L2 experience leads to greater proceduralization of grammatical processes (Ullman, 2001, 2005, 2016). Even our oldest L2 participants showed a symmetrical attraction effect, unlike the asymmetrical one that we observed in our older L1 speakers.

One possible reason for this may be that even after many years of exposure L2 speakers might not be able to create accurate or sufficiently detailed encodings of complex subject NPs during processing. This could either be due to a reduced ability to identify an NP's number feature under processing pressure, as we suggested above, or because our L2 speakers were more likely to assign a shallow or "flat" rather than a fully-fledged hierarchical structure to the subject NP (cf. Shallow-Structure Hypothesis; Clahsen & Felser, 2006a, 2006b, 2018; Clahsen, Felser, Neubauer, Sato, & Silva, 2010). This would then make it difficult to retrieve the subject head from memory when a verb form must be chosen, with singular and plural attractors being equally likely to be mistaken for the agreement controller. Tendencies for shallow representations have been observed in L2 morphological processing in both younger and older adults (Hahne, Mueller, & Clahsen, 2006; Jacob, Fleischhauer, & Clahsen, 2013; Reifegerste et al., 2019).

Thus, it appears that chronological age affected general measures of L2 language performance in our sentence completion task: Older L2 speakers were overall more accurate and faster than their younger counterparts, possibly due to increased L2 experience in the older compared to the younger adults. However, it appears that the underlying linguistic mechanisms at play during agreement processing (e.g., whether or not hierarchically structured NP representations were created) may be less affected by age, and that older L2 speakers may have employed the same processing mechanisms as the younger ones, despite having had more exposure to the language. See Reifegerste et al. (2019) for similar observations in L2 inflectional processing.

The role of cognitive factors

Both our L1 and our L2 speakers' performance was affected by cognitive factors, albeit to different degrees. There was little evidence that L1 speakers' ability to process agreement correctly and efficiently was affected by their cognitive abilities as measured by our test battery. While we found age-related increases in the size of the attraction effect, this could not specifically be linked to any one of the cognitive abilities that we assessed. What we did find was a main effect of INTERFERENCE CONTROL on L1 speakers' overall RTs. Given that this effect indicated that individuals with better interference control showed faster responses across all experimental conditions, this effect most likely reflects a general facilitation for participants who were able to focus well on the task at hand, rather than any effect specific to agreement attraction.

In contrast, the size of L2 speakers' attraction effects were modulated by both WM (accuracy rates) and by INTERFERENCE CONTROL (RTs). L2 speakers' WM scores were found to modulate their attraction rates: While performance with matching NPs (singular - singular, plural - plural) was not affected by WM scores, increasing WM scores were associated with greater accuracy in choosing the correct verb for mismatching NPs (singular - plural, plural - singular). L2 speakers' RTs were also affected by nonverbal cognitive measures. A main effect of general PROCESSING SPEED yielded faster RTs across experimental conditions. More interestingly, a significant interaction between INTERFERENCE CONTROL and MATCH indicated that better INTERFERENCE CONTROL was associated with increased performance with mismatching (but not matching) noun phrases. These findings support the notion that that L2 grammar processing is more cognitively demanding than L1 processing (see, e.g., Cunnings, 2017), as we found greater modulation of performance by specific cognitive abilities in L2 than in L1.

Limitations and future studies

While we believe that the present study examines a widely studied grammatical phenomenon in a well-balanced and carefully controlled sample of participants using a commonly-employed task, there are a few limitations that may inspire future studies.

First, we used a presentation time of 500 ms for all participants in order to account for older adults' and L2 speakers' slower reading times. This might have rendered the task easier for some participant groups as compared to others. Future studies may want to consider using adjusted presentation rates (e.g., as a function of average accuracy during a practice task, or based on performance speed during practice) to account for both group as well as individual differences in processing speed.

Second, a possible alternative to comparing two different groups of people (L1 vs. L2 speakers) would have been to test bilingual individuals in both of their languages. While we tried to control for various cognitive background variables, any group comparison involves the comparison of different people who may very well differ in various ways beyond the factors that were assessed. This is of particular importance given the purported positive effects of multilingualism on cognitive health, particularly at greater age (see Introduction). Future studies may address this possibility by comparing multilingual speakers in both their L1 and their L2.

Third, one may argue that effects of aging only start to come into play at a very high age (e.g., over the age of 80), as they are counteracted by increases in vocabulary size (in both L1 and L2 speakers; Bowles & Salthouse, 2008; Facal et al., 2012; Keuleers, Stevens, Mandera, & Brysbaert, 2015; Verhaeghen, 2003) before that. It is, however, worth noting that all words used in our study were of reasonably high frequency, so that they could be expected to be well-known to all speakers. Nonetheless, future studies might want to expand the age range beyond what was tested in the present study, though care needs to be taken that such old-old participants are still cognitively healthy.

Fourth, as the majority of previous work examining effects of aging, our study design treated chronological age as a between-subject factor and compared participants of different ages with one another. The underlying assumption of such cross-sectional studies is that participants differ as little as possible from one another, except in their age (and age-related changes to cognition that might be of interest). However, there may of course have been other factors that our participants differed on, such as motivation or type of L2 instruction, which have been shown to affect success at language learning (Bowden, Steinhauer, Sanz, & Ullman, 2013; Dörnyei, 1998). Future studies may want to combine cross-sectional studies with longitudinal studies to arrive at a clearer picture of the effects of aging on sentence-processing abilities (see Cohen-Shikora & Balota, 2016; Connor et al., 2004; Rönnlund, Nyberg, Bäckman, & Nilsson, 2005, for such work on word-finding abilities in L1).

Fifth, the type of data that was acquired in the present study lends itself very well to approaches employing computational modeling. For example, the drift diffusion model put forward by Ratcliff and colleagues has been applied to various aspects of language processing in aging (e.g., letter discrimination, lexical decision), and has been influential in examining the speed-accuracy trade-off and how it might be affected by aging (Hammerly et al., 2019; Ratcliff et al., 2004; Ratcliff, 1978; Ratcliff, Spieler, & McKoon, 2000; Ratcliff, Thapar, & McKoon, 2007). While outside of the scope of the present study, future studies utilizing computational modeling may help elucidate the effects of aging on linguistic abilities further.

Lastly, we would like to point out that with 12 trials per condition, the number of items per condition in our sentence-processing experiment was on the low side. This was done largely so as to avoid effects of fatigue in our older participants, since such effects could have confounded any age effects on sentence-processing abilities. However, small item numbers pose the risk of low reliability, which is particularly problematic if one is trying to assess the correlation between two measures with low reliability, as is often the case in studies assessing individual differences (Hedge, Powell, & Sumner, 2018). In the present study, Intraclass Correlation Coefficients (ICC) indicated varied reliability across the different conditions, falling on average on the lower end of "fair" reliability (see Table B2 in the Appendix). For this reason, caution is warranted especially when assessing the relationship between the cognitive variables (which themselves may be prone to lower reliability than is generally assumed) and performance in the sentence-completion task. Thus, future studies should strive to maximize the number of items per condition (and use individual-differences tests with high reliability), especially if they seek to correlate performance to individual differences measures.

Conclusion

The present study assessed native and non-native agreement processing across the lifespan. The grammatical phenomenon of interest was susceptibility to subject-verb agreement attraction errors, assessed using a sentence-completion task, while we also examined various cognitive abilities that have been argued to underlie sentence processing. Our results revealed some interesting differences between native and non-native agreement processing, both when comparing the two groups' developmental trajectories as well as across the full age range. First, unlike what we observed for our L1 participants, the L2 participants' performance actually improved with increasing age, providing an additional building block to the still understudied notion that aging is associated not just with declines, but also potentially with improvements. Second, it appears that even after decades of immersive exposure to a late-learned L2, highly-proficient L2 speakers do not behave like age-matched L1 speakers. Despite not differing in general measures of overall language performance (e.g., accuracy rates), our findings indicated differences in the finer-grained linguistic mechanisms underlying said performance, possibly because L2 speakers assigned shallower or less detailed structures to the subject NPs, compared to L1 speakers' more fully-fledged hierarchically-structured representations. Third, it appears that L2 agreement processing requires greater cognitive control of the sort assessed in the present study (e.g., WM, inhibitory control) than L1 processing.

CRediT authorship contribution statement

Jana Reifegerste: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Rebecca Jarvis:** Software, Formal analysis, Investigation. **Claudia Felser:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We are grateful to the individuals who participated in the experiments, and we thank Tiffany M. Dyson and Marilena Tsopanidi for their contributions to the data collection, as well as the members of the Potsdam Research Institute for Multilingualism, Adrian Staub, and two anonymous reviewers for valuable input. This work was supported by an Alexander-von-Humboldt-Professorship awarded to Harald Clahsen. The authors declare no competing financial interest.

Appendix A

Experimental sentences (and gloss). Note that unlike in the English gloss, in the German sentences noun number is encoded not just by the noun suffix, but also by the article. (Subject NP: singular *der* or *das*, plural *die*; attractor NP: singular *des*, plural *der*.)

1. *Der/die Napf/Näpfe des/der süßen Hundes/Hunde HAT/HABEN sehr stark gerochen.*
   The food bowl/food bowls of the cute dog/dogs HAS/HAVE smelled very strongly.
2. *Der/die Brief/ Briefe des/der diplomatischen Anwalts/ Anwälte HAT/HABEN zu viele Fehler.*
   The letter/letters from the diplomatic lawyer/lawyers HAS/HAVE too many mistakes.
3. *Der/die Redakteur/Redakteure des/der billigen Magazins/Magazine HAT/HABEN sehr hart gearbeitet.*
   The editor/editors of the cheap magazine/magazines HAS/HAVE worked very hard.
4. *Der/die Ertrag/Erträge des/der globalen Konzerns/Konzerne HAT/HABEN die Mitarbeiter erfreut.*
   The revenue/revenues of the global corporation/corporations HAS/HAVE delighted the employees.
5. *Das/die Ende/Enden des/der blauen Fadens/Fäden HAT/HABEN den Jungen geärgert.*
   The end/ends of the blue thread/threads HAS/HAVE annoyed the boy.
6. *Der/die Erfolg/Erfolge des/der bekannten Films/Filme HAT/HABEN die Kritiker überrascht.*
   The success/successes of the popular movie/movies HAS/HAVE surprised the critics.
7. *Das/die Foto/Fotos des/der schönen Pferdes/Pferde HAT/HABEN zu viel gekostet.*
   The photo/photos of the beautiful horse/horses HAS/HAVE cost too much.
8. *Der/die Plan/Pläne des/der neugierigen Detektivs/Detektive HAT/HABEN leider nicht funktioniert.*
   The plan/plans of the curious detective/detectives HAS/HAVE unfortunately not worked.
9. *Das/die Geräusch/Geräusche des/der wilden Tieres/Tiere HAT/HABEN die Kinder erschreckt.*
   The noise/noises of the wild animal/animals HAS/HAVE startled the children.
10. *Der/die Zaun/Zäune des/der grünen Gartens/Gärten HAT/HABEN viele kleine Löcher.*
    The fence/fences of the green garden/gardens HAS/HAVE many small holes.
11. *Der/die Gewinn/Gewinne des/der erfolgreichen Betriebs/Betriebe HAT/HABEN den Chef erfreut.*
    The profit/profits of the successful business/businesses HAS/HAVE delighted the boss.
12. *Das/die Bild/Bilder des/der teuren Bootes/Boote HAT/HABEN viele schöne Farben.*
    The image/images of the expensive boat/boats HAS/HAVE many nice colors.
13. *Der/die Angriff/Angriffe des/der vorsichtigen Feindes/Feinde HAT/HABEN alle sehr überrascht.*
    The attack/attacks of the careful enemy/enemies HAS/HAVE surprised everyone.
14. *Das/die Buch/Bücher des/der berühmten Professors/Professoren HAT/HABEN viele schöne Bilder.*
    The book/books by the famous professor/professors HAS/HAVE many nice images.
15. *Das/die Auto/Autos des/der faulen Sohnes/Söhne HAT/HABEN sehr schmutzige Scheiben.*
    The car/cars of the lazy son/sons HAS/HAVE very dirty windows.
16. *Der/die Moderator/Moderatoren des/der langweiligen Programms/Programme HAT/HABEN gar keinen Humor.*
    The host/hosts of the boring show/shows HAS/HAVE absolutely no humor.
17. *Der/die Roman/Romane des/der begabten Autors/Autoren HAT/HABEN viele interessierte Leser.*
    The novel/novels by the talented author/authors HAS/HAVE many interested readers.
18. *Der/die Koch/Köche des/der mächtigen Königs/König HAT/HABEN wohl kaum Erfahrung.*
    The cook/cooks of the powerful king/kings HAS/HAVE apparently little experience.

19. *Der/die Parkplatz/Parkplätze des/der neuen Supermarkts/Supermärkte HAT/HABEN eine hohe Schranke.*
    The parking lot/parking lots of the new supermarket/supermarkets HAS/HAVE a high barrier.
20. *Der/die Stoff/Stoffe des/der modischen Hutes/Hüte HAT/HABEN mich nicht überzeugt.*
    The fabric/fabrics of the stylish hat/hats HAS/HAVE not convinced me.
21. *Der/die Geruch/Gerüche des/der schmutzigen Stalls/Ställe HAT/HABEN den Urlauber gestört.*
    The smell/smells of the dirty stable/stables HAS/HAVE disturbed the vacationer.
22. *Das/die Problem/Probleme des/der engen Freundes/Freunde HAT/HABEN mich sehr besorgt.*
    The problem/problems of the close friend/friends HAS/HAVE greatly worried me.
23. *Das/die Video/Videos des/der älteren Mannes/Männer HAT/HABEN uns alle amüsiert.*
    The video/videos of the older man/men HAS/HAVE amused all of us.
24. *Der/die Architekt/Architekten des/der modernen Museums/Museen HAT/HABEN viele gute Ideen.*
    The architect/architects of the modern museum/museums HAS/HAVE many good ideas.
25. *Der/die Vorschlag/Vorschläge des/der klugen Direktors/Direktoren HAT/HABEN leider niemanden interessiert.*
    The suggestion/suggestions by the smart manager/managers HAS/HAVE not interested anyone.
26. *Der/die Auftritt/Auftritte des/der talentierten Chores/Chöre HAT/HABEN das Publikum beeindruckt.*
    The performance/performances of the talented choir/choirs HAS/HAVE impressed the audience.
27. *Das/die Spielzeug/Spielzeuge des/der kleinen Kindes/Kinder HAT/HABEN viel Freude bereitet.*
    The toy/toys of the small child/children HAS/HAVE made them happy.
28. *Das/die Konzert/Konzerte des/der beliebten Tenors/Tenöre HAT/HABEN alle Zuschauer begeistert.*
    The concert/concerts of the popular tenor/tenors HAS/HAVE excited all spectators.
29. *Das/die Gespräch/Gespräche des/der unerfahrenen Lehrlings/Lehrlinge HAT/HABEN für Ärger gesorgt.*
    The conversation/conversations of the inexperienced trainee/trainees HAS/HAVE caused some trouble.
30. *Der/der Kapitän/Kapitäne des/der defekten Schiffs/Schiffe HAT/HABEN nicht viel Ahnung.*
    The captain/captains of the defective ship/ships HAS/HAVE no clue.
31. *Das/die Geschenk/Geschenke des/der freundlichen Mönchs/Mönche HAT/HABEN uns sehr erfreut.*
    The gift/gifts from the friendly monk/monks HAS/HAVE delighted us a lot.
32. *Das/Die Ergebnis/Ergebnisse des/der spannenden Spiels/Spiele HAT/HABEN alle sehr überrascht.*
    The result/results from the exciting game/games HAS/HAVE greatly surprised everybody.
33. *Der/die Anruf/Anrufe des/der erfahrenen Arztes/Ärzte HAT/HABEN den Patienten beruhigt.*
    The call/calls from the experienced doctor/doctors HAS/HAVE calmed the patient down.
34. *Der/die Acker/Äcker des/der fleißigen Landwirts/Landwirte HAT/HABEN viel Arbeit gemacht.*
    The field/fields of the hardworking farmer/farmers HAS/HAVE meant a lot of work.
35. *Der/die Wunsch/Wünsche des/der jungen Vaters/Väter HAT/HABEN nur selten Vorrang.*
    The wish/wishes of the young father/fathers HAS/HAVE only rarely priority.
36. *Das/die Resultat/Resultate des/der wichtigen Projekts/Projekte HAT/HABEN den Forscher interessiert.*
    The result/results of the important project/projects HAS/HAVE interested the researcher.
37. *Der/die Hinweis/Hinweise des/der schlauen Generals/Generäle HAT/HABEN allen sehr geholfen.*
    The tip/tips of the smart general/generals HAS/HAVE helped everybody a lot.
38. *Der/die Hintergrund/Hintergründe des/der schlimmen Mordes/Morde HAT/HABEN die Polizei verwirrt.*
    The background/backgrounds of the terrible murder/murders HAS/HAVE confused the police.
39. *Das/die Gebet/Gebete des/der weisen Bischofs/Bischöfe HAT/HABEN wohl nicht geholfen.*
    The prayer/prayers of the wise bishop/bishops HAS/HAVE apparently not helped.
40. *Der/die Ausflug/Ausflüge des/der müden Gastes/Gäste HAT/HABEN doch Spaß gemacht.*
    The excursion/excursions of the tired guest/guests HAS/HAVE been fun after all.
41. *Der/die Bericht/Berichte des/der erfahrenen Kommissars/Kommissare HAT/HABEN die Arbeit erleichtert.*
    The report/reports of the experienced commissioner/commissioners HAS/HAVE facilitated the work.
42. *Das/die Büro/Büros des/der ehrgeizigen Ingenieurs/Ingenieure HAT/HABEN sehr große Fenster.*
    The office/offices of the ambitious engineer/engineers HAS/HAVE very big windows.
43. *Der/die Eingang/Eingänge des/der großen Hauses/Häuser HAT/HABEN sehr große Türen.*
    The entrance/entrances of the big house/houses HAS/HAVE very big doors.
44. *Der/die Klang/Klänge des/der alten Instruments/Instrumente HAT/HABEN die Zuhörer begeistert.*
    The sound/sounds of the old instrument/instruments HAS/HAVE excited the listeners.
45. *Das/die Ergebnis/Ergebnisse des/der anstrengenden Kampfes/Kämpfe HAT/HABEN niemanden mehr überrascht.*
    The result/results of the exhausting fight/fights HAS/HAVE not surprised anyone.
46. *Der/die Kern/Kerne des/der harten Pfirsichs/Pfirsiche HAT/HAS große Schmerzen verursacht.*
    The pit/pits of the hard peach/peaches HAS/HAVE caused a lot of pain.
47. *Der/die Einwand/Einwände des/der aufmerksamen Notars/Notare HAT/HABEN den Vater verärgert.*
    The objection/objections by the attentive notary/notaries HAS/HAVE angered the father.
48. *Das/die Telefonat/Telefonate des/der netten Pastors/Pastoren HAT/HABEN alle Fragen geklärt.*
    The phone call/phone calls from the nice priest/priests HAS/HAVE clarified all questions.

Appendix B

See Fig. B1 for performance in the sentence-completion task by condition and age, including data points for each participant by condition. See Tables B1 and B2 for descriptive performance data by age group and split-half reliability metrics.

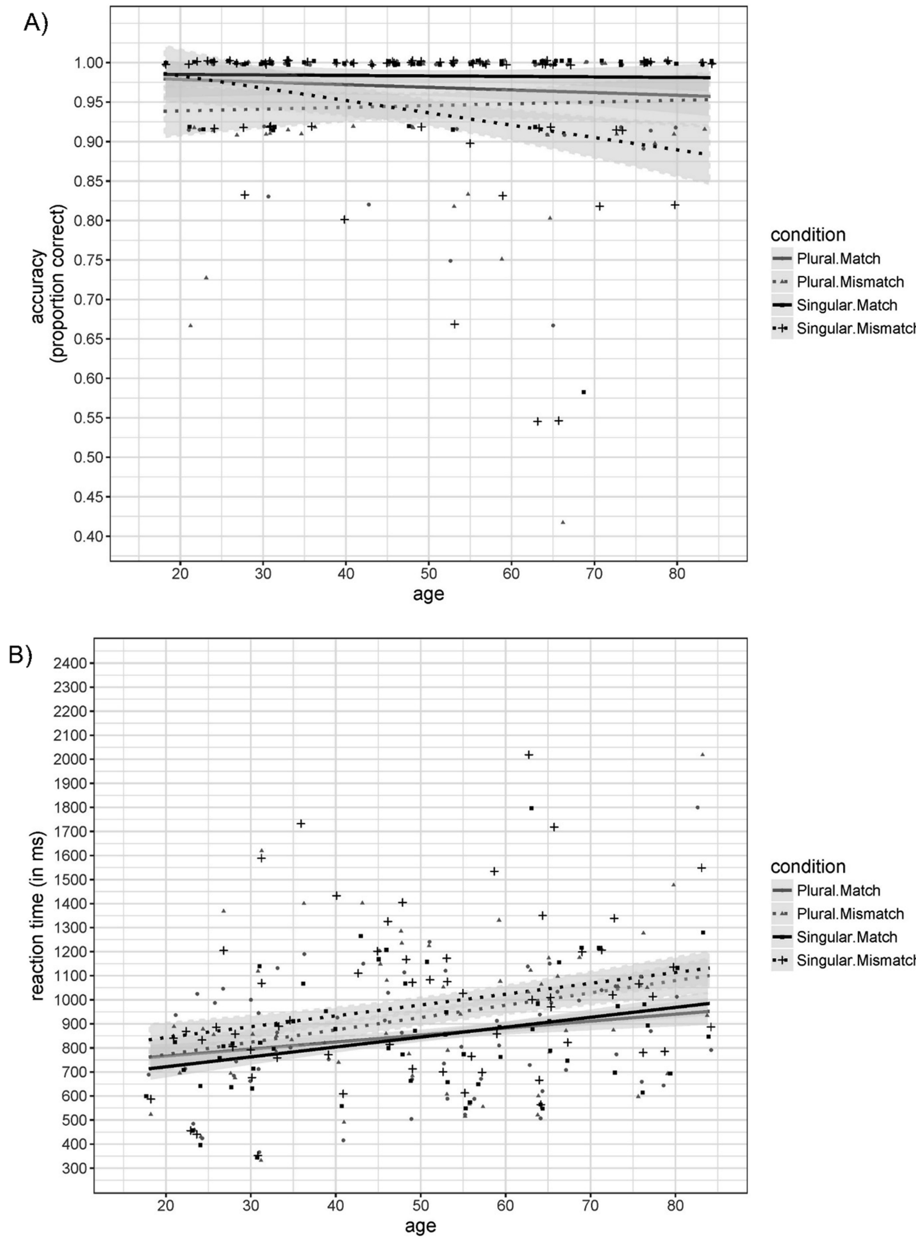

Fig. B1. Performance (A: L1 accuracy rates; B: L1 RTs; C: L2 accuracy rates; D: L2 RTs) in the sentence-completion task as a function of age, across conditions (lines) and broken down by participant by condition (data points).

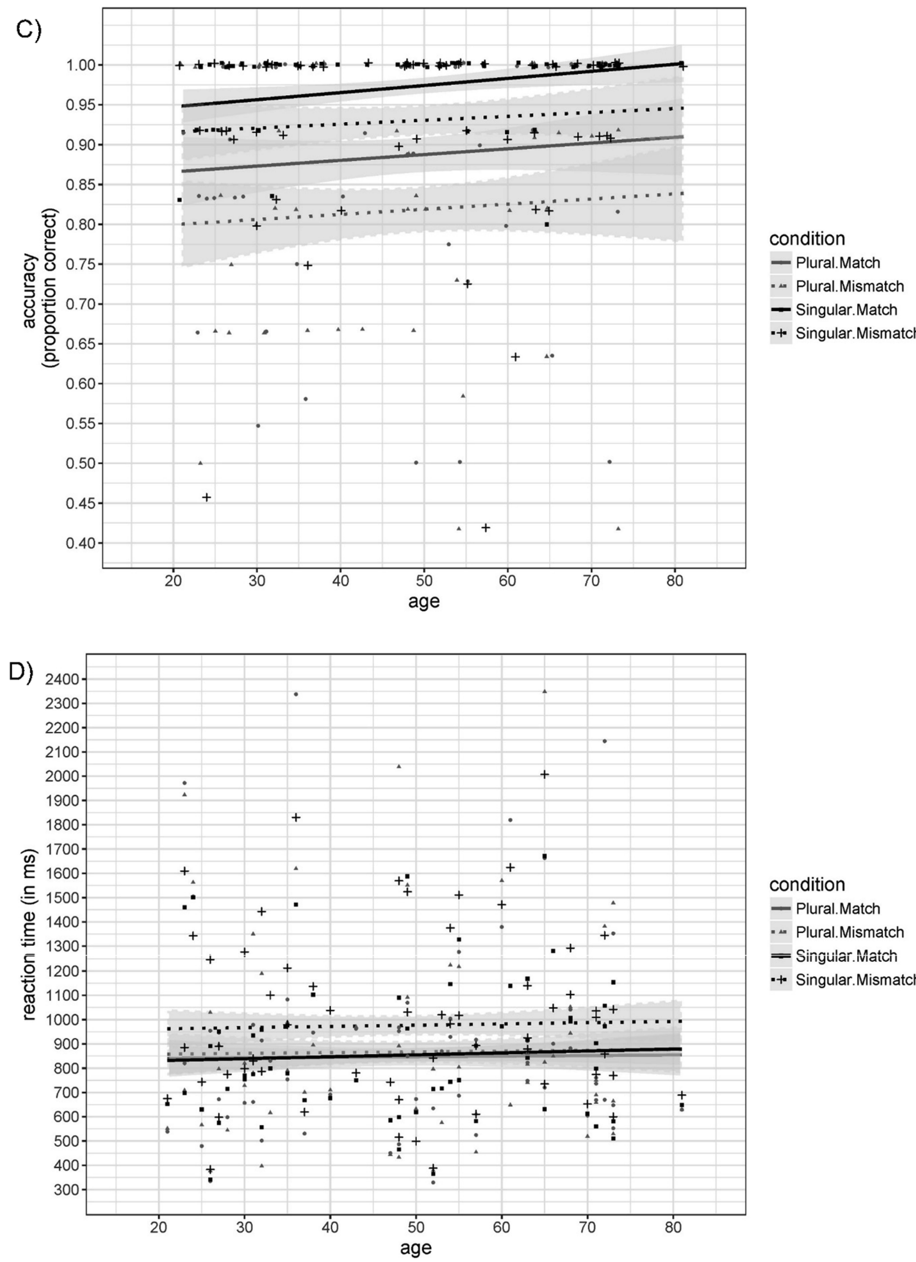

Fig. B1. (*continued*)

Table B1
Descriptive accuracy and RT data, broken down by language group, condition, and age group (younger adults, middle-aged adults, older adults).

			younger adults	middle-aged adults	older adults	Average
L1	Accuracy	Singular-Match	0.983 (0.129)	0.992 (0.091)	0.975 (0.156)	0.983 (0.128)
		Singular-Mismatch	0.970 (0.171)	0.957 (0.203)	0.882 (0.324)	0.937 (0.244)
		Plural-Match	0.979 (0.144)	0.970 (0.170)	0.957 (0.203)	0.969 (0.174)
		Plural-Mismatch	0.936 (0.246)	0.962 (0.192)	0.939 (0.239)	0.946 (0.227)
		Average	0.967 (0.178)	0.970 (0.170)	0.939 (0.240)	0.959 (0.199)
	RTs	Singular-Match	746 (313)	852 (347)	936 (415)	845 (369)
		Singular-Mismatch	866 (470)	1008 (447)	1062 (437)	975 (459)
		Plural-Match	806 (395)	833 (385)	921 (419)	852 (402)
		Plural-Mismatch	839 (462)	915 (457)	1024 (491)	926 (475)
		Average	813 (415)	901 (416)	983 (444)	0.902 (0.297)
L2	Accuracy	Singular-Match	0.946 (0.226)	0.995 (0.069)	0.983 (0.130)	0.973 (0.162)
		Singular-Mismatch	0.925 (0.264)	0.927 (0.261)	0.938 (0.241)	0.930 (0.255)
		Plural-Match	0.872 (0.335)	0.885 (0.320)	0.904 (0.295)	0.887 (0.317)
		Plural-Mismatch	0.831 (0.375)	0.780 (0.415)	0.837 (0.370)	0.818 (0.386)
		Average	0.894 (0.308)	0.898 (0.303)	0.916 (0.278)	898 (430)
	RTs	Singular-Match	847 (455)	804 (468)	907 (437)	854 (455)
		Singular-Mismatch	975 (546)	928 (544)	1022 (530)	977 (540)
		Plural-Match	857 (610)	761 (423)	910 (547)	947 (542)
		Plural-Mismatch	847 (537)	861 (619)	893 (519)	867 (555)
		Average	882 (540)	838 (517)	933 (510)	887 (524)

Table B2
Overview of split-half reliabilities (odd vs. even item numbers), calculated as Intraclass Correlation Coefficients (*ICC*).

	L1 accuracy	L2 accuracy	L1 RTs	L2 RTs
Singular attraction effect	*ICC* = .563	*ICC* = .370	*ICC* = .404	*ICC* = .385
Plural attraction effect	*ICC* = .405	*ICC* = .381	*ICC* = .323	*ICC* = .450

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.104083.

References

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition, 19*(4), 689–698. https://doi.org/10.1017/S1366728916000225.

Acuña-Fariña, J. C. (2012). Agreement, attraction and architectural opportunism. *Journal of Linguistics, 48*(2), 257–295.

Altmann, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Language and Cognitive Processes, 21*(1–3), 322–354. https://doi.org/10.1080/0169096054400006.

American Council on the Teaching of Foreign Languages. (2015). Assigning CEFR Ratings to ACTFL Assessments. Alexandria, VA. Retrieved from https://www.actfl.org/publications/additional-resources/assigning-cefr-ratings-actfl-assessments.

Armstrong, A., Bulkes, N. Z., & Tanner, D. S. (2016). *ERPs show processing similarities of English subject- verb agreement violations in native speakers and L1 Chinese learners. Poster Presented at the 2016 Neurobiology of Language Conference, London, UK*.

Au, R., Joung, P., Nicholas, M., Obler, L. K., Kass, R., & Albert, M. L. (1995). Naming ability across the adult life span. *Aging and Cognition, 2*(4), 300–311. https://doi.org/10.1080/13825589508256605.

Baayen, R. H. (2013). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics." R package version 1.4.1.

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*, 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/10.1287/mksc.12.4.395.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database. Linguistic data consortium.* Philadelphia: University of Pennsylvania.

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*, 65–85. https://doi.org/10.1016/j.jml.2006.08.004.
Bak, T. H., Vega-Mendoza, M., & Sorace, A. (2014). Never too late? An advantage on tests of auditory attention extends to late bilinguals. *Frontiers in Psychology, 5*, 1–6. https://doi.org/10.3389/fpsyg.2014.00485.
Baker, C., & Prys Jones, S. (1998). *Encyclopedia of bilingualism and bilingual education.* Clevedon, UK: Multilingual Matters.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001.
Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.
Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia, 45*(2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009.
Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859.
Bock, K. (2004). Psycholinguistically speaking: Some matters of meaning, marking and morphing. In B. H. Ross (Vol. Ed.), *The psychology of learning and motivation: Vol. 44*, (pp. 109–144). San Diego, CA: Elsevier. https://doi.org/10.1016/S0079-7421(03)44004-8.
Bock, K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language, 66*(1), 17–37. https://doi.org/10.1016/j.jml.2011.07.008.
Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language, 31*(1), 99–127. https://doi.org/10.1016/0749-596X(92)90007-K.
Bock, K., & Eberhard, K. M. (1993). Meaning, sound and Syntax in English number agreement. *Language and Cognitive Processes, 8*(1), 57–99. https://doi.org/10.1080/01690969308406949.
Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45–93.
Bock, K., Nicol, J. L., & Cutting, J. C. (1999). The ties that bind: creating number agreement in speech. *Journal of Memory and Language, 40*, 330–346.
Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia, 51*(13), 2492–2511. https://doi.org/10.1016/j.neuropsychologia.2013.09.004.
Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychology and Aging, 23*(2), 366–376.
Brébion, G. (2001). Language processing, slowing, and speed/accuracy trade-off in the elderly. *Experimental Aging Research, 27*, 137–150.
Caplan, D., & Waters, G. S. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory, 13*(3–4), 403–413. https://doi.org/10.1080/09658210344000459.
Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*(2), 161–174. https://doi.org/10.1017/S136672890700291X.
Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(01), 3–42 https://doi.org/10.1017.S0142716406060024.
Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*, 564–570. https://doi.org/10.1016/j.tics.2006.10.002.
Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition, 40*(3), 693–706.
Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning, 60*(1), 21–43.
Clahsen, H., & Reifegerste, J. (2017). Morphological processing in old-age bilinguals. In M. Libben, M. Goral, & G. Libben (Eds.). *Bilingualism: A Framework for Understanding the Mental Lexicon* (pp. 217–248). Amsterdam: Benjamins.
Cohen-Shikora, E. R., & Balota, D. A. (2016). Visual word recognition across the adult lifespan. *Psychology and Aging, 31*(5), 488–502. https://doi.org/10.1037/pag0000100.
Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 59*(5), P203–P209. https://doi.org/10.1093/geronb/59.5.P203.
Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.* McGill University.
Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition, 20*(4), 659–678. https://doi.org/10.1017/S1366728916000675.
Davidson, D. J., Zacks, R. T., & Ferreira, F. (2003). Age preservation of the syntactic processor in production. *Journal of Psycholinguistic Research, 32*(5), 541–566. https://doi.org/10.1023/A:1025402517111.
Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language, 60*(1), 112–143. https://doi.org/10.1016/j.jml.2008.07.001.
Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching, 31*, 117–135. https://doi.org/10.1017/S026144480001315X.
Dronjic, V. (2013). *Concurrent memory load, working memory span, and morphological processing in L1 and L2 English.* University of Toronto.
du Boisgueheneuc, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., ... Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. *Brain, 129*(12), 3315–3328. https://doi.org/10.1093/brain/awl244.
Eberhard, K. M. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language, 36*, 147–164.
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.112.3.000.
Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149.
Facal, D., Juncos-Rabadán, O., Rodríguez, M. S., & Pereiro, A. X. (2012). Tip-of-the-tongue in aging: Influence of vocabulary, working memory and processing speed. *Aging Clinical and Experimental Research, 24*(6), 647–656. https://doi.org/10.3275/8586.
Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State" A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.
Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics, 32*, 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.
Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *The Journal of Neuroscience, 31*(47), 17242–17249. https://doi.org/10.1523/JNEUROSCI.0309-11.2011.
Franck, J., Vigliocco, G., & Nicol, J. L. (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.
Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001.
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377.
Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191–215. https://doi.org/10.1177/0267658312437990.
Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121–134.
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.
Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences, 115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115.
Hartsuiker, R. J., Antón-Méndez, I., & Van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language, 45*(4), 546–572. https://doi.org/10.1006/jmla.2000.2787.
Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject–verb agreement. *Memory & Cognition, 31*(8), 1316–1326.
Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: Evidence From subject-verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 891–904. https://doi.org/10.1037/0278-7393.31.5.891.
Häussler, J. (2012). *The emergence of attraction errors during sentence comprehension.* University of Konstanz.
Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.
Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22*(3), 369–397. https://doi.org/10.1191/0267658306sr272oa.
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.
Jacob, G., Fleischhauer, E., & Clahsen, H. (2013). Allomorphy and affixation in morphological processing: A cross-modal priming study with late bilinguals. *Bilingualism: Language and Cognition, 16*(4), 924–933. https://doi.org/10.1017/S1366728913000291.
Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603–634 https://doi.org/10.1017.S0142716404001298.
Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology, 7*, 703. https://doi.org/10.3389/fpsyg.2016.00703.
Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and non-native Spanish: An eye-movement investigation. *Language Learning, 59*(3), 503–535.
Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics, 7*(3), 277–288. https://doi.org/10.1017/S0142716400007578.
Kemper, S. (1988). Geriatric psycholinguistics: Syntactic limitations of oral and written language. In L. L. Light, & D. M. Burke (Eds.). *Language, memory, and aging* (pp. 58–76). Cambridge, United Kingdom: Cambridge University Press.
Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging, 12*(2), 362–371. https://doi.org/10.1037//0882-7974.12.2.362.
Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the

crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology, 68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560.
Krampe, R. T., & Charness, N. (2006). Aging and Expertise. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.). *The Cambridge handbook of expertise and expert performance* (pp. 723–742). New York, NY: Cambridge University Press.
Kynette, D., & Kemper, S. (1986). Aging and the loss of grammatical forms: A cross-sectional study of language performance. *Language and Communication, 6*(1–2), 65–72.
Lago, S., & Felser, C. (2018). Agreement attraction in native and non-native speakers of German. *Applied Psycholinguistics, 39*(1), 619–647.
Lago, S., Gra, M., Şafak, D. F., Demir, O., Kırkıcı, B., & Felser, C. (2019). Straight from the horse's mouth: Agreement attraction effects with Turkish possessors. *Linguistic Approaches to Bilingualism, 9*(3), 398–426. https://doi.org/10.1075/lab.17019.lag.
Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language, 82*, 133–149. https://doi.org/10.1016/j.jml.2015.02.002.
Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14*(4), 359–375. https://doi.org/10.1191/026765898672500216.
Lawrence, B., Myerson, J., & Hale, S. (1998). Differential decline of verbal and visuospatial processing speed across the adult life span. *Aging, Neuropsychology, and Cognition, 5*(2), 129–146.
Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419. https://doi.org/10.1207/s15516709cog0000_25.
Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Developmental Psychology, 22*(4), 580–585. https://doi.org/10.1037/0012-1649.22.4.580.
Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics, 36*, 1283–1315.
Lima, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and Aging, 6*(3), 416–425.
Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes, 23*(6), 769–799. https://doi.org/10.1080/01690960701774182.
Madden, D. J. (1992). Four to Ten milliseconds per year: Age-related slowing of visual word identification. *Journal of Gerontology: Psychological Sciences, 47*(2), 59–68.
Mallinson, G., & Blake, B. (1981). *Language typology: Crosslinguistic studies in syntax.* Amsterdam: North Holland.
McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006.
McDonald, J. L., & Roussel, C. C. (2010). Past tense grammaticality judgment and production in non-native and stressed native English speakers. *Bilingualism: Language and Cognition, 13*(4), 429–448. https://doi.org/10.1017/S1366728909990599.
McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111–123.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91.
Mueller, S. T. (2012). The Psychology Experiment Building Language. Retrieved from http://pebl.sourceforge.net/.
Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods, 222*, 250–259. https://doi.org/10.1007/s11103-011-9767-z.Plastid.
Myerson, J., Ferraro, F. R., Hale, S., & Lima, S. D. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging, 7*(2), 257–270.
Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech, 48*(2), 123–156. https://doi.org/10.1177/00238309050480020101.
Nicol, J. L. (1995). Effects of clausal structure on subject-verb agreement errors. *Journal of Psycholinguistic Research, 24*(6), 507–516.
Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject-Verb agreement processes in comprehension. *Journal of Memory and Language, 36*, 569–587. https://doi.org/10.1006/jmla.1996.2497.
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (1998). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*(3), 263–281.
Obler, L. K., Fein, D., Nicholas, M., & Albert, M. L. (1991). Auditory comprehension and aging: Decline in syntactic processing. *Applied Psycholinguistics, 12*, 433–452. https://doi.org/10.1017/S0142716400005865.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97–113.
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.
Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2010). Neural processing during older adults' comprehension of spoken sentences: Age differences in resource allocation and connectivity. *Cerebral Cortex, 20*(4), 773–782. https://doi.org/10.1093/cercor/bhp142.
Phillips, C. (2013). Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes, 28*(1–2), 156–187. https://doi.org/10.1080/01690965.2010.530960.
Poeppel, D., & Wexler, K. (1993). The full competence hypothesis of clause structure in early German. *Language, 69*, 1–33.
Prehn, K., Taud, B., Reifegerste, J., Clahsen, H., & Flöel, A. (2018). Neural correlates of grammatical inflection in older native and second-language speakers. *Bilingualism: Language and Cognition, 21*(1), 1–12. https://doi.org/10.1017/S1366728916001206.
Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology, 70*, 305–311.
Randolph, C., Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Hermann, B. P. (1999). Determinants of confrontation naming performance. *Archives of Clinical Neuropsychology, 14*(6), 489–496. https://doi.org/10.1016/S0887-6177(98)00023-7.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108.
Ratcliff, R., & McKoon, G. (2015). Aging effects in item and associative recognition memory for pictures and words. *Psychology and Aging, 30*(3), 669–674. https://doi.org/10.1037/pag0000030.
Ratcliff, R., Spieler, D. H., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review, 7*(1), 1–25.
Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the Lexical Decision Task. *Psychology and Aging, 19*(2), 278–289. https://doi.org/10.1037/0882- 7974.19.2.278.
Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the Diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging, 22*(1), 56–66. https://doi.org/10.1037/0882-7974.22.1.56.
Rechel, B., Grundy, E., Robine, J.-M., Cylus, J., Mackenbach, J. P., Knai, C., & McKee, M. (2013). Ageing in the European Union. *The Lancet, 381*(9874), 1312–1322. https://doi.org/10.1016/S0140-6736(12)62087-X.
Reifegerste, J. (2014). *Morphological processing in younger and older people: Evidence for flexible dual-route access.* Radboud University Nijmegen.
Reifegerste, J., Elin, K., & Clahsen, H. (2019). Persistent differences between native speakers and bilinguals: Evidence from inflectional and derivational processing in older speakers. *Bilingualism: Language and Cognition, 22*, 425–440.
Reifegerste, J., & Felser, C. (2017). Effects of aging on interference during pronoun resolution. *Journal of Speech, Language, and Hearing Research, 60*(12), https://doi.org/10.1044/2017_JSLHR-L-17-0183.
Reifegerste, J., Hauer, F., & Felser, C. (2017). Agreement processing and attraction errors in aging: Evidence from subject-verb agreement in German. *Aging, Neuropsychology, and Cognition, 24*(6), 672–702. https://doi.org/10.1080/13825585.2016.1251550.
Reifegerste, J., Meyer, A. S., & Zwitserlood, P. (2017). Inflectional complexity and experience affect plural processing in younger and older readers of Dutch and German. *Language, Cognition and Neuroscience, 32*(4), 471–487. https://doi.org/10.1080/23273798.2016.1247213.
Ristić, B., Molinaro, N., & Mancini, S. (2016). Agreement attraction in Serbian. *The Mental Lexicon, 11*(2), 242–276. https://doi.org/10.1075/ml.11.2.04ris.
Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging, 20*(1), 3–18. https://doi.org/10.1037/0882-7974.20.1.3.
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.
Salthouse, T. A. (1992). Influence of processing speed on adult age differences in working memory. *Acta Psychologica, 79*, 155–170. https://doi.org/10.1016/0001-6918(92)90030-H.
Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403–428. https://doi.org/10.1037/0033-295X.103.3.403.
Sato, M., & Felser, C. (2010). Sensitivity to morphosyntactic violations in English as a second language sensitivity to morphosyntactic violations in English as a second language. *Second Language, 9*, 101–118.
Schel, M. A., Kühn, S., Brass, M., Haggard, P., Ridderinkhof, K. R., & Crone, E. A. (2014). Neural correlates of intentional and stimulus-driven inhibition: A comparison. *Frontiers in Human Neuroscience, 8*, 27. https://doi.org/10.3389/fnhum.2014.00027.
Schlueter, A. L. Z. (2017). *Memory retrieval in parsing and interpretation.* University of Maryland.
Shibuya, M., & Wakabayashi, S. (2008). Why are L2 learners not always sensitive to subject-verb agreement? *EUROSLA Yearbook, 8*(1), 235–258.
Slevc, L. R., & Martin, R. C. (2016). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology, 28*(7), 773–790. https://doi.org/10.1016/j.jml.2006.08.004.
Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject-verb number agreement. *Cognitive Science, 42*, 1043–1074. https://doi.org/10.1111/cogs.12591.
Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging, 25*(2), 377–390. https://doi.org/10.1037/a0018022.The.
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. S., & Bulkes, N. Z. (2015). Cues, quantification, and agreement in language comprehension. *Psychonomic Bulletin and Review, 22*(6), 1753–1763. https://doi.org/10.3758/s13423-015-0850-3.
Tanner, D. S., Nicol, J. L., & Brehm, L. (2014). 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.
Tanner, D. S., Nicol, J. L., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. *Proceedings of the 36th Boston university conference on language development*

(pp. 594–606). Somerville, MA: Cascadilla Press.
Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan, & F. Restle (Vol. Eds.), *Cognitive Theory: Vol. III*, (pp. 200–239). Hillsdale, NJ: Erlbaum.
Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge: Cambridge University Press.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5.
Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology, 6*, 347. https://doi.org/10.3389/fpsyg.2015.00347.
Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(1), 105–122.
Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.). *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141–178). Washington, DC: Georgetown University Press.
Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In G. Hickok, & S. Small (Eds.). *The neurobiology of language* (pp. 953–968). Amsterdam: Elsevier.
Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/ procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research, 34*(1), 39–65. https://doi.org/10.1177/0267658316675195.
Van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods, 39*(4), 973–978. https://doi.org/10.3758/BF03192992.
VanPatten, B., Keating, G. D., & Leeser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism, 2*, 109–140. https://doi.org/10.1075/lab.2.2.01pat.
Veenstra, A., Acheson, D. J., Bock, K., & Meyer, A. S. (2014). Effects of semantic integration on subject–verb agreement: Evidence from Dutch. *Language, Cognition and Neuroscience, 29*(3), 355–380. https://doi.org/10.1080/01690965.2013.862284.
Veenstra, A., Acheson, D. J., & Meyer, A. S. (2014). Keeping it simple: Studying grammatical encoding with lexically reduced item sets. *Frontiers in Psychology, 5*, 783. https://doi.org/10.3389/fpsyg.2014.00783.
Veenstra, A., Antoniou, K., Katsos, N., & Kissine, M. (2017). The Role of executive control in agreement attraction in monolingual and bilingual children. In M. LaMendola, & J. Scott (Eds.). *Proceedings of the 41st annual Boston University Conference on Language Development* (pp. 706–717). Somerville, MA: Cascadilla Press.
Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition, 137*, 40–46. https://doi.org/10.1016/j.cognition.2014.12.008.
Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging, 18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332.
Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge University Press.
Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61*(3), 261–298. https://doi.org/10.1016/S0010-0277(96)00713-5.
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*, 186–215.
Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*(4), 407–442. https://doi.org/10.1080/016909696387169.
Vigliocco, G., & Nicol, J. L. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition, 68*(1), B13–B29. https://doi.org/10.1016/S0010-0277(98)00041-9.
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.
Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging, 16*(1), 128–144. https://doi.org/10.1037//0882-7974.16.1.128.
Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M., & Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer's disease. *Bilingualism: Language and Cognition, 18*(3), 568–574. https://doi.org/10.1017/S136672891400087X.
Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72*, 37–48. https://doi.org/10.1016/j.jml.2013.12.003.
Yoon, J., Campanelli, L., Goral, M., Marton, K., Eichorn, N., & Obler, L. K. (2015). The effect of plausibility on sentence comprehension among older adults and its relation to cognitive functions. *Experimental Aging Research, 41*(3), 272–302. https://doi.org/10.1080/0361073X.2015.1021646.