Sina Bosch | João Veríssimo | Harald Clahsen

Inflectional morphology in bilingual language processing

An age-of-acquisition study

Suggested citation referring to the original publication:
Language Acquisition 26 (2019) 3, pp. 339–360
DOI https://doi.org/10.1080/10489223.2019.1570204
ISSN (print) 1048-9223
ISSN (online) 1532-7817

LANGUAGE ACQUISITION
2019, VOL. 26, NO. 3, 339–360
https://doi.org/10.1080/10489223.2019.1570204

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Inflectional morphology in bilingual language processing: An age-of-acquisition study

Sina Bosch, João Veríssimo, and Harald Clahsen

University of Potsdam

ABSTRACT
This study addresses the question of how age of acquisition (AoA) affects grammatical processing, specifically with respect to inflectional morphology, in bilinguals. We examined experimental data of more than 100 participants from the Russian/German community in Berlin, all of whom acquired Russian from birth and German at different ages. Using the cross-modal lexical priming technique, we investigated stem allomorphs of German verbs that encode multiple morphosyntactic features. The results revealed a striking AoA modulation of observed priming patterns, indicating efficient access to morphosyntactic features for early AoAs and a gradual decline with increasing AoAs. In addition, we found a discontinuity in the function relating AoA to morphosyntactic feature access, suggesting a sensitive period for the development of morphosyntax.

ARTICLE HISTORY
Received 7 March 2018
Accepted 18 December 2018

1. Introduction

Much research on grammatical processing in bilinguals has focused on the question of how the processing of grammatical phenomena differs between a native language (L1) acquired from birth and a second or nonnative language (L2) acquired later in life. Some researchers have claimed that L1 and L2 speakers employ the same system of linguistic representations and processing mechanisms for comprehending and producing sentences and morphologically complex words and that L1/L2 performance differences are essentially reducible to peripheral factors, such as decoding problems, working memory limitations, retrieval interference, slower processing speed (e.g., Cunnings 2017; McDonald 2006), difficulties with lexical access and retrieval (e.g., Hopp 2016), or a reduced ability to predict during L2 processing (see Kaan 2014). Other researchers have argued that reductionist accounts of this kind only provide partial explanations and have proposed more substantial differences between native and nonnative processing of morphological and syntactic phenomena. This line of research is most prominently represented by the Shallow-Structure Hypothesis (SSH)—originally proposed by Clahsen and Felser (2006a, 2006b)—and much subsequent research; see Clahsen and Felser (2017) for a recent review. The SSH holds that nonnative processing relies less on grammatical and more on nongrammatical information sources in comparison to native processing. A range of subtle L1/L2 differences obtained in studies of morphological processing have been attributed to this contrast, for example, that L2 learners exhibit native-like priming patterns for derived word forms but not for inflected forms (Kirkici & Clahsen 2013; Jacob, Heyer, & Veríssimo 2018) and that lexical constraints on word-formation processes affect L2 processing in the same way as L1 processing, unlike morphological constraints that showed reduced effects in L2 processing (Clahsen et al. 2013, 2015). Regarding the reason for *why* L2 processing differs from L1 processing, the SSH points to maturation, a sensitive period for the acquisition of grammar, as the crucial variable—

CONTACT Sina Bosch bosch@uni-potsdam.de University of Potsdam, Potsdam Research Institute for Multilingualism, 14476 Potsdam, Germany.

without, however, specifying how and from which age of acquisition (AoA) onwards a native speaker becomes nonnative. The current study addresses this question.

1.1. *AoA effects on language performance*

AoA has been found to be a crucial predictor of linguistic performance. Birdsong (2009:404) noted that "earlier is better," suggesting that later onsets of acquisition may lead to decreased levels of ultimate attainment. AoA effects have also featured prominently as evidence indicating a critical or sensitive period for language acquisition. However, despite several decades of research, serious controversies persist as to the existence, shape, and source of sensitive periods for second language acquisition. Even among those who acknowledge the existence of sensitive periods (e.g., Long 1990, 2013; Johnson & Newport 1989), there is considerable disagreement as to what domains the sensitive period applies to. On the one hand, proponents of some form of sensitive period for linguistic attainment have promoted an age-attainment function that resembles the shape of a sharp drop in ultimate proficiency during a relatively short time span around puberty, without further decline with increasing AoA (e.g., Johnson & Newport 1989; Flege 1999; DeKeyser 2000, 2010). On the other hand, previous research has also shown an age-attainment function that does not exhibit a steep slope around puberty but rather shows a gradual decline across the entire life span (e.g., Bialystok & Hakuta 1999; Bialystok & Miller 1999; Birdsong 2006; Hakuta et al. 2003). Pinker (1994:293), for example, claimed that "… acquisition of a normal language is guaranteed for children up to the age of six, is steadily compromised from then until shortly after puberty, and is rare thereafter." Similarly, according to Long (2013), sensitive periods for human language learning start at early developmental stages (not during puberty as often claimed), possibly even from birth onwards, and are followed by a gradual (not abrupt) decline before their eventual closure—a developmental trajectory that is reflected by a discontinuity in linguistic attainment as a function of AoA. Afterwards, the slope of the decline flattens for the rest of the life span and is thus less (if at all) noticeable. Such an AoA modulation of linguistic attainment would result in a "stretched Z" (Long 2013) or "stretched L" distribution of a sensitive period (Birdsong 2005; Birdsong & Vanhoeve 2016), which suggests that only child L2 learners seem capable of achieving full levels of L2 attainment.

If there are indeed sensitive periods of language development, the ability to acquire language seems susceptible to (possibly innately specified) maturational changes, a possibility that has raised general interest beyond the specific field of language acquisition research. Considerable attention in previous research has been devoted to grammatical skills and their potential modulation by AoA. To take a well-known study, Johnson & Newport (1989) reported a linear performance decline in Chinese and Korean L1 speakers' grammaticality judgements of English up to an AoA of 16 years, and no further AoA modulation thereafter; see also a number of follow-up grammaticality judgement studies that essentially replicated Johnson & Newport's (1989) findings (DeKeyser 2000; DeKeyser et al. 2010; Granena & Long 2013). Another source of evidence for AoA effects in the domain of grammar comes from studies that compared types of successive language acquisition at different ages of onset; see Meisel (2013) for a review of these studies. In spoken French, for example, subject clitic pronouns form a tight morphosyntactic dependency with the finite verb functioning much like subject-verb agreement markers. In line with that, subject clitic pronouns are not combined with nonfinite verbs in children who learn French from birth (e.g., **je conduire* … "I drive$_{+INF}$ … "). Yet adult L2 learners as well as successive child learners of French were found to produce such constructions after an AoA of about four years, a clear contrast taken to indicate an early sensitive phase for the acquisition of grammar. However, despite its theoretical significance, the empirical evidence supposedly supporting critical or sensitive periods in language development has not convinced everyone; see Birdsong (2014) and Mayberry & Kluender (2018) for reviews. One concern is that while robust linear declines are often reported to coincide with later AoAs, reliable discontinuities in how the AoA function maps onto grammatical skill—an important requirement for positing sensitive periods (Hakuta et al. 2003)—are much more difficult to demonstrate.

Furthermore, it is possible that what appear to be AoA effects on grammar are in fact effects of non-age-related factors such as reduced general language skill in a late-learned L2, the possibility of L1 transfer, decreased exposure, less practice and use, etc. How these learning factors contribute to L1/L2 performance differences and how they can be distinguished from genuine AoA effects is still subject to controversy.

A fresh look on the topic has recently been given by Veríssimo, Heyer, Jacob, & Clahsen (2018). Using the masked-priming technique, Veríssimo et al. investigated priming effects from inflected and derived forms (of German) in a large sample of Turkish/German bilinguals who were exposed to Turkish from birth and had varying ages of onset of German acquisition (range: 0 to 38 years). The study directly compared two morphological phenomena (of German) that on the surface appear to be highly similar, regular – *t* participle inflection (e.g., *geöffnet* 'opened') and productive – *ung* nominalization (e.g., *Öffnung* 'opening'), both used as (masked) primes on the same targets (e.g., *öffnen* '(to) open'). The results showed that derivational priming was unaffected by AoA, whereas inflectional priming gradually dissipated with increasing AoA of German, from 5 to 6 years of age onwards. Importantly, the modulation of inflectional priming by AoA was obtained even when other predictors of interest (viz. proficiency, length of exposure, and use of German) were controlled for, suggesting a genuine effect of AoA on grammatical processing in bilinguals. According to Veríssimo et al., these findings are indicative of a sensitive period for the acquisition of grammar, specifically the ability to extract inflectional rules from the input, which is progressively compromised after early childhood.

Against this background, the current study examines how AoA affects the way morphosyntactic information is accessed from complex lexical entries during online language processing. Morphosyntax involves inflectional processes that encode tense, number, case, person, and other grammatical functions. Previous research indicates that successful acquisition and efficient use of morphosyntax may be challenging areas for late learners, unlike for native L1 speakers (see White 2003). Hence, the study of morphosyntactic skill provides a promising opportunity for discovering genuine AoA effects. Using experimental data from a large sample of bilingual (Russian/German) speakers who acquired Russian from birth and German at different ages, we present new evidence for long-term consequences of AoA on grammatical processing in bilinguals.

1.2. *Morphosyntax in the German mental lexicon*

A prominent case from the Germanic languages of the representation of morphosyntactic information in the mental lexicon is stem allomorphy in verbs and nouns with specific morphosyntactic feature sets requiring marked (irregular) stems (e.g., English: *keep* – *kept* [+Past] or *child* – *children* [+Plur]. These kinds of marked stems represent a case of "lexically conditioned suppletive allomorphy" (Paster 2016:181), that is, they are not determined by regular phonological rules (hence "suppletive") and are idiosyncratic to particular lexical items (hence "lexically conditioned"). In German, there are about 200 base verbs with marked stems, which are selected for various inflected verb forms encoding a range of morphosyntactic features, for example, past tense, past participle, and subjunctive (*sterb*- 'die-', [+Past]: *starb*-, [+Part.]: (*ge*)*storb*(*en*), [+Subj.]: *stürb*-). The particular phenomenon we examined is marked verbal stems of so-called strong and mixed verbs, specifically the two variants of so-called secondary present-tense stems (Wiese 2008), which either have a fronted or a raised stem vowel (e.g., *schlaf*- → *schläf*- 'sleep,' *geb*- → *gib*- 'give'). Among present-tense forms, these marked stems are required for second and third person singular indicative forms paired with the corresponding regular suffixes -*st* and -*t*, for example *schläf-st* '(you) sleep' or *gib-t* '(s/he) gives.' While these marked stems can be historically derived from phonological ("ablaut") rules, these rules are unproductive in modern German with many exceptions (Durrell 1980, 2001; Wiese 1996).

As regards the representation of morphosyntactic information and specifically of verbal stem allomorphy in the German mental lexicon, we can distinguish between associative and hierarchically structured approaches. The first type of approach emphasizes that stem morphemes (both marked and unmarked ones) are associatively connected by semantic and/or phonological links encoding degrees of similarity. Bittner (1996) and Köpcke (1998), for example, observed that the stem-alternation patterns of the strong and mixed verbs of German form phonological similarity clusters. For instance, unmarked stems with a medial high front vowel /ɪ/ and a velar nasal /ŋ/ (e.g., *singen* 'to sing,' *sinken* 'to sink,' *ringen* 'to wrestle,' *wringen* 'to wring,' etc.) have marked stems with the same vowel changes. Smolka et al. (2007), on the other hand, focused on semantic similarity and noted that the various stem forms of strong verbs in German (e.g., *sterb-*, *starb-*, (*ge*)*storb*(*en*), *stürb-*) form a semantic cluster as they activate a shared concept (e.g., "die"). What is common among these proposals is that stem forms are supposed to constitute associative clusters held together by similarity.

A second approach conceives of the lexicon not only as associative links but also as sets of hierarchically structured entries and lexical templates, for example, for verbs with marked stems. One implementation of this idea comes from Wunderlich (1996; Wunderlich & Fabri, 1995), who specifically applied the notion of default inheritance networks (Corbett & Fraser 1993; Hippisley 2016) to German. The purpose of these representations is to capture the morphological relationships between lexical items and at the same time to prevent the lexicon from listing redundant lexical information. The main idea of this implementation is that stem variants (e.g., *stirb-* and *starb-*) of the same lexeme (e.g., *sterben*) are not repeatedly listed in the mental lexicon but that some stems have an impoverished, i.e., underspecified, lexical entry with minimally specified analyses. When these forms are used, their full form and interpretation is complemented from the properties of the base entry. An example illustration of these kinds of structured entries is shown in Figure 1 for the various stems of the German verb *sterben* 'to die.' The base stem (*sterb-*) at the top is the unmarked, least specified stem form and represents the mother node of the inheritance hierarchy tree implementation. The subnodes inherit all information from their respective mother node except for the features they replace or add. Hence, to avoid redundancy they are only specified for phonological changes and/or morphosyntactic feature values. The leftmost subnode is specified for the vowel change (e.g., *sterb-* → *stirb-*) plus the feature [−1] for second and third person; the imperative form (+IMP) is inherited from this subnode, capturing the fact that strong verbs that have marked stems in the imperative also have marked second and third person forms but not necessarily vice versa; compare, for example, *geben* – *gib!* – *gibst* 'to give – give! – give-2nd-sg.,' but *werden* – *werde!* – *wirst* 'to become – become! – become-2nd sg.' The subnode $[\ldots a\ldots]_{+PRET}$ is for preterit stems (e.g., *starb-*), from which subjunctives (e.g., *stürb-*) are inherited, and finally the stem $[..o..n]_{+PART}$ for (irregular) participle forms (e.g., *(ge)storben*).

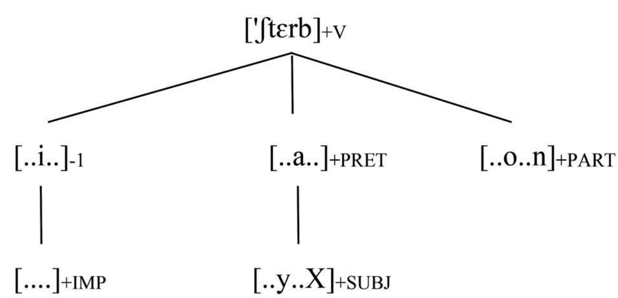

Figure 1. The stem *sterb-* and its subentries in a default inheritance network.

Previous experimental studies examined the processing of these kinds of lexically conditioned stem forms in L1 German speakers (Clahsen et al. 2001) and in groups of advanced nonnative L2 learners (Krause et al. 2015; Bosch & Clahsen 2015). All of these studies made use of the cross-modal priming paradigm in which prime-target pairs with marked and unmarked stem forms were compared and the direction of priming was manipulated. In one condition, priming went from marked to unmarked stems, e.g., *starb-* → *sterb-*, and in the other condition in the reverse order, from unmarked to marked stems, e.g., *sterb-* → *starb-*, (Clahsen et al. 2001). To determine priming effects for these conditions, they were compared to corresponding repetition-priming control conditions (i.e., with identical primes and targets), for example, *starb-* → *starb-* and *sterb-* → *sterb-*, which provide a measure of maximal facilitation as the prime activates the target word in its entirety. In L1 German speakers, marked stems (e.g., *starb-*) led to efficient (near-repetition) priming on the recognition of target forms of the same verbs that contained unmarked stems (e.g., *sterb-*). When the priming direction was reversed (e.g., *sterb-* → *starb-*), however, verb forms with unmarked stems as primes produced significantly less facilitation on target forms with marked stems. Clahsen et al. (2001) attributed this contrast to the different pairing of morphosyntactic features in the two conditions. In the "marked-to-unmarked" condition (*starb-* → *sterb-*), the target does not contain any information that is not already available from the prime, yielding efficient recognition of the target. In contrast, the targets in the "unmarked-to-marked" condition contain information that is not available from the prime, namely, the feature [+past], and this yielded significantly reduced repetition priming. Clahsen et al. (2001) interpreted these priming asymmetries in terms of hierarchically structured lexical entries (see Figure 1; Wunderlich 1996; Hippisley 2016). These complex entries consist of (underspecified) subentries for the different stem allomorphs that are defined in terms of phonological forms and morphosyntactic features and in which subentries inherit information from higher nodes but not vice versa. Given these representations, the priming asymmetry that Clahsen et al. (2001) reported results from the fact that a marked stem (e.g., *starb*) constitutes a specific subentry that inherits all the information from the higher node, which makes a marked stem a highly efficient prime for the corresponding unmarked target stem. In the reverse case ("unmarked-to-marked"), however, there is no such inheritance, with the specific, unprimed [+past] feature of the target (*starb*) reducing repetition priming.

More recently, Krause et al. (2015) extended this line of research, (i) by examining a different set of German verbal stem allomorphs, namely *stirb-* → *sterb-* in the "marked-to-unmarked" and *sterb-* → *stirb-* in the "unmarked-to-marked" condition; and (ii) by testing a group of fluent late bilinguals (mean AoA: 11.73 years; range: 6–21) with Russian as L1 in addition to a control group of L1 German speakers. While the L1 data replicated the priming pattern from Clahsen et al. (2001), the opposite pattern was found for the L2 group: less efficient priming from verb forms that contain marked stems (e.g., *stirb-*) than from prime words with the base stem (e.g., *sterb-*). Krause et al. (2015) argued that the L2 mental lexicon represents stem allomorphs as associatively linked entities on the basis of overlapping phonological form and meaning (e.g., Smolka et al. 2007; Lukatela et al. 1987). Furthermore, accessing marked forms such as *stirb-* is claimed to be more challenging in the L2 than retrieving unmarked forms, hence the reduced priming effect obtained for L2 speakers in the "marked-to-unmarked" condition. Krause et al. (2015) posited a sharp L1 versus L2 contrast in morpholexical representation as the source of the different priming effects. While the L1 lexicon employs structured lexical entries with morphosyntactic features, the L2 lexicon, for the same kinds of items, is supposed to be associatively organized using (surface) form and meaning. Note, however, that in terms of age of acquisition (AoA), Krause et al. (2015) compared two groups of extremes, the L1 speakers who acquired German from birth (AoA: 0) and L2 learners who learned German relatively late in life (mean AoA: of 11;07 years). This raises the question of how earlier onsets of acquisition modulate morpholexical representation and the kinds of priming patterns reported. Perhaps the division between L1 and L2 morpholexical representation and processing becomes less sharp and more gradient than previously envisaged once a spectrum of bilinguals is included covering a wider range of AoAs.

1.3. *The present study*

The purpose of the present study is to determine how different ages of onset of acquisition affect bilingual language performance, focusing on the long-term consequences of AoA on a bilingual speaker's morpholexical representation and processing. We investigated grammatical inflection, viz. German verb forms of German that encode morphosyntactic features via stem allomorphy, a phenomenon that is known to be challenging, at least for late bilinguals (e.g., Krause et al. 2015). We tested adult bilingual speakers who acquired Russian from birth and German at different ages. The experimental technique we employed was cross-modal lexical priming, which is supposed to tap into the mental representations of the "central-level" lexicon (Marslen-Wilson 2007). Previous research suggests that grammatical inflection is challenging, at least for late learners (Johnson & Newport 1989; Blom et al. 2006; Prévost & White 2000). Furthermore, inflectional processing has been argued to deteriorate with later AoAs (Veríssimo et al. 2018). On the basis of these findings, we may find efficient morphosyntactic feature priming in bilinguals, but only for those with an early AoA.

2. Methods

The following predictors of interest were collected: (i) *AoA of German*, that is, the self-reported age of onset of the acquisition of German (in years); (ii) *Skill in German*, a general measure of competence in the German language, as assessed by a paper-and-pencil version of the Goethe Institute Placement Test (a 30-item multiple-choice test assessing lexicon and grammar that can be mapped to the levels of the CEFR "Common European Framework for Languages"); and (iii) *Use of German*, a measure of everyday use and exposure to German, calculated as an average of self-reported percentages of use in a typical week (see Birdsong et al. 2012; Marian et al. 2007).

2.1. *Participants*

The data for the present study come from 106 Russian-German bilinguals (88 females) between the ages of 17 and 44 (mean: 25.29 years), all of whom had acquired Russian from birth and were recruited from a large Russian-German community living in the Berlin/Potsdam area in Germany. This sample of 106 participants included 26 late learners of German (tested by Krause et al. 2015), plus a new group of 80 participants. Table 1 presents biographic background information for all participants included in the present study.

One participant achieved only 10 out of 30 points in the Goethe Institute Placement Test, our measure of Skill in German (equivalent to an A2 skill level) and was excluded prior to any data analyses. The remaining participants were, as a whole, high-proficiency, advanced learners of German. In our measure of Skill in German, they had an average score of 26.44/30, with 58 participants scoring between 27 and 30 points, corresponding to a C2 level ("mastery/proficiency," the highest CEFR level), 42 participants scoring 22 to 26 points, corresponding to a C1 level ("effective operational proficiency/advanced"), and 5 participants achieving 20 to 22 points, which placed them at the upper end of the B2 level ("vantage/upper intermediate").

Table 1. Biographic information of participants (means; standard deviations, and range values in parentheses).

M age (in years)	Goethe placement test score (out of 30)	*M* age of acquisition (in years)	*M* length of exposure (in years)	*M* use of German (in percent)
25.29 (3.20;17–44)	26.44 (2.33; 20–30)	8.86 (5.94; 0–22)	16.5 (6.59; 3–36)	65.08 (15.27; 30–94)

The 105 participants whose data we analyzed reported an AoA of German between 0 years (i.e., from birth) and 22 years (mean: 8.86). Their average Use of German as self-reported for a typical week was 65.08%. All participants were right-handed and had normal or corrected-to normal vision. None of them reported having any kind of psychological or neurological disorder and/or language-related difficulty. They remained naïve with respect to the goals of the experiment until they finished the experimental session and received a small fee for their participation.

2.2. *Materials*

The materials were taken from Experiment 2 of Krause et al. (2015). The 32 critical items were the second- and third-person singular present-tense forms of so-called strong verbs of German, which have marked secondary present-tense stems. Two variants of these stems were included: (i) 18 verbs with a stem-vowel raising, from *-e* in the infinitive to *-i* in the second- and third-person singular present tense (e.g., *sterben* 'to die' vs. *du stirbst/er stirbt* 'you die/he dies'); and (ii) 14 verbs with stem-vowel fronting, from an *-a-* stem in the infinitive to an umlauted *-ä-* stem (e.g., *schlafen* 'to sleep' vs. *du schläfst/er schläft* 'you sleep/he sleeps'). Experimental primes and targets were presented with unmarked stems (viz., on the infinitive form) and with marked stems (viz., on the third-person singular present-tense form), yielding four types of prime-target combinations (see Table 2 for illustration).

Following previous studies (Clahsen et al. 2001; Leminen & Clahsen 2014; Bosch & Clahsen 2015), each of the two test conditions with morphologically related prime-target pairs (e.g., *sterben* → *stirbt* and *stirbt* → *sterben*) was compared against a corresponding control condition with identical prime-target pairs (e.g., *stirbt* → *stirbt* and *sterben* → *sterben*). For each target type (i.e., targets with marked and unmarked stems), we compared lexical decision RTs following test primes with those following identity primes; for example, RTs for the target *sterben* in *stirbt* → *sterben* were compared with RTs for the same target in *sterben* → *sterben*. We will refer to this measure as relative-to-identity (RTI) priming, defined as $RT_{identity} - RT_{test}$. Identical repetition of a word constitutes a baseline of maximal priming, since the prime activates the target in its entirety, including all its morphosyntactic information. We expect RTs following morphologically related (but not identical) prime words to be longer than those following identical primes, yielding negative RTI magnitudes ($RT_{identity}$ - RT_{test}). Hence, highly efficient (near-repetition) priming corresponds to RTI magnitudes close to 0. We expect that RTI priming depends on the properties of the target word that are and those that are not available from the prime word (cf. Clahsen et al. 2001). The more properties or features of target words are already available from morphologically related primes, the smaller in magnitude (i.e., less negative) should the RTI effect be; more specifically, RTs following test primes should be closer to identity priming in cases of relatively large prime-target overlap. On the other hand, if target words contain properties or features that are not available from test primes, these unprimed properties or features should lead to larger (i.e., more negative) RTI effects (i.e., to a greater difference between RTs folllowing test primes and following identity primes). In this way, RTI priming provides a measure of the extent to which our participants accessed the morphosyntactic features that are encoded in inflected forms.

Since morphologically related primes (e.g., *stirbt*) and targets (e.g., *sterben*) differed only in that they were different inflected forms of the same verbs, their semantic, phonological, and orthographic overlap, as well as their lemma frequencies, were perfectly parallel. It is true that the *-en* affix (e.g., in

Table 2. An example stimulus set (*sterben* 'to die').

		Target form	
		Marked stems	Unmarked stems
Prime type	Test	*sterben* → *stirbt*	*stirbt* → *sterben*
	Identity	*stirbt* → *stirbt*	*sterben* → *sterben*

sterb-en) is ambiguous in that it may encode first-/third-person plural or infinitives. However, if presented in isolation, these -*en* forms are most likely to be identified as an infinitival form. Because the infinitival form is not specified for any of the person/number features that finite forms such as *stirbt* are specified for, feature conflicts between these forms and infinitives should not arise.

Critical prime and target forms were matched for mean word-form frequency, such that third-person singular present-tense forms with marked stems (e.g., *stirbt*) exhibited a similar mean word-form frequency (57.6 per million, in the dlex database, Heister et al. 2011) as infinitival forms with unmarked stems (e.g., *sterben*, 59.6 per million) (repeated measures Cohen's $d = 0.13$). The items had a wide range of base-stem frequencies (2–3,436 occurrences per million), with a mean of 603 -per million ($SD = 884.41$); see Krause et al. (2015) for further details on the materials and a complete list of experimental items. All critical items were distributed in a pseudo-randomized manner over four experimental lists, such that each participant saw each verb only once. A set of 256 filler items (112 word-word fillers and 144 word-nonword pairs) was added to the experimental items.

2.3. *Procedure*

The procedures were parallel to Experiment 2 of Krause et al. (2015). All participants were tested individually in a quiet and dimly lit room. For the actual experiment, they were seated in front of a 24-inch computer screen and provided with stereo headphones. For both stimulus presentation and data collection, we employed the DMDX reaction time software (Forster & Forster 2003). Each participant was randomly assigned to one experimental version. In each trial, a fixation cross (800 ms) was followed by an auditory attention tone (200 ms) that aimed at focusing participants' visual and auditory attention. At the offset of the attention tone, the prime word was presented via headphones. The auditory stimuli were spoken by a female native speaker of Standard High German prerecorded in a sound studio at the University of Potsdam. Immediately at the offset of the spoken prime, the visual targets were presented on the computer screen and remained there for 500 ms. Participants were asked to perform a lexical decision as quickly and accurately as possible. After the targets' disappearance, participants were given an extra 2000 ms with a blank screen to respond. RT measurement started from the presentation of the targets onwards. After the response, the next trial started automatically with the presentation of the fixation cross. The experimental session was preceded by a practice phase including 20 trial items with 10 word and 10 nonword targets. Before the experiment, participants filled out a short biographical questionnaire. After the experiment, participants completed the Goethe Institute Placement test. The whole experimental session lasted approximately 35 minutes per participant.

2.4. *Data analysis*

No participant or item had to be removed due to low accuracy scores (range of by-participant accuracy: 78.1%–100%; range of by-item accuracy: 88.6%–100%). Incorrect responses (3.39%) and timeouts (0.12%) were removed from the data set. To normalize the distribution of RTs and reduce the influence of extremely long responses, data points that were slower than 2,000 ms were removed (4 data points; 0.12%); in addition, all analyses were conducted on log-transformed RTs (Baayen & Milin, 2010; Ratcliff 1993).

Following Verissimo et al. (2018), two sets of regression analyses were performed on the RT data: (i) linear mixed-effects models to determine how AoA is associated with RTI priming effects, and (ii) nonlinear models (viz., regression-with-breakpoints) to determine potential discontinuities in the function relating AoA to RTI priming. These analytical techniques are detailed in the following.

2.4.1. *Mixed-effects models*

The RT data were analyzed with linear mixed-effects regression models with crossed random effects for participants and items (e.g., Baayen et al. 2008). As categorical fixed effect variables, the models included

Prime Type (identity vs. test), Target Type ("marked-to-unmarked" and "unmarked-to-marked"), and the interaction between these two factors. The models further included the continuous predictors AoA of German as well as Skill in German and Use of German. Each of these three continuous predictors was allowed to interact with the factorial fixed effects and, in particular, with the critical Prime Type x Target Type interaction, yielding three 3-way interactions (as well as all subsumed 2-way interactions). Thus, the AoA effects reported in the following are statistically controlled for Skill in German and Use of German: They represent "purer" effects of AoA once any correlations with language skill or frequency of use have been removed (see, e.g., Wurm & Fisicaro 2014). Finally, Trial Position (i.e., the rank of items in the task) was also included as a continuous predictor to control for task-related effects and to remove autocorrelation of residuals (Baayen & Milin, 2010). Treatment contrasts (i.e., dummy coding) were employed for the factorial predictors and the comparisons of interest were obtained by relevelling factors and refitting the model. All continuous predictors (AoA of German, Skill in German, Use of German, Trial Position) were centered around their means.

Regarding the models' random effects structure, we followed the recommendation of Matuschek et al. (2017) and included "random slopes" (which capture variation in the magnitudes of effects across participants and items) only if they resulted in models with a lower AIC, indicating greater goodness of fit. Each of the key predictors, Prime Type, Target Type, and AoA of German was tested for inclusion as a by-subject and/or by-item slope (as appropriate) and the AIC of the resulting model was recorded. The slope that provided the largest drop in AIC was included first and all other slopes were retested for inclusion, with this process being repeated for as long as a better model could be achieved. Using the intercept-only model as the basis (AIC = −1.48), the following random slopes were consecutively added: Target Type by item (AIC = −109.48), Prime Type by item (AIC = −113.70), the Target Type by Prime Type interaction (AIC = −142.81), and Prime Type by subject (AIC = −142.94). All the following results were obtained from models with this final random structure.

2.4.2. *Regression-with-breakpoints*

To test for potential discontinuities in the effects of AoA on RTI priming, we fitted a number of *regression-with-breakpoints* models. Regression-with breakpoints (also called "segmented" or "piecewise" regression) allows combining two linear regressions in a single model, by estimating two different slopes at each side of a breakpoint (see, e.g., Baayen 2008; Neter et al. 1996). In particular, we estimated (i) AoA effects that are present in an *initial* age band, from an AoA of 0 up to a breakpoint in the AoA scale, and (ii) AoA effects that take place *after* the breakpoint, until an AoA of 22 (i.e., the maximum AoA represented in our sample). If the two regression slopes are significantly different, then there is evidence for a break or discontinuity in the function being assessed (in our case, in the function relating AoA and RTI priming, for the "marked-to-unmarked" and "unmarked-to-marked" conditions). In other words, if the data presents a discontinuity that is strong enough so that the AoA effects on each side of a breakpoint are of different magnitudes, then the greater complexity of estimating these two slopes is justified—in which case the regression-with-breakpoints is to be preferred as a more adequate model of the data.

One question that arises when employing this statistical technique is how to estimate the precise location of the breakpoint. Here we follow a *breakpoint discovery* procedure, similarly to the approach outlined in Veríssimo et al. (2018) (in turn following Baayen 2008; Vanhove 2013). In this approach, a series of models containing breakpoints at each value in the AoA scale is separately fit to the data and the model that best fits the data (specifically, the model with the lower deviance) is picked out as containing the most likely location for a discontinuity.

In what concerns other model characteristics, note that the regression-with-breakpoints models that we have employed included only minimal modifications to the previous linear mixed-effects model, namely, the terms that allow estimating AoA effects to the "left" and "right" of the break-point, as well as their difference. However, other model characteristics were kept the same as in the linear models: (i) we have maintained the two additional predictors Skill in German and Use of

German (as well as their interactions with Prime Type, Target Type and the Prime Type x Target Type interaction), and (ii) the breakpoint models had the same random effects structure that was described previously for the linear models.

3. Results

Mean RTs, SDs and accuracy rates for lexical (word/nonword) decisions in the four experimental conditions (across all participants and items) are shown in Table 3. As accuracy rates were close to ceiling in all conditions, we did not perform any further analyses on these data. The mean RTs indicate faster responses for identity than for test primes and faster responses for target verb forms with unmarked stems than for those with marked stems.

The mixed-effects regression model revealed that in both the "marked-to-unmarked" condition (e.g., *stirbt* → *sterben* vs. *sterben* → *sterben*) and the "unmarked-to-marked" condition (e.g., *sterben* → *stirbt* vs. *stirbt* → *stirbt*), RTs following test primes were significantly longer than following identity primes (unmarked targets: $b = 0.1079$, $t = 6.38$, $p < .001$; marked targets: $b = 0.0944$, $t = 4.70$, $p < .001$). More importantly, a significant three-way interaction of Prime Type, Target Type, and AoA of German was obtained ($b = -0.0064$, $t = -2.29$, $p = .022$), which indicates that differences in RTI priming in the "marked-to-unmarked" versus "unmarked-to-marked" conditions were modulated by AoA of German. Figure 2

Table 3. Overall mean RTs (in ms; standard deviations in parentheses).

	Visual targets	
Auditory primes	Marked stem	Unmarked stem
Test	624 (129)	604 (139)
Identity	565 (143)	538 (109)

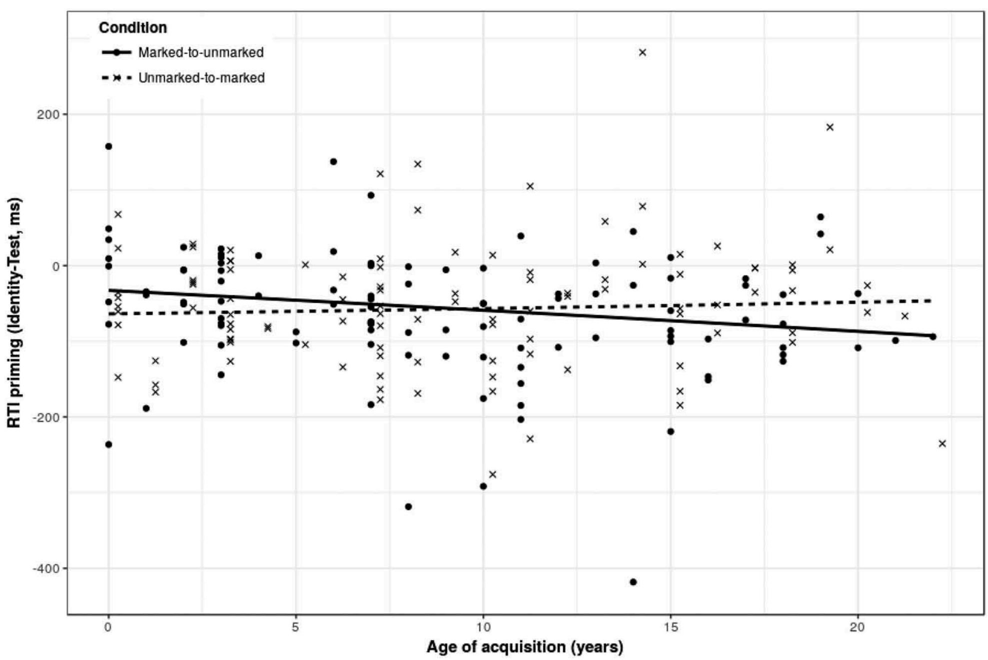

Figure 2. Linear regressions for the "marked-to-unmarked" condition (solid line) and the "unmarked-to-marked" condition (dashed line). Black dots (for the "marked-to-unmarked" condition) and gray crosses (for the "unmarked-to-marked" condition) display RTI priming (in ms) for individual participants with respective AoAs. Note that the AoA scale for the "unmarked-to-marked" condition is slightly shifted to the right for illustration purposes.

displays this three-way interaction by plotting the linear AoA effects on the magnitudes of RTI priming, separately for the "marked-to-unmarked" and the "unmarked-to-marked" conditions. In the "marked-to-unmarked" condition (e.g., *stirbt* → *sterben* vs. *sterben* → *sterben*), there was a significant interaction of Prime Type by AoA of German ($b = 0.0044$, $t = 2.13$, $p = .034$), with increasing AoA being associated with a larger (more negative) RTI effect (see Figure 2, full line). Recall that a more negative RTI effect indicates a greater difference between (longer) RTs following test primes (e.g., *stirbt* → *sterben*), relatively to (shorter) RTs following identity primes (e.g., *sterben* → *sterben*). In contrast, for the "unmarked-to-marked" condition (e.g., *sterben* → *stirbt* vs. *stirbt* → *stirbt*), no interaction between Prime Type and AoA of German was obtained ($b = -0.0020$, $t = -0.96$, $p = .338$), that is, the RTI difference between test and identity conditions did not change with AoA (see Figure 2, dashed line).

These results indicate that AoA is associated with a larger difference between morphologically related and identity priming for "marked-to-unmarked" prime-target pairs. However, given that AoA may be correlated with other speaker attributes, the question arises of whether the modulation of priming effects can be attributed to AoA or instead has other sources; see Table 4 for correlations between participant-level variables.

It is true that participants' AoA exhibits a weak negative correlation with the variables Skill in German ($r = -.23$) and Use of German ($r = -.38$), indicating that with increasing AoA, the speakers' level of skill in German as well as the amount of use of the German language in their everyday lives slightly decrease. There are, however, two reasons for why AoA is a more likely source of the obtained effect than other participant-level variables related to proficiency or use. Firstly, recall that our statistical model also included the variables Skill in German and Use of German (both interacting with Prime Type and Target Type), which means that the reported AoA effects have been controlled for these covariates. Thus, the effect that we obtained is an effect of the "unique" part of AoA, that is, of the portion of its variance that cannot be attributed to Skill in German or Use of German (Wurm & Fisicaro 2014). Secondly, unlike AoA, the two additional predictors did not modulate RTI priming neither for the "marked-to-unmarked" condition (Skill in German: $b = 0.0013$, $t = 0.24$, $p = .812$; Use of German: $b = -0.0004$, $t = -0.41$, $p = .680$), nor for "unmarked-to-marked" prime-target pairs (Skill in German: $b = -0.0008$, $t = -0.14$, $p = .886$; Use of German: $b = 0.0001$, $t = 0.13$, $p = .896$). Accordingly, no three-way interactions were obtained (Skill in German: $b = -0.0021$, $t = -0.28$, $p = .776$; Use of German: $b = 0.0005$, $t = 0.40$, $p = .688$). Thus, of the three predictors we examined, only AoA produced a reliable linear effect on the different RTI priming patterns.

3.1. *Analysis of AoA vs. exposure*

Although these results allow us to conclude that between-participant differences in language skill and use cannot explain the obtained AoA effects on RTI priming, AoA is also closely related to a different participant-level variable: the number of years that participants have been exposed to the German language. Naturally, those participants who acquired German earlier in life have also been speaking it and hearing it for a longer time, which raises the possibility that effects that appear to result from AoA may in fact be effects of the length of exposure to the second language. To assess AoA effects independently of length of exposure (and vice versa), we first fitted a mixed-effects model that included not only the variables AoA, Skill in German, and Use of German, but also Length of Exposure to German (defined as chronological age minus AoA). In this model, neither

Table 4. Correlation coefficients (r) between participant-level variables.

	Age of acquisition	Skill in German	Use of German
Skill in German	−0.23		
Use of German	−0.38	0.41	
Length of exposure	−0.74	0.21	0.28

AoA nor Length of Exposure had an effect on "marked-to-unmarked" RTI priming (AoA: $t = 1.16$; Exposure: $t = -0.45$), and neither produced a significant three-way interaction (AoA: $t = -0.79$; Exposure: $t = 1.14$). The lack of significant effects when both AoA and Length of Exposure were included in the same regression is likely to result from the strong negative correlation between these two variables ($r = -.74$), such that their unique (uncorrelated) parts are small. That is, in the presence of such high levels of multicollinearity, the likelihood of detecting either effect becomes much lower (Friedman & Wall 2005; Wurm & Fisicaro 2014).

We have made use of two different statistical approaches that enabled us to avoid the problem of multicollinearity and ascertain which of the two variables (AoA or Length of Exposure) is a better predictor of "marked-to-unmarked" RTI effects. Firstly, we employed backwards stepwise regression, a technique in which predictors are sequentially removed from a regression model, whenever the removal leads to a model with greater "goodness of fit" (this is commonly assessed by AIC, a measure that penalizes complexity and leads to predictors being kept only when they make a substantial contribution to explaining variance in the data; Venables & Ripley 2002). We calculated AIC values for the previous "full" model (containing AoA, Length of Exposure, Use of German, and Skill in German) and for models in which each of these predictors (as well as their interactions with Prime Type and Target Type) were removed. The model with the lower AIC (i.e., better fit) was then selected, and this procedure was sequentially repeated until removal of predictors did not produce a better model. Beginning with the "full" model (AIC = −136.8), Use of German was removed in the first step (AIC = −142.9), followed by Length of Exposure (AIC = −150.5), and finally, Skill in German, leaving a model that, of the initial four biographical variables, contained only AoA in interaction with RTI priming effects (AIC = −153.9). However, removal of AoA (and its interactions) would produce a *worse* model (AIC = −140.5). That is, when the four biographical variables were pitted against one another and sequentially removed, then AoA emerged as the only important predictor.

A second approach to avoiding multicollinearity and disentangling the contributions of AoA and length of exposure is to directly compare the two "rival" candidate models on the same data set: one model with the predictors AoA, Skill in German, and Use of German (as well as their interactions with Prime Type and Target Type) and another model with the predictors Length of Exposure, Skill in German, and Use of German (with the same interactions). That is, instead of including both AoA and Length of Exposure in a single regression model (each controlled for the other), we instead attributed the common variance of AoA and Length of Exposure to one or the other predictor (i.e., to AoA, uncontrolled for Exposure, and to Exposure, uncontrolled for AoA) and compared the two resulting models in terms of how well they fit the data. The first indication that AoA, rather than Length of Exposure, is the more important predictor of RTI priming effects was that the critical effect of AoA reported previously (with increasing AoA being associated with a larger RTI difference for "marked-to-unmarked" items) did not reach significance in the model in which AoA was replaced by Length of Exposure ($b = 0.0026$, $t = 1.45$, $p = .148$). Secondly, and more importantly, it was also possible to directly compare these two models by calculating their goodness of fit. The regression model with AoA (plus covariates and interactions) had an AIC of −142.9, whereas the rival model in which AoA was replaced with Length of Exposure had an AIC of −135.3 (as mentioned previously, smaller values indicate better fit). Finally, to intepret the magnitude of this difference, we made use of a related measure of fit, the Bayesian Information Criterion (BIC; Schwarz 1978). In turn, this measure allowed us to quantifiy the relative evidence in favor of one or the other model, expressed as a Bayes Factor (see Wagenmakers 2007). Specifically, the difference in fit between the two models (delta BIC = 7.6) corresponded to a Bayes Factor of 45.7, which means that a model with AoA is ~46 times more likely given the obtained data than a model with Length of Exposure (98% vs. 2% posterior probabilities). The data thus constitute "strong evidence" in favor of the model with AoA (Raftery 1995).[1]

[1]This, of course, presupposes that the two models are equally plausible a priori, a standard assumption in such analyses.

To conclude, the results of a stepwise regression procedure, as well as of a direct comparison of models using measures of goodness of fit, allow us to conclude that differences between participants on "marked-to-unmarked" priming are better explained by AoA (i.e., by the age at the onset of German acquisition) than by length of exposure to German.

3.2. *Nonlinear breakpoint analyses*

Our second main analysis sought to determine whether nonlinear discontinuities can be found in the modulation of RTI priming by AoA for "marked-to-unmarked" prime-target pairs. Effects of AoA may, for example, be present until a certain age but decrease in magnitude or disappear afterwards (e.g., Johnson & Newport 1989; DeKeyser 2005), or alternatively, our processing measures may be unaffected by AoA during an early "window" but show a modulation by AoA only after a certain age (e.g., Veríssimo et al. 2018). These nonlinear analysis made use of regression-with-breakpoints, also conducted as mixed-effects models on all trials (i.e., without prior averaging).

As explained previously, we followed a breakpoint discovery procedure to determine the most likely location for a discontinuity in the AoA–RTI priming function. In particular, we have fitted models containing breakpoints at each year in the AoA scale (between the AoAs of 1 and 21 years) and recorded each model's goodness of fit, as quantified by its deviance. The results of this procedure are shown in Figure 3, which shows the deviance of models with breakpoints at different points of the AoA scale. The results show that the best model (i.e., with the lowest deviance) contained a breakpoint at an AoA of 11 years (followed closely by a model with a breakpoint at an AoA of 10).

The results of the best model (with a breakpoint at an AoA of 11) are displayed in Figure 4, separately for the "marked-to-unmarked" and "unmarked-to-marked" conditions. AoA effects on RTI priming were found to be significantly different between the two priming directions, but only before an AoA of 11 as revealed by a three-way interaction between AoA, Prime Type, and Target

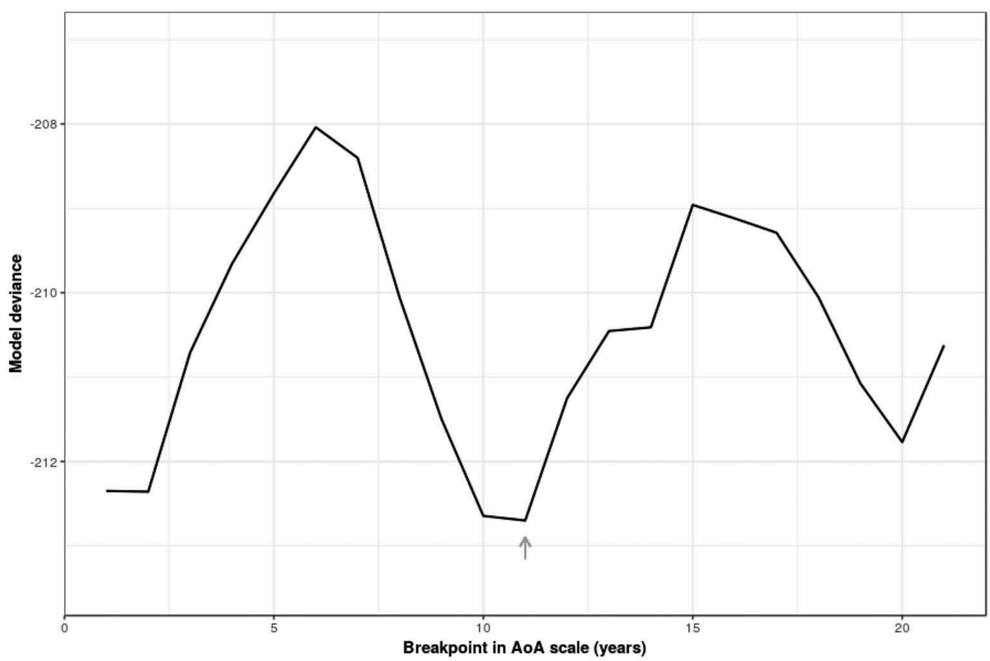

Figure 3. Model deviance (i.e., goodness of fit) for regression-with-breakpoint models, with successive breakpoints at different values of the AoA scale.

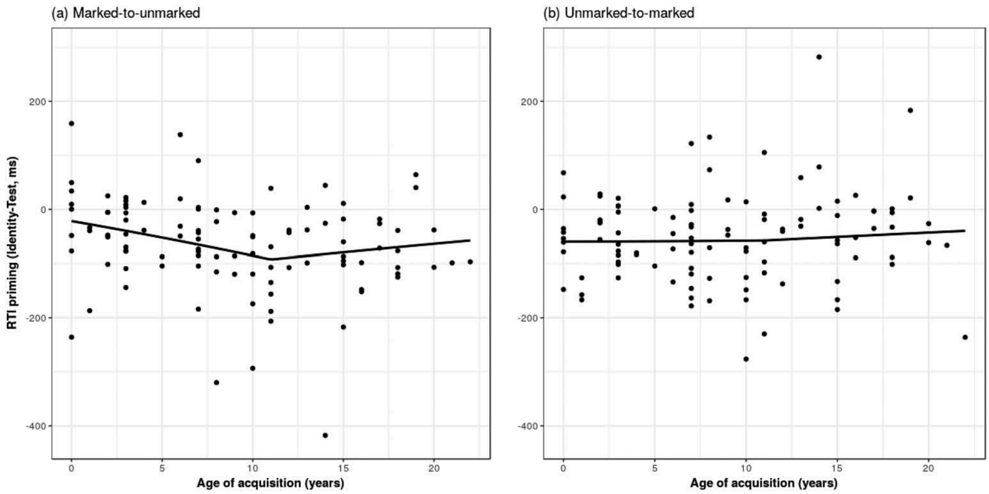

Figure 4. Regression-with-breakpoints model for the effect of AoA on the "marked-to-unmarked" condition (panel a) and the "unmarked-to-marked" condition (panel b). Black dots display RTI priming (in ms) for individual participants with respective AoAs.

Type, which was only present before an AoA of 11 (i.e., AoAs of 0 to 11; $b = -0.0122$, $t = -2.50$, $p = .012$) but not after (i.e., AoAs of 11 to 22; $b = 0.0023$, $t = 0.35$, $p = .728$). In the "marked-to-unmarked" condition (Figure 4, panel a), a significant difference was obtained between the AoA slopes on the left and the right sides of an AoA of 11 years, indicating a *breakpoint discontinuity* (i.e., a significant three-way interaction was obtained between the adjusted AoA variable, the indicator variable, and Prime Type; $b = -0.0150$, $t = -2.02$, $p = .044$). Up until an AoA of 11, the results closely resemble the pattern that was obtained in the linear model presented previously. Specifically, increasing AoA was associated with an increasingly greater difference between RTs following test (e.g., *stirbt* → *sterben*) and identity primes (e.g., *sterben* → *sterben*), that is, with a more negative RTI effect for "marked-to-unmarked" prime-target pairs ($b = 0.0104$, $t = 2.87$, $p = .004$) . In contrast, after an AoA of 11 years, the effect of AoA on RTI priming was no longer present (and even numerically reversed, $b = -0.0046$, $t = -0.94$, $p = .348$). In other words, in the "marked-to-unmarked" condition, the RTI effect was constant throughout the AoA range of 11 to 22 years.

In contrast, the RTI priming effect for the "unmarked-to-marked" condition (e.g., *sterben* → *stirbt*, relatively to *stirbt* → *stirbt*) revealed a very different AoA function (see Figure 4, panel b) without any evidence of a breakpoint discontinuity ($b = -0.0005$, $t = -0.07$, $p = .948$) and no effects of AoA on RTI priming, neither before ($b = -0.0018$, $t = -0.50$, $p = .616$) nor after an AoA of 11 years ($b = -0.0023$, $t = -0.46$, $p = .642$). Thus, the regression line in Figure 4, panel b, is continuous and approximately flat: It presents no obvious discontinuities and shows a similar RTI priming effect throughout the AoA scale.

As in the linear analyses reported previously, the additional predictors Skill in German and Use of German showed no effects on RTI priming, neither for the "marked-to-unmarked" nor for the "unmarked-to-marked" condition and no interactions (all $|t|$s < 0.43, all ps$>.671$).

4. Discussion

The main findings of the current study can be summarized in two points. Firstly, AoA was found to selectively affect morphological priming such that facilitation from marked stems as primes to unmarked stems as targets (e.g., *stirbt* → *sterben*) gradually declined with increasing AoA, while for the reverse condition ("unmarked-to-marked," e.g., *sterben* → *stirbt*), priming effects were found to be constant, irrespective of the age at which German was acquired.

Secondly, this AoA effect on "marked-to-unmarked" priming was discontinuous (rather than linear), with a gradual decline from an AoA of 0 up to an AoA of 11 and a flattening of AoA effects thereafter (parallel to the "unmarked-to-marked" condition). Hence, AoA effects on priming are confined to a particular age band. In the following, these findings will be discussed in greater detail.

4.1. *Gradient activation of lexical entries*

The effect of age of onset of language acquisition we obtained in the "marked-to-unmarked" condition is indicative of a modulation of morphosyntactic feature access across the AoA scale. Recall that previous L1 research on the processing and representation of verbal stem allomorphs revealed near-repetition priming effects on the recognition of unmarked targets when preceded by marked primes (e.g., *starb-* → *sterb-*) compared to when the order of prime and target was reversed (e.g., *sterb-* → *starb*) (Clahsen et al. 2001). For late bilinguals, however, the opposite pattern was found, with less efficient repetition priming from marked stem primes than from prime words, including the base stem (Krause et al. 2015). These findings were attributed to the marked stems' specific morphosyntactic feature content, which yielded near-repetition priming for "marked-to-unmarked" items in the L1, but not for late bilinguals. Krause et al. (2015) attributed this contrast to different morpholexical representations of these forms in the L1 versus the L2 lexicon, structured lexical entries with morphosyntactic features such as those proposed by Wunderlich (1996) for the L1 (see Figure 1), and an associative lexicon that encodes (surface) form and meaning but not morphosyntax for the L2.

Our current results demonstrate that it was (RTI) priming from marked stems that was affected by AoA, with a significant decrease in facilitation between an AoA of 0 and 11 (see Figure 4, panel a), while (RTI) priming in the reverse "unmarked-to-marked" condition was not affected by AoA (see Figure 4, panel b). In other words, the extent to which a marked stem's morphosyntactic features facilitate (target) word recognition *gradually* changes across the AoA scale. This finding suggests that there might be intermediate states of morpholexical representation (depending on AoA) between a lexicon consisting of structured entries with morphosyntactic features (such as those illustrated in Figure 1) and an associative lexicon of direct form-meaning mappings without morphosyntactic features (as proposed by Krause et al. (2015) for a late-learned L2).

To be more precise, we propose to maintain the notions of structured lexical entries with underspecified subnodes held together by default inheritance, while at the same time introducing the possibility of gradient activation for the subnodes of a lexical entry; see Veríssimo (2019) for a computational model of lexical representation that implements these notions. Here, we present a brief sketch of this model with special reference to allomorphic stems of German verbs (see Figure 5). Note that (for simplicity) not all subnodes of the verb *sterben* are shown here.

As is common in interactive activation models, both the nodes (and subnodes) as well as the links between them have different weights or strengths (as indicated by circles and curved arrows). In addition, subnodes are supposed to be underspecified and linked by default inheritance, as in Wunderlich's (1996) original account. One consequence is that the recognition of the base stem *sterb-* directly activates the base node (i.e., STERB-; see Figure 5) and indirectly the subnode STIRB- (by virtue of the two stems belonging to the same lexical entry). By contrast, the recognition of the marked stem *stirb-* directly activates both the subnode containing the specific features of the marked stem (STIRB-) and the base node (i.e., STERB-; see Figure 5) that contains general properties of the lexical entry inherited by the subnode. In this way, the priming asymmetries reported for L1 German speakers can be explained (Clahsen et al. 2001; Krause et al. 2015), in that the prime in a prime-target pair such as *stirb-* → *sterb-* directly activates the target, whereas in the reverse case (*sterb-* → *stirb-*) the prime only indirectly activates the target. In addition, the various nodes and their links have weighted activation levels and connection strengths, changes of which may lead to differences in priming. Reduced priming from marked-to-unmarked stems, for example, the pattern found for

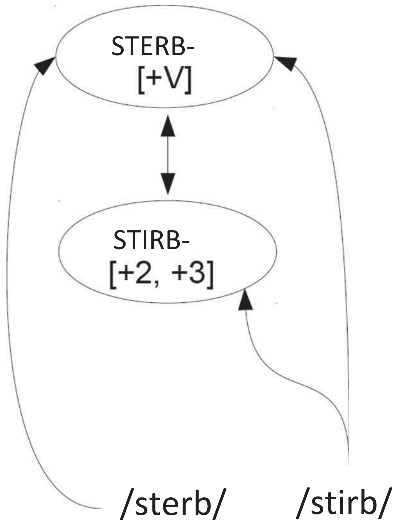

Figure 5. Model of a structured lexical entry with gradient activation for subnodes for the German strong verb *sterben* ("die").

bilinguals with late AoAs, may arise from weaker links between a marked stem's phonological form and its morphosyntactic features. Alternatively, the link from the marked to the base stem may be weakened, such that the activation propagating from a subnode like STIRB- to the base node STERB- is reduced. In any case, by introducing activation and connection weights into structured lexical entries, we account for graded AoA-related priming effects. The specific findings from the current study suggest that until the "breakpoint," i.e., within a sensitive period from 0 to 11 years, the earlier the onset of acquisition the stronger the activation level of the marked stem, its morphosyntactic features, and/or its connection weights, yielding efficient "marked-to-unmarked" priming for bilinguals with early AoAs. Later exposure to the second language during this period leads to less stable morpholexical representations and consequently, reduced priming effects.

4.2. *Sensitive periods for specific linguistic domains*

The present data suggest that the development from native-like to nonnative performance is discontinuous across the AoA scale. Up to an AoA of 11, "marked-to-unmarked" priming gradually declines, after which it levels off and remains flat throughout later AoAs (see Figure 4, panel a). While our finding of a gradual decline rather than a "catastrophic one-time event" (Long 1990:251) is consistent with previous research on sensitive periods for language acquisition, a number of studies have reported shorter durations of sensitive periods for language. Long (1990, 2013) argued, for example, that native-like performance for grammar (morphology and syntax) is most likely for an AoA range of 0 to about 6 years and decreasingly likely at later AoAs (from 6 to the midteens), see also Hyltenstam and Abrahamsson (2003) and Veríssimo et al. (2018) for similar findings. Meisel (2013), using evidence from early bilingual children's language development, proposed a sensitive period for morphsyntax with an even earlier offset of around 4 years of age. By contrast, several other studies reported later offsets of sensitive periods for grammar. Johnson and Newport (1989), for example, obtained a strong correlation between ultimate attainment in a second language and AoA before 16 years and a much weaker one thereafter (r = .87 vs. r = .16); see also DeKeyser (2005) and Granena and Long (2012). Most recently, Hartshorne et al. (2018) collected grammaticality judgments from an unusually large set of two-thirds of a million native and nonnative English speakers covering a wide range of AoAs. They argue that their results indicate a sharply defined critical period for language acquisition, but with a rather late offset. Native-like attainment is still within reach up until an AoA of approximately 17 years, according to Hartshorne et al. (2018).

How can these differences in offsets of sensitive periods be explained? One crucial factor is that sensitive periods may be linguistically *selective*, in that particular linguistic domains, skills, or phenomena may have sensitive periods with their own specific age bands. Consequently, the observed variability in ages of offset of sensitive periods might be due to the different phenomena tested in previous studies. A related source of variability in the reported length of sensitive periods may come from the different linguistic tasks and measures used in previous studies to determine AoA effects for language or grammar. While many studies relied on coarse-grained measures of linguistic proficiency or ultimate attainment, such as accuracy scores in sentence recall, picture matching, and acceptability and grammaticality judgments across a broad range of phenomena, other studies used more fine-grained measures, e.g., response latencies, to determine potential AoA effects on more specific aspects of linguistic knowledge and performance. AoA bands can be expected to differ depending on which tasks and measures are employed. Most probably, global linguistic proficiency or ultimate attainment tasks and measures yield broader sensitivity periods with wider AoA bands than studies using more subtle measures that tap into more specialized linguistic skills. A case in point of the former is Hartshorne et al.'s (2018) study for which an average general performance measure (viz. accuracy log-odds) was calculated from grammaticality judgments of a wide range of syntactic and morphosyntactic phenomena (e.g., passives, clefts, agreement, relative clauses, verb syntactic subcategorization, *wh*-movement). For this measure, Hartshorne et al. discovered a critical period of broad grammatical sensitivity that is preserved until about 17 years and then declines. One study illustrating that more subtle linguistic measures may yield more confined sensitive periods is Veríssimo et al.'s (2018) study comparing masked morphological priming effects for inflection and derivation, which revealed a highly selective AoA effect with a nonlinear trajectory (indicative of a sensitive period) for *inflection only*, and no AoA modulation for the same participants' performance on derivation.

That sensitive periods for language and grammar are linguistically selective has been shown in a number of previous studies. Granena and Long (2012), for example, argued for "windows of opportunity" with specific AoA bands for different linguistic domains, with an early offset for phonology, followed by the lexicon and collocations, and finally, with an offset in the midteens, for morphosyntax. Likewise, Huang (2014) reported distinct AoA effects with different age bands for grammaticality judgment versus spoken production tasks. Even more subtle sensitive periods were reported by Werker and Hensch (2015) for phonology, who found distinct AoAs bands, for example, for phoneme perception, audiovisual matching, and phonemic integration. Consider, for example, the well-known case of phoneme discrimination in young infants that is supposed to be under maturational control. Until about 10 to 12 months of age, infants are capable of discriminating consonant distinctions that are not testified in their input. Later, through listening to the language(s) in their environment, these particular languages' distinctions are strengthened, and the ability to discriminate phonemic distinctions that are not available from the environment is subject to a steady decline; see Werker and Hensch (2015) for a review. If sensitive periods are indeed specific to particular linguistic domains and phenomena, even within grammar, the diverse AoA bands reported in previous research do not come as a surprise, given that different tasks, measures, and phenomena were employed in these studies.

How can the results of the present study be reconciled with previous research on sensitive periods for language and grammar? Our results are unusual in that we found a sensitive period with a *late offset* of 11 years, even though a *highly specific* linguistic phenomenon (viz. stem allomorphy) was examined using a *fine-grained* measure (viz. cross-modal priming) that taps into subtle properties of morpholexical representation and processing. The study that is most closely comparable to the present one is Veríssimo et al. (2018) with Turkish/German bilinguals. As in our current study, Veríssimo et al. (2018) investigated a specific inflectional phenomenon of German (viz. – *t* past participle formation, relative to derivation) using a morphological priming technique (albeit masked rather than cross-modal priming). Nevertheless, Veríssimo et al. (2018) obtained a considerably narrower sensitive period for inflectional priming from with a much earlier offset (at about 5 to 6 years) than what we found in the current study. Note, however, that although both studies examined inflectional phenomena of German, there is an important difference between – *t* participle priming (e.g., *gewarnt*) and priming from the kinds of inflected verb forms with marked stems (e.g., *wirft*) that were tested in the current study. The latter represents a case of *lexically conditioned* inflection that applies to a specific subset of so-called "strong" German verbs that have marked stems with internal stem changes to encode particular morphosyntactic feature sets. By contrast, – *t* participle formation represents a completely regular—*lexically unconditioned*—affixation process that applies to any member of the category "verb" to form a past participle. Veríssimo et al. (2018) attributed the AoA effect they obtained for – *t* participle priming to paradigm-based learning mechanisms, which are subject to a sensitive period during which inflectional rules can be efficiently extracted from the input.

Matters for lexically conditioned inflectional forms are different, however, since these forms are stored in lexical memory, rather than being derived by inflectional rules. The lexical memory representations of the kinds of strong verbs and their marked stems that were investigated in the present study, for example, may be conceived of in terms of structured lexical entries with various linked subnodes and weighted activation and connection links; see Figure 5 for illustration. Although our results provide support for a sensitive period and more generally, for maturational control of morphosyntax (Granena & Long 2012), complex lexical entries of this kind may need time and exposure to get fully established. As a result, later ages of onset of acquisition within the sensitive period will yield reduced activation levels of the marked stem, its morphosyntactic features, and/or its connection weights. A long-term consequence of weaker morpholexical representations is the AoA-related gradual decline in morphological priming from inflected forms with marked stems that we found in the present study.

Supporting evidence that the development of stable morpholexical representations for irregularly inflected word forms takes time comes from the study of morphological overregularization errors in child language production. Marcus et al.'s (1992) large-scale study of the regular and irregular past-tense forms in English child language revealed that overregularizations (i.e., **strived* instead of *strove*) rarely occur (with less than 4% of all irregular forms) but are found during an extended period of time and well into school age. Even adults occasionally make overregularization errors in their spontaneous speech (Stemberger 1982). These kinds of errors are attributable to weak lexical entries (for specific irregular forms), which cause memory-retrieval failures and as a result a fall-back on the regular form (Marcus et al. 1992). Lengthy periods with occasional overregularization errors into late childhood have also been reported for the acquisition of other languages. For 8–13-year-old German children, for example, Jessen et al. (2017) reported – *t* overregularization rates between 4% and 6% in their (elicited) production of irregular participle. In line with these findings from child language research, we interpret the observed discontinuity at an AoA of 11 years in our present data as reflecting an extended period of sensitivity for the development and stabilization of complex lexical entries with their irregular subentries and corresponding morphosyntactic features.

5. Conclusion

The current study investigated long-term consequences of the onset of acquisition of a second language on grammatical processing in bilingual speakers. Our focus was on how AoA affects bilinguals' language performance with respect to how morphosyntactic features are accessed from inflected verb forms with marked (irregular) stems during online word recognition. The results showed a *linguistically selective* and *nonlinear* AoA effect in the observed priming patterns. AoA was specifically modulated for inflected forms with morphosyntactically marked stems. We also found a nonlinear discontinuity for this AoA effect, suggesting a sensitive period for the development of this type of inflectional morphology. The specific findings from the present study have implications for two closely related more general issues in language acquisition and bilingualism research: (i) the nature of native versus nonnative (L1/L2) differences in morphological processing and (ii) the role of sensitive critical periods in bilingual language development.

Firstly, as regards L1/L2 differences in morphological processing, many studies have compared native speakers with late bilinguals and reported sharp differences between these two groups. For the specific phenomenon under study here, for example, Krause et al. (2015) proposed distinct morpholexical representations for the L1 and the L2 lexicon, a set of structured entries with morphosyntactic features in the L1, and an associative lexicon of direct form-meaning mappings without morphosyntactic features for a late-learned L2. Secondly, the results of the current study confirm that sensitive/critical periods of development exist for both L1 and L2 language acquisition. We found a long-term effect of AoA on participants' linguistic performance, which provides new evidence for a sensitive period in bilingual language development. For future research on sensitive/critical periods, it is perhaps worth mentioning how this was achieved (see also Veríssimo 2018): (i) by targeting a specific linguistic domain (inflectional morphology), (ii) by using a fine-grained measure of (morpholexical) representation and processing, and (iii) by testing AoA effects for nonlinear discontinuous trajectories. We suggest that this approach is beneficial for determining sensitive/critical periods in language development, since it goes beyond the more familiar studies of global linguistic outcomes, general proficiency, or ultimate attainment that are perhaps not sensitive enough to discover linguistically highly selective sensitive periods of language development.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by an Alexander von Humboldt-Stiftung award to Harald Clahsen. In addition, this study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Collaborative Research Centre SFB 1287, Project Numbers A02 and B04.

ORCID

João Veríssimo http://orcid.org/0000-0002-1264-3017
Harald Clahsen http://orcid.org/0000-0002-6190-1318

References

Baayen, H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

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

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

Bialystok, E. & B. Miller. 1999. The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition* 2. 127–145. doi:10.1017/S1366728999000231

Bialystok, E. & K. Hakuta. 1999. Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In Birdsong David (ed.), *Second language acquisition and the Critical Period Hypothesis*, 161–181. Mahwah, NJ: Lawrence Erlbaum.

Birdsong, D. 2005. Interpreting age effects in second language acquisition. In Kroll J., De Groot A. (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 109–127. Oxford: Oxford University Press.

Birdsong, D. 2006. Age and second language acquisition and processing: A selective overview. *Language Learning* 56. 9–49. doi:10.1111/j.1467-9922.2006.00353.x

Birdsong, D. 2009. Age and the end state of second language acquisition. *The New Handbook of Second Language Acquisition* 17. 401–424.

Birdsong, D. 2014. Dominance and age in bilingualism. *Applied Linguistics* 35. 374–392. doi:10.1093/applin/amu031
Birdsong, D. & J. Vanhoeve. 2016. Age of second language acquisition: Critical periods and social concerns. In Nicoladis E., Montanari S. (eds.), *Bilingualism across the lifespan: Factors moderating language proficiency*, 163–182. Berlin: De Gruyter.
Birdsong, D., L. Gertken & M. Amengual. 2012. Bilingual language profile: An easy-to-use instrument to assess bilingualism. https://sites.la.utexas.edu/bilingual/.
Bittner, A. 1996. *Starke "schwache" Verben–Schwache "starke" Verben. Struktur des deutschen Verbsystems* [Strong "weak" verbs – Weak "strong" verbs. The structure of the German verb system]. Tübingen, Germany: Stauffenburg.
Blom, E., D. Polisenska & F. Weerman. 2006. Effects of age on the acquisition of agreement inflection. *Morphology* 16. 313–336. doi:10.1007/s11525-007-9110-1
Bosch, S. & H. Clahsen. 2015. Accessing morphosyntactic information in L1 and L2 word recognition: A masked priming study. In Grillo E., Jepson K. (eds.), *BUCLD 39: Proceedings of the 39th annual Boston University conference on Language development* (Vol. 1), 101–112. Somerville, MA: Cascadilla Press.
Clahsen, H. & C. Felser. 2006a. Grammatical processing in language learners. *Applied Psycholinguistics* 27. 3–42. doi:10.1017/S0142716406060024
Clahsen, H. & C. Felser. 2006b. How native-like is non-native language processing? *Trends in Cognitive Sciences* 10. 564–570. doi:10.1016/j.tics.2006.10.002
Clahsen, H. & C. Felser. 2017. Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*. doi:10.1017/S0272263117000250
Clahsen, H., L. Balkhair, J.-S. Schutter & I. Cunnings. 2013. The time course of morphological processing in a second language. *Second Language Research* 29. 7–31. doi:10.1177/0267658312464970
Clahsen, H., S. Eisenbeiß, M. Hadler & I. Sonnenstuhl. 2001. The mental representation of infl ected words: An experimental study of adjectives and verbs in German. *Language* 77. 510–543. doi:10.1353/lan.2001.0140
Clahsen, H., S. Gerth, V. Heyer & E. Schott. 2015. Morphology constrains native and non-native word formation in different ways: Evidence from plurals inside compounds. *The Mental Lexicon* 10. 53–87. doi:10.1075/ml
Corbett, G. & N. Fraser. 1993. Network morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29. 113–142. doi:10.1017/S0022226700000074
Cunnings, I. 2017. Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20. 659–678. doi:10.1017/S1366728916000675
DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22. 499–533.
DeKeyser, R. 2005. What makes learning second-language grammar difficult? A review of issues. *Language Learning* 55 (Supplement 1). 1–25. doi:10.1111/j.0023-8333.2005.00294.x
DeKeyser, R. 2010. Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals* 43. 80–92. doi:10.1111/(ISSN)1944-9720
DeKeyser, R., I. Alfi-Shabtay & D. Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31. 413–438. doi:10.1017/S0142716410000056
Durrell, M. 1980. Morphonologische und morpholexische Regelmäßigkeiten im deutschen Ablautsystem. *Jahrbuch für internationale Germanistik, Reihe A* 8. 19–28.
Durrell, M. 2001. Strong verb Ablaut in the West Germanic languages. *Zur Verbmorphologie germanischer Sprachen*.
Flege, J. 1999. Age of learning and second language speech. In Birdsong David (ed.), *Second language acquisition and the critical period hypothesis*, 101–131. Mahwah, NJ: Lawrence Erlbaum.
Forster, K. I. & J.C. Forster. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35. 116–124. doi:10.3758/BF03195503
Friedman, L. & M. Wall. 2005. Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician* 59. 127–136. doi:10.1198/000313005X41337
Granena, G. & M. Long. 2012. Age of onset, length of residence, language aptitude and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29. 311–343. doi:10.1177/0267658312461497
Granena, G. & M. Long. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29. 311–343. doi:10.1177/0267658312461497
Hakuta, K., E. Bialystok & E. Wiley. 2003. Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science* 14. 31–38. doi:10.1111/1467-9280.01415
Hartshorne, J.K., J.B. Tenenbaum & S. Pinker. 2018. A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition* 177. 263–277. doi:10.1016/j.cognition.2018.04.007
Heister, J., K. Würzner & J. Bubenzer. 2011. dlexdb - eine lexikalische Datenbank für die psychologische und linguistische Forschung [dlexdb - a lexical database for psychological and linguistic research]. *Psychologische Rundschau* 62. 10–20. doi:10.1026/0033-3042/a000029
Hippisley, A. 2016. Network morphology. In Hippisley A., Stump G. (eds.), *The Cambridge handbook of morphology*, 482–510. Cambridge: Cambridge University Press.

Hopp, H. 2016. The timing of lexical and syntactic processes in L2 sentence comprehension. *Applied Psycholinguistics* 37. 1253–1280. doi:10.1017/S0142716415000569
Huang, Becky. 2014. The effects of age on second language grammar and speech production. *Journal of Psycholinguistic Research* 43. 397–420. doi:10.1007/s10936-013-9261-7
Hyltenstam, K. & N. Abrahamsson. 2003. Maturational Constraints in SLA. In Doughty C. J., Long M. H. (eds.), *The handbook of second language acquisition*, 539–588. Oxford: Blackwell.
Jacob, G., V. Heyer & J. Veríssimo.2018. Aiming at the same target: A masked priming study directly comparing derivation and inflection in the second language. International Journal of Bilingualism 22.619–637.
Jessen, A., E. Fleischhauer & H. Clahsen. 2017. Morphological encoding in German children's language production: Evidence from event-related brain potentials. *Journal of Child Language* 44. 427–456. doi:10.1017/S0305000916000118
Johnson, J. & E. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21. 60–99. doi:10.1016/0010-0285(89)90003-0
Kaan, E. 2014. Predictive sentence processing in L1 and L2. What's different? *Linguistic Approaches to Bilingualism* 4. 257–282. doi:10.1075/lab.4.2.05kaa
Kirkici, B. & H. Clahsen. 2013. Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition* 16. 776–794. doi:10.1017/S1366728912000648
Köpcke, K.-M. 1998. Prototypisch starke und schwache Verben der deutschen Gegenwartssprache [Prototypically strong and weak verbs in present-day German]. *Germanistische Linguistik 141/142*. 45–60.
Krause, H., S. Bosch & H. Clahsen. 2015. Morphosyntax in the bilingual mental lexicon: An experimental study of strong stems in German. *Studies in Second Language Acquisition* 37. 597–621. doi:10.1017/S0272263114000564
Leminen, A. & H. Clahsen. 2014. Brain potentials to inflected adjectives: Beyond storage and decomposition. *Brain Research* 1543. 223–234. doi:10.1016/j.brainres.2013.10.038
Long, M. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12. 251–285. doi:10.1017/S0272263100009165
Long, M. 2013. Maturational constraints on child and adult SLA. In Granena Gisela, Long Mike (eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment*, 3–42. Amsterdam: John Benjamins.
Lukatela, G., C. Carello & M. Turvey. 1987. Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes* 2. 1–17. doi:10.1080/01690968708406349
Marcus, G. F., S. Pinker, M. Ullman, M. Hollander, T. J. Rosen & F. Xu. 1992. Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development* 57. 1–182.
Marian, V., H. Blumenfeld & M. Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50. 940–967. doi:10.1044/1092-4388(2007/067)
Marslen-Wilson, W. 2007. Morphological processes in language comprehension. In Gaskel G. (ed.), *The Oxford handbook of psycholinguistics*, 175–193. Oxford: Oxford University Press.
Matuschek, H., R. Kliegl, S. Vasishth, H. Baayen & D. Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94. 305–315. doi:10.1016/j.jml.2017.01.001
Mayberry, R.I. & R. Kluender. 2018.Rethinking the critical period for language: New insights into an old question from American Sign Language. Bilingualism: Language and Cognition 21(5). 886-905.
McDonald, J. 2006. Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55. 381–401. doi:10.1016/j.jml.2006.06.006
Meisel, J. 2013. Sensitive phases in successive language acquisition: The critical period hypothesis revisted. In Boeck C., Grohmann K. (eds.), *The Cambridge handbook of biolinguistics*, 69–85. Cambridge: Cambridge University Press.
Neter, J., M. Kutner, C. Nachtsheim & W. Wasserman. 1996. *Applied linear statistical models*. Chicago: Irwin.
Paster, M. 2016. Alternations: Stems and allomorphy. In Hippisley A., Stump G. (eds.), *The Cambridge handbook of morphology*, 93–116. Cambridge: Cambridge University Press.
Pinker, S. 1994. *The language instinct: How the mind creates language*. New York: Morrow.
Prévost, P. & L. White. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16. 103–133. doi:10.1191/026765800677556046
Raftery, A. 1995. Bayesian model selection in social research. *Sociological Methodology* 25. 111–163. doi:10.2307/271063
Ratcliff, R. 1993. Methods for dealing with reaction time outliers. *Psychological Bulletin* 114. 510–532. doi:10.1037/0033-2909.114.3.510
Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6. 461–464. doi:10.1214/aos/1176344136
Smolka, E., P. Zwitserlood & F. Rösler. 2007. Stem access in regular and irregular inflection: Evidence from German participles. *Journal of Memory and Language* 57. 325–347. doi:10.1016/j.jml.2007.04.005

Stemberger, J. P. 1982. The nature of segments in the lexicon: Evidence from speech errors. *Lingua* 56. 235–259. doi:10.1016/0024-3841(82)90012-2

Vanhove, J. 2013. The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PloS one* 8. e69172. doi:10.1371/journal.pone.0069172

Venables, W.N. & B.D. Ripley. 2002. *Modern applied statistics with S-plus*. New York: Springer-Verlag.

Veríssimo, J. 2018. Sensitive periods in both L1 *and* L2: Some conceptual and methodological suggestions [Commentary on Mayberry & Kluender, 2017]. *Bilingualism: Language and Cognition*. doi:10.1017/S1366728918000275

Veríssimo, J. 2019 *Taking it a level higher: A lemma-extended interactive activation model of complex word recognition*. Manuscript submitted for publication.

Veríssimo, J., V. Heyer, G. Jacob & H. Clahsen. 2018. Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition*. doi:10.1080/10489223.2017.1346104

Wagenmakers, E.-J. 2007. A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review* 14. 779–804. doi:10.3758/BF03194105

Werker, J.F. & T.K. Hensch. 2015. Critical periods in speech perception: New directions. *Annual Review of Psychology* 66. 173–196. doi:10.1146/annurev-psych-010814-015104

White, L. 2003. *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Wiese, B. 2008. Form and function of verbal ablaut in contemporary standard German. In Sackmann R. (ed.), *Explorations in integrational linguistics. Four essays on German, French, and Guaraní*, 97–151. Amsterdam/Philadelphia: Benjamins.

Wiese, R. 1996. *The phonology of German*. Oxford: Clarendon Press.

Wunderlich, D. 1996. Minimalist morphology: The role of paradigms. In Booij G., van Marle J. (eds.), *Yearbook of morphology 1995*, 93–114. Dordrecht: Kluwer Academic Publishers.

Wunderlich, D. & R. Fabri. 1995. Minimalist morphology: An approach to inflection. *Zeitschrift für Sprachwissenschaft* 14. 236–294. doi:10.1515/zfsw.1995.14.2.236

Wurm, L. & S. Fisicaro. 2014. What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language* 72. 37–48. doi:10.1016/j.jml.2013.12.003