Language Acquisition, 21: 386–410, 2014
Copyright © Taylor & Francis Group, LLC
ISSN: 1048-9223 print / 1532-7817 online
DOI: 10.1080/10489223.2014.892947

Routledge
Taylor & Francis Group

The Debate on Maturational Constraints in Bilingual Development: A Perspective from First-Language Attrition

Monika S. Schmid
University of Essex and University of Groningen

A controversial topic in research on second-language acquisition is whether residual variability and optionality in high-proficiency late second-language (L2) learners is merely the outcome of cross-linguistic transfer, competition, and processing limitations, or whether late learners have an underlying representational deficit due to maturational constraints on ultimate attainment in L2. This study argues that insights into this question can be gained by comparing advanced late L2 learners with late bilinguals who grew up with the language under investigation as their first language (L1), prior to emigrating to another country. The latter group, who use the language of the host country predominantly in their daily lives, typically exhibit increased optionality in their native language as a result of cross-linguistic transfer and L1 attrition. They do not, however, have a representational deficit in their L1, having acquired it monolingually during childhood. Such a comparison has the potential to distinguish grammatical features that are prone to bilingualism effects from those that natives can maintain but with which L2ers struggle persistently, possibly due to maturational limitations. This study compares 20 long-term attriters (English L2) with 20 highly advanced immersed learners of German (English L1) and 20 predominantly monolingual controls. The bilingual populations are matched for proficiency and for their use of German in daily life. The analysis comprises a group comparison and an investigation of individual performance, to assess whether there are L2 speakers who perform within the accuracy ranges of a larger population of attriters ($n = 53$) on all features, and similarly, whether any of the attriters perform within the accuracy range of a population of native controls ($n = 53$). The findings indicate that there are some areas of grammar (e.g., obligatory word order) where the L2 speakers are similar to the L1 attriters, and others (in particular noun phrase morphology) where attriters and monolinguals behave differently from the L2ers. This finding is interpreted as being consistent with an account that assumes some form of maturational constraint on language learning.

1. INTRODUCTION

One of the most contested issues in the field of second-language acquisition is the question of maturational constraints. The key question is whether the representation and processing of a second language (L2) is qualitatively different from that of the native language if the L2 has been learned after a certain age (often assumed to lie around puberty). The present article proposes that

Correspondence should be sent to Monika S. Schmid, Department of English Language and Linguistics, Faculty of Arts, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands. E-mail: m.s.schmid@rug.nl

insight can be gained into this question through studies that go beyond the comparison between monolinguals and second-language learners (L2ers), and include first-language (L1) speakers who are long-term bilinguals and use their L1 on an infrequent basis (attriters). Both attriters and L2ers experience cross-linguistic transfer from their stronger language (the L1 in the case of second-language learners, the L2 for attriters) to their weaker one, and these transfer phenomena have been argued to be of a similar nature (e.g., Sorace 2005). However, age-related constraints only apply in second-language acquisition, as first-language attrition affects the language that was acquired in childhood.

At present, there are two types of empirical approaches that have attempted to determine whether there are maturational constraints on second-language learning. Firstly, a number of studies have been conducted on relatively large populations, with a wide range of ages at onset and proficiency levels (e.g., Hakuta, Bialystok & Wiley 2003; Johnson & Newport 1989), in order to determine whether variance in proficiency levels is linearly correlated with age or whether there is some kind of discontinuity around puberty, indicating a maturational effect. The second approach focuses on carefully selected learner populations at very advanced levels of proficiency and attempts to determine whether these speakers have become fully nativelike or whether there are residual areas of nonnativeness (e.g., Hopp 2007; Abrahamsson & Hyltenstam 2009).

One of the earliest and most influential investigations of the former type of study (Johnson & Newport 1989) did indeed find a pattern interpreted to be consistent with a maturational effect: proficiency correlated with age of arrival (AoA) in a sample of 46 L2 learners of English for the younger segment of the population (AoA 3–15), but not among speakers who had learned the L2 after the age of 17. Other investigations have also taken the approach of dividing their population into AoA bins and establishing that among younger subjects, but not in the older groups, AoA correlates with eventual proficiency (DeKeyser 2003; DeKeyser, Alfi-Shabtay & Ravid 2010). This method of investigating the AoA-ultimate proficiency relationship separately and independently for different segments of the population has frequently been criticized, as such an approach has to make relatively arbitrary decisions on where the boundaries between the older and the younger population are to be set.[1] For example, changing the age cutoff in the data investigated by Johnson & Newport (1989) from 15 to 20 makes the nonlinear effect disappear (Bialystok & Hakuta 1994). An investigation of 240 Korean learners of English with varying ages of onset showed that an apparent maturational discontinuity was strongly affected by confounding variables and disappeared when these were controlled for in the statistical analysis (Flege, Yeni-Komshian & Liu 1999). Such findings suggest that while there is an undisputed correlation between age and ultimate attainment, the causative effect of age as an independent variable cannot be proven by these kinds of studies.

These findings point to a possible underlying problem with investigations attempting to probe age-of-onset mechanisms on the basis of random samples. Ultimate proficiency levels of L2 speakers vary enormously, even under conditions that are conducive to L2 acquisition (e.g., immersed learning among migrants). This is true for L2 populations of all ages: Even simultaneous bilinguals may end up with low proficiency levels in one of their languages, and often siblings who are brought up under similar conditions concerning the language policy at home, the

[1] It should be noted here that statistical tests of linear relationships have limited power when it comes to investigating correspondences that, by their very nature, are assumed not to be linear. Recent statistical advances, e.g., the work of Harald Baayen, have the potential of advancing these kinds of investigations.

education in both the environmental and the parents' language, and so on, will differ dramatically from each other in terms of their language skills (Montrul 2008; Seton 2011). It has often been pointed out that the main difference between L1 acquisition and bilingual acquisition is that monolinguals exhibit far less variance in their learning paths, learning rate, and ultimate grammar than bilinguals (Birdsong 2004).

It is uncontroversial that success in L2 learning is to some extent constrained by personal background variables such as motivation, aptitude, frequency and contexts of input, time and effort that the speaker is willing and able to devote to the language-learning process, as well as learning mechanisms and strategies (implicit vs. explicit learning, e.g., Paradis 2004; assimilation vs. accommodation, e.g., Bialystok 2001), identity and identification, etc. These factors partly determine how far along the path toward becoming a successful L2 speaker an individual learner will progress. They also covary with age, and their impact on success in L2 learning is not in doubt. What is unclear is whether beyond these factors, AoA also *independently* limits the potential *endpoint* that very advanced L2 learners may reach (but which most L2ers will not even get close to) through some form of maturational constraint on the learning mechanisms. This latter factor would imply that even the most successful late L2 learners may be unable to fully acquire certain features that are unproblematic for native speakers.

The first set of predictors—factors limiting *progress*—will impact on the full range of a bilingual's skills, while independent AoA effects limiting *potential* (if they exist at all) have been predicted to be subtle and to affect relatively few linguistic phenomena (see the studies in Snape, Leung & Sharwood Smith 2009). This suggests that in any large-scale study covering a representative sample of bilinguals, the strong and holistic impact of background factors will, in all probability, mask any possible independent contribution of age that might have led to a Critical Period effect for the most-advanced speakers. In other words, even if there is a discontinuity in ultimate attainment in the AoA range around puberty, it is likely to be lost in the noise of variance caused by other factors that also vary with age but for reasons not linked to language-specific biological or neurological maturation processes. Any potential discontinuity is therefore undetectable in large-scale studies of populations with varying proficiency levels.

This is particularly true for studies relying on self-ratings or census data, such as the one presented by Hakuta, Blalystok & Wiley (2003), where individual bilingual proficiency in English was based on self-ratings on a 3-point scale of "not well"–"well"–very well."[2] This scale cannot capture the intricacies of relatively small benefits predicted by the CP for the highest proficiency ranges (Stevens 2004). Furthermore, in such self-evaluations, speakers may not assess their own proficiency against an abstract model of nativeness but within the frame of reference of their own cohort: Given the prevalence of the belief in CP effects outside the linguistic community, and the uncontested statistical impact of AoA on L2 success, an L2 learner who became immersed at age 8 may gauge his or her proficiency against a very different standard than someone who migrated in their twenties.

In view of these limitations of cohort studies on AoA effects, investigations of maturational constraints to ultimate attainment should focus their efforts on the highest proficiency ranges. A number of recent studies have attempted to probe ultimate L2 proficiency among the "cream of

[2]The original scale used in the census includes 5 points, but since the extreme ends refer to monolingual proficiency in another language (1) and in English (5), these do not apply to the bilingual population, which is limited to the range 2–4.

the crop" by scrutinizing very advanced L2 learners and comparing their performance to that of a monolingual native reference population (Hopp 2007; Abrahamsson & Hyltenstam 2009). The goal of these investigations is to establish whether there are individual second-language learners who have demonstrably reached native levels of proficiency on certain grammatical features or linguistic measures, since evidence for the fact that some L2 learners *have* reached this level would strongly suggest that all of them *could* although few of them actually *do* (for reasons linked to the extralinguistic factors listed previously). The findings from these and other investigations are controversial, as some find nativelike performance among their speakers (Boxtel, Bongaerts & Coppen 2005; Hopp 2007), while others argue for persisting limitations, modulated by personal factors such as language aptitude (Abrahamsson & Hyltenstam 2009).

A problem for comparisons of advanced L2ers and monolinguals lies in the fact that there are inherent differences between monolinguals and bilinguals (Grosjean 1989; Cook 1995, 2002). For example, it was first demonstrated more than 30 years ago that proficient bilinguals experience phonetic assimilation toward the other language in both their linguistic systems (Flege 1987), and a similar assimilatory process toward phonetic settings supplied by the L2 has recently been found in the L1 of even novice instructed learners (Chang 2012). Furthermore, bidirectional transfer is not confined to phonetics but can be found across all linguistic levels (Jarvis & Pavlenko 2008; Schmid 2013b).

Thus, like the possible confound described previously in the discussion on ultimate attainment (concerning the uncontested impact of external, personal background factors on *success* in language learning and the controversial notion of maturational constraints), there may be a similar confound affecting linguistic representations. Here, too, two sets of factors can potentially account for the fact that monolingual natives and very-advanced L2ers still tend to perform differently when it comes to certain phonetic, phonological, and grammatical features. The first concerns those phenomena that demonstrably and uncontroversially affect bilinguals' processing and use of all of their languages, namely cross-linguistic interference and/or transfer affecting both production and processing (henceforth: *the bilingualism effect*). That this effect can be present and observable even in cases where there is no deficit in underlying linguistic representations is evident from the fact that even mature, strongly L1-dominant bilinguals experience it in their native language. It is therefore in all probability consistent with models of bilingual production that assume problems related to mapping due to computational limitations of the linguistic-cognitive system (e.g., the Missing Surface Inflection Hypothesis [MSIH] Prévost & White 2000, L. White et al. 2004). In other words, even though the underlying grammar may be intact, nontargetlike forms or responses may be observed if processing load exceeds capabilities, and in such cases, the speaker may rely on resources supplied by a language other than the one s/he is currently using or processing, particularly if the target language is not the dominant one.

While such processes of cross-linguistic interference undoubtedly underlie some instances of "deviant" linguistic production or processing (particularly where they are observed in a speaker's dominant language), a second assumption regarding late L2 grammars is strongly contested. It is expressed in Representational Deficit approaches to second-language learning (e.g., Hawkins & Chan 1997; Hawkins & Hattori 2006; Hawkins & Tsimpli 2009), which take L2 acquisition to be maturationally constrained in that late learners are unable to establish some grammatical features unless these are supplied by the L1. It is important to note here that representational deficit accounts do *not* assume that such features cannot be learned, as is often assumed (e.g., L. White et al. 2004; E. White, Genesee & Steinhauer 2012), but that in the absence of features

supplied by the L1, late learners will have to resort to compensatory, nongrammatical strategies (e.g., lexical context dependencies) in order to acquire them (Hawkins & Tsimpli 2009).

The problem in resolving this controversy lies in determining whether, in addition to an undisputed bilingualism effect, late L2 learners also have some form of representational deficit. This cannot be achieved through a comparison of late learners and monolinguals, since the latter not only demonstrably do not have an representational deficit but also do not experience cross-linguistic transfer, so differences between the populations could be attributed to either effect. Only comparisons involving populations impacted by one but not by the other can help resolve this issue.

It has thus recently been proposed that a more realistic way of assessing the impact of maturational limitations on ultimate attainment is to compare late L2 learners not to monolinguals but to simultaneous or early bilinguals (Singleton 2013). While this approach does have the benefit of allowing for variance caused by bilingual transfer (which a monolingual reference group does not experience), it does not eliminate the confound of potential versus actual attainment. Bilingual populations across the entire range of AoAs show considerably higher levels of variability than monolingual native controls. It is impossible to determine how much of this variation is the outcome of bilingual interference, causing online errors on a structure that has actually been fully acquired, or fossilization below a learner's actual potential (the mere fact that a learner *has not* learned something, for example, due to lack of time or motivation, does not necessarily mean that s/he *cannot*). For example, in a recent study of a population of 16 early or simultaneous Turkish-Dutch bilinguals, all of whom rated themselves as Dutch-dominant (with the exception of one speaker who considered himself a balanced bilingual), 50% performed at or below chance at detecting gender violations on an off-line judgment task in Dutch and also showed no electrophysiological response to such violations, while the Dutch monolingual native controls were at ceiling in the judgment task and had a strong P600 effect (Seton 2011). Given such levels of variance, it may be extremely difficult to differentiate fossilization from limitations imposed by maturational constraints based on comparisons of early and late bilinguals.

The solution to this dilemma is to test a reference population for whom acquisition of the language in question demonstrably reached the target levels of proficiency typically attained by monolinguals. Such a population might conceivably consist of carefully selected early or simultaneous bilinguals at advanced levels of proficiency, but here again it is extremely difficult to control for the confounding extralinguistic factors that impact bilingual development. Furthermore, it is doubtful that processes of cross-linguistic competition and interference operate in the same way for sequential acquisition as they do in simultaneous acquisition, where two or more languages are acquired at more or less the same time and the resulting system may be merged to some extent (and may become differentiated at a later age; Singleton 2013).

A more appropriate point of reference for measuring ultimate attainment and its possible limitations in late L2 acquisition are late bilinguals who have experienced language-dominance reversal (Hopp & Schmid 2013; Montrul 2008; Schmid 2013a). This process, also known as L1 attrition (L1A), can be observed in migrants who have lived in an L2 environment for extended periods of time, have become very proficient in the L2, and in many cases only rarely use their L1. Speakers for whom the L2 has become dominant can be assumed to experience competition effects and processing limitations arising from communication demands in their L1 that are similar to those that constrain L2 accuracy under the MSIH, and L1 attrition has been shown to lead to interference and nontargetlike language use akin to that found in L2 speakers (Sorace 2005).

L1 attrition has been the focus of a considerable amount of research over the past few decades, and it has been established that attriters exhibit phenomena of tranfer from their L2 into their native language across all linguistic levels. From early on in the process of second-language acquisition, L1 phonetic categories appear to shift toward the L2 values (Chang 2012), and this process of transfer and adaptation appears to increase with prolonged exposure (Flege 1987). Similar phenomena of shift toward values and preferences specified in the L2 can be observed for grammatical categories (see, e.g., Dussias & Sagarra 2007 for relative clause attachment, Gürel & Yılmaz 2011 for reference in Turkish overt and null pronouns, or Montrul 2008 for a variety of grammatical phenomena in L1 Spanish attriters and heritage speakers). The question of whether attrition may lead to an emerging representational deficit for such speakers has been addressed, and it is generally accepted that this does not usually happen in postpuberty attriters (Schmid 2013a). This implies that L1 attriters can supply the missing link for investigations of ultimate attainment in SLA, in that they experience one of the phenomena that may constrain this (bilingual interference) but not the other (maturational constraints or representational deficits).

Investigations that compare highly advanced (near-native) L2 speakers with L1 attriters at similar levels of proficiency have the potential to reveal areas of grammar that are susceptible to bilingualism effects in contrast to those where potential maturational constraints might limit ultimate attainment for L2ers but not for L1Aers. Both populations should show similar deviations from the monolingual native norm in the case of the former type of feature, but only the L2ers, and not the L1Aers, should differ from monolinguals on features whose acquisition is maturationally constrained. For linguistic phenomena impacted by bilingual interference, the two bilingual populations should differ from the monolingual controls, but not (or less so) from each other. For phenomena that are constrained by AoA, the two bilingual populations should be clearly differentiated from each other with little to no overlap, indicating failure of the late learners to approximate the performance of the L1 attriters. Attriters and monolinguals, on the other hand, should be similar to each other.

The present study tests these assumptions among a population of long-term L1 attriters and advanced immersed L2 learners of German. The attriters are immersed in an English-speaking environment in the Vancouver area, while the L2 learners all grew up with English as their native language but have been living in Germany for an extended period of time. English and German offer interesting grammatical contrasts that provide a challenge to both attriters and L2 learners, as German is morphologically more complex than English and has a different underlying word order.

2. GERMAN MORPHOLOGY AND WORD ORDER IN ACQUISITION AND ATTRITION

This section briefly summarizes the description of eight morphological and syntactic features of German and their acquisition in L1 and L2, as provided in Schmid (2002, 2004). These studies also investigate the accuracy and distribution of these features in free-speech data collected from 35 German-Jewish refugees who had lived in an English-speaking environment since the 1930s and were interviewed for an Oral History project in the 1990s. All of these grammatical features have been shown to be affected by L1 attrition to some extent, and error rates across these

grammatical categories are highly similar across different corpora and investigations (Stolberg & Münch 2010; Schmid 2010).

(1) **NP morphology**. German NPs are morphologically considerably more complex than English ones. Determiners, quantifiers, adjectives, and other elements within the NP are marked for case, gender, and number, and all these properties interact not only with each other but also with definiteness. The system contains a fair number of syncretisms, making it particularly challenging for L2 learners. The individual grammatical properties are discussed subsequently, and the overall system is summarized in Table 1.

- *Case*. German has four cases, and all NPs are obligatorily marked for case, usually on determiners, pronouns, adjectives, and only rarely on the noun itself. Children progress through regular stages of acquisition, mastering case marking by age 4 (Tracy 1984). English learners of German, who do have case in their L1 but mark it only on pronouns, tend to associate the sentence-initial position with the nominative and the postverbal position with the oblique, and to associate definiteness and animacy with the nominative/subject function (Jordens 1992).
- *Gender*. German nouns belong to one of three genders, marked on determiners, (anaphoric) pronouns, adjectives, quantifiers, and others. Only in rare cases can the gender of the noun be predicted based on its phonological or morphological form. Despite this complex and opaque system of gender concord, children master it early. In SLA, gender is one of the most frequently researched topics, and it has

TABLE 1
The Inflectional Paradigm of German NPs

	Masculine		*Feminine*		*Neuter*	
Sg						
	definite	**indefinite**	**definite**	**indefinite**	**definite**	**indefinite**
N	**der** grosse Berg	**ein** grosser Berg	**die** grosse Strasse	**eine** grosse Strasse	**das** grosse Tal	**ein** grosses Tal
G	**des** grossen Berges	**eines** grossen Berges	**der** grossen Strasse	**einer** grossen Strasse	**des** grossen Tales	**eines** grossen Tales
D	**dem** grossen Berg	**einem** grossen Berg	**der** grossen Strasse	**einer** grossen Strasse	**dem** grossen Tal	**einem** grossen Tal
A	**den** grossen Berg	**einen** grossen Berg	**die** grosse Strasse	**eine** grosse Strasse	**das** grosse Tal	**ein** grosses Tal
Pl						
	definite			**indefinite**		
N	**die** grossen Berge/Strassen/Täler			grosse Berge/Strassen/Täler		
G	**der** grossen Berge/Strassen/ Täler			grosser Berge/Strassen/Täler		
D	**den** grossen Bergen/Strassen/Tälern			grossen Bergen/Strassen/Tälern		
A	**die** grossen Berge/Strassen/Täler			grosse Berge/Strassen/Täler		

der grosse Berg (masc.) 'the big mountain,' *die grosse Strasse* (fem.) 'the big street,' *das grosse Tal* (neut.), 'the big valley'; *Sg = Singular, Pl = Plural, N = Nominative, G = Genitive, D = Dative, A = Accusative.*

been shown to be persistently difficult in late L2 acquisition in particular for speakers whose L1, like English, does not encode grammatical gender (Rogers 1987; Hopp 2013).

- *Plural*. Plural marking on German nouns is characterized by a high degree of allomorphy. There are five different suffixes, three of which may combine with umlauting of the stem vowel. The definite plural determiner is identical to the definite feminine singular and does not encode gender. Which allomorph an individual noun takes is difficult to predict in most cases, and plural allomorphy is problematic for both L1 and L2 learners. Children continue to use nontargetlike suffixes until the age of 6 (Phillips & Bouma 1980; Schaner-Wolles 1988) and plural allomorphy also presents a persistent problem in late L2 acquisition (Clahsen 1995).

(2) **VP morphology**. Formally, the German system of VP inflection is very similar to that of English. There are three main differences: Firstly, while the verbal inflectional paradigm, comprising weak verbs (which form their past tense through predictable suffixation) and strong verbs (which are inflected by ablauting), is similar to the English one, English has relatively few highly frequent strong verbs, but in German, about half of all verbs (of both high and low frequency) are strong. Both in L1 and in L2 acquisition, weak forms are overgeneralized to strong contexts. Secondly, German has two past tense auxiliaries, *haben* 'have' and *sein* 'be.' Thirdly, German verbs are marked for person and number in both present and preterite.

(3) **Obligatory word order**. German is a largely free word-order language, allowing topicalization and scrambling of most sentence constituents. Obligatory rules concern the position of the finite verb, leading to three distinct patterns with no equivalent in English (which has fixed SVO word order in assertive sentences):

- *V2 in main clauses*. The preverbal position of the main clause can be occupied by any constituent but may not contain more than one constituent. If an element other than the subject is topicalized, the finite verb has to precede the subject.
- *Discontinuous word order (DWO)*: In constructions involving a (modal) auxiliary and a nonfinite main verb form (infinitive, participle) or verbs with a separable particle, other sentence constituents may intervene between the finite and nonfinite parts of the verb.
- *Verb-final in subordinate clauses*: In subordinate clauses in which the clause-initial position is occupied by a subordinating conjunction or relativizer, the finite verb has to appear in sentence-final position.

German word order has been shown to be acquired in a reliable, fixed sequence in children, who use verbal elements in second and final position from the start. L2 learners, on the other hand, start out with SVO, consecutively mastering the separable particle rule, subject-verb inversion, and verb final in subordinate clauses (Clahsen 1982).

German morphology and word order thus pose a number of challenges for speakers acquiring it from a native English background or maintaining it alongside an English L2 that is used frequently in daily life.

3. RESEARCH QUESTIONS

The present study investigates cross-linguistic interference and instances of nontargetlike structures in the German of late learners of German who are native speakers of English and L1 speakers of German who have been immersed in an L2 English environment for an extended period. The following research questions are addressed:

RQ1: Are there grammatical features of German that long-term immersed bilinguals for whom German is either the L1 or the L2 use in a similar manner, but for which their performance differs from that of a monolingual reference population (bilingualism effect)?

RQ2: Are there grammatical features of German on which long-term immersed bilinguals for whom German is the L1 differ markedly from speakers for whom it is the L2 (maturational constraints)?

RQ3: If areas of grammar in which the two bilingual populations differ from each other (see RQ2) are found, are there individuals in the L2 German population who score within the range of a large population of L1 attriters (no maturational constraints)?

4. MATERIALS AND METHODOLOGY

The following analysis consists of two parts. The first presents a comparison of two bilingual populations (highly advanced, immersed late L2ers and largely L2-dominant long-term L1Aers of German with English as the contact language) matched on measures of general proficiency, self-evaluation, and length of residence in a German-speaking environment. The performance of the groups on a range of controlled tasks as well as their accuracy in two spontaneous speech samples per speaker are compared to each other and to a predominantly monolingual reference group, in order to determine on which grammatical features the bilingual populations are similar to each other but different from the monolinguals (RQ1), and in which areas the attriters perform more like the controls and less like the L2ers (RQ2). The second part seeks to determine whether at least some of the L2ers manage to achieve an accuracy score that falls within the range of a wider sample of attriters on those features that were shown to be more difficult for the L2 population (RQ3). The goal of this analysis is to determine whether there are any grammatical features that elude even the most advanced among the L2 learners.

All participants were healthy, adult volunteers who contacted the author voluntarily in answer to calls for participation disseminated through local newspapers, German clubs, churches, word of mouth, etc.

4.1. Participants

This study investigates three populations of German speakers (see Table 2 for the biographical data):

- Reference population (henceforth RP, $n = 53$). This population consists of predominantly monolingual speakers of German who grew up with German as their only native language. It is becoming increasingly difficult to find fully monolingual speakers in present-day

TABLE 2
Participant Characteristics

	RP			*L1A*			*L2*		
	Mean	*SD*	*Range*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*
Age at testing	60.9	11.6	39–91	63.2	10.9	37–88	51.90	12.7	31–79
Age at emigration				26.2	7.1	14–47	25.3	4.0	20–39
Length of residence				37.1	12.4	9–54	26.6	12.0	7–53
Length of residence in Germany				26.2	7.12	14–47	26.6	12.0	7–53

European countries, as foreign-language teaching has become an obligatory part of the school curricula. The speakers in this population mostly had some instruction in English (and, in some cases, in other languages), and attained varying levels of proficiency in these languages. None of them uses a language other than German regularly in their daily lives, and none of them has ever lived outside Germany for any length of time. Educational levels were diverse: 13 participants had completed the German *Volksschule* or *Hauptschule* (the minimum schooling requirement, usually followed by an apprenticeship), 23 obtained the certificate necessary for most clerical apprenticeships (*Realschule* or *Mittlere Reife*), 6 had completed high school, and 11 had a university education.

- Late first-language attriters of German ($n = 53$). All speakers in this population grew up in Germany with German as their only native language. They emigrated to Canada after age 17 (with the exception of one speaker, the wife of one of the originally recruited participants, who was 14 at the time of emigration). The minimum period of residence in Canada was 9 years. The attriters were matched with the reference population for age and educational background (*Volksschule/Hauptschule* $n = 13$, *Realschule/Mittlere Reife* $n = 22$, high school $n = 5$, university $n = 13$).
- Advanced late second-language learners of German ($n = 20$). This population consisted of speakers who grew up in an English-speaking environment with English as their only native language (with the exception of two speakers who also spoke Gaelic at home). They emigrated to Germany after age 20, and the minimum period of residence there was 7 years. The participants in this group were more highly educated than the others, all of them having either completed high school ($n = 6$) or university ($n = 14$). The recruitment procedure explicitly specified that very advanced to near-native speakers of German were sought.

As is evident, the match between populations on biographical factors is not perfect. Since the aim of the present study is to test whether L2 learners under favorable conditions can match the performance of L1 speakers under unfavorable conditions, it was deemed acceptable for populations to vary on criteria that are usually considered impact performance of linguistic tasks as long as the difference favored the L2 population—for example, the average higher level of education and lower age at testing among these speakers. While the L1Aers have lived longer in a

bilingual environment than the L2ers, both populations spent on average the same amount of time in Germany (the attriters before, the L2ers since their migration).

4.1.1. Use of German by the Bilingual Populations

Input and use of both L1 and L2 are usually seen as critical factors for success in L2 acquisition and L1 maintenance. The linguistic habits of the bilingual populations in daily life were assessed by means of a questionnaire comprising 78 items on personal background, language use, and language attitude (the full questionnaire is available at http://www.let.rug.nl/languageattrition/SQ). For the purpose of matching the populations, overall average use of German in daily, informal contexts was calculated based on nine individual questions (Cronbach $\alpha = .78$):

- overall frequency of use of German (1 question)
- frequency of use of German within the family (5 questions)
- frequency of use of German with friends (3 questions)

All questions were 5-point Likert-Scale items coded on a scale of 0 to 1 (1 = *extremely frequent or exclusive use of German in this context;* 0 = *no or extremely little use of German in this context*). A comparison of the two populations showed that the average use of German in these contexts was very similar, with .46 for the attriters ($SD = .25$) and .45 for the L2ers ($SD = .21$). Unsurprisingly, the difference between groups is not significant, $t\ (71) = .064$, $p = .949$. The scatterplot in Figure 1 shows the frequency of use by length of residence. Although this figure suggests a decrease in L1 use with longer periods of residence for the L1Aers, this correlation is not significant for either population (L1A $r = –.173$, $p = .216$; L2 $r = .004$, $p = .985$).

The impact of the L1 use on language attrition and maintenance has been shown to be a complex matter (Opitz 2013; Schmid 2007; Schmid & Dusseldorp 2010). For example, Schmid (2011b) suggests that both very frequent and very infrequent L1 use may accelerate attrition, as her study found the strongest attrition effects among those speakers who used their L1 the most and the least, with intermediate populations having higher levels of language maintenance. She suggests that in the former context, speakers receive a great deal of input from other bilinguals who may not be targetlike in a consistent way and that this unreliability of the input may lead to accelerated language change as suggested by Grosjean & Py (1991). Speakers who use their L1 very infrequently or not at all, on the other hand, may experience attrition through a process of atrophy.

It is not the purpose of the present investigation to contribute to the debate on the impact of external factors on attrition but to assess the amount of interference that any speaker might likely experience after a long time in an L2 environment. Given the long periods of residence of some of the attriters and the large variety in language use, we have to assume that there are at least some attriters here for whom the L1 has deteriorated as much as one might reasonably expect.

4.1.2. Selection of Populations Matched for Proficiency in German

Part I of the present investigation aims to determine whether some areas of grammar in spontaneous speech production are more problematic to acquire for L2ers than they are to maintain for L1Aers. In order to ensure that any differences found could not be ascribed to effects of general proficiency, a subsample from the L1A group was established that would match the L2 group on

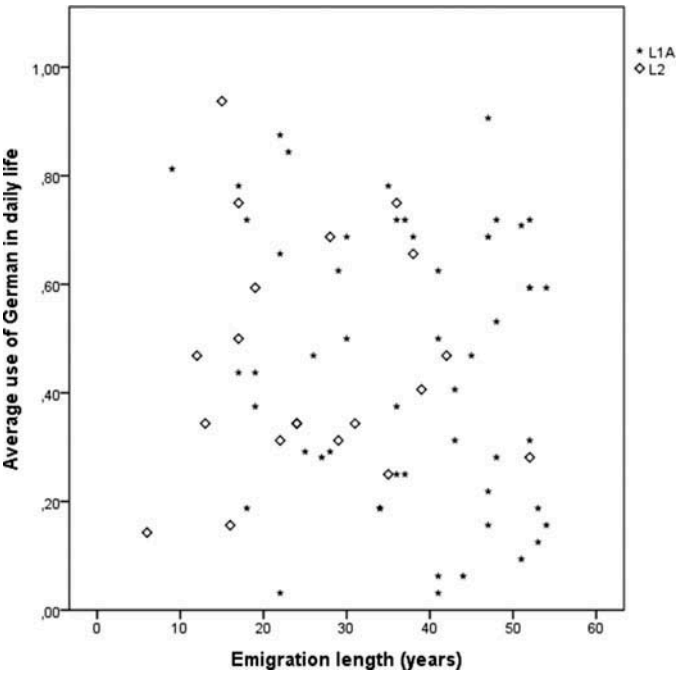

FIGURE 1 Self-reported use of German in informal contexts by length of residence.

this variable. This match was initially established on the basis of the scores that each individual participant had achieved on a C-Test and then confirmed on the basis of two more tasks, a self-evaluation and a verbal-fluency task.

The C-Test is a version of a fill-in or cloze task in which parts of words are deleted from a text according to a predetermined schema. The first sentence of the text is left intact, in order to establish context. Beginning with the second sentence, the second half of every second word is deleted (in words with uneven numbers of letters one more is deleted than is left). The C-Test is generally accepted to be a good indicator of proficiency in more-advanced learner groups and has often been used in L1 attrition studies (for an overview see Schmid 2011a). The present study used five short texts of 80–100 words in length. Each text contained 20 gaps, so that the maximum total score was 100.

On this task the RP group scored 83.43 (*SD* 9.0), the L1A group scored 77.08 (*SD* 11.5), and the L2 group 74.30 (*SD* 13.0). A matched subpopulation of L1Aers was established by pairing each L2 speaker with an attriter who had received as close as possible to the same score. A perfect match could not always be found since, for example, two L2ers scored 85 but only one attriter had this score, so one L2er was paired with an attriter whose score was 84 (this was compensated for in another case with a similar conflict where an attriter with the nearest higher score was chosen). In total, 20 L1Aers with a mean C-Test score of 74.25 (*SD* 13.1) were chosen. A subsample (n = 20) of the RP group was then selected by ranking all attriters and all controls in order of their C-Test score and selecting those controls whose rank matched that of the attriters who had been

TABLE 3
Proficiency Measures across the Three Populations

		C-Test			*Verbal fluency*			*Can-Do scales*		
Population	*n*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*
RP	53	83.5	8.9	59–99	25.8	4.6	16.5–37.5	3.8	.5	2.8–4.7
RP subset	20	80.3	9.9	59–94	26.1	5.1	17.5–37.5	3.7	.6	2.8–4.7
L1A	53	77.1	11.5	48–95	21.8	4.7	12–34.5	3.9	.6	2.6–4.7
L1A subset	20	74.3	13.1	51–95	22.6	3.6	17–32.5	4.2	.5	3.2–4.7
L2	20	74.3	13.0	54–95	22.9	4.7	16.5–32	4.2	.5	3.1–5.0

selected (e.g., if the third-best attriter had been matched with an L2er, the third-best control was also used). This subpopulation of the RP group had a mean C-Test score of 80.32 (*SD* 9.9).

In order to assess whether the proficiency match between the bilingual populations was valid, two more-formal language measures were used (for a full overview see Table 3). The first consisted of a set of two verbal-fluency tasks assessing the speakers' productive vocabulary. In this task, the speaker is cued to name as many items belonging to a certain semantic or phonetic category (e.g., animals, or words beginning with the letter *s*) as he or she is able to within a given time period. The present study used two semantic categories ("animals" and "fruit and vegetables," in line with previous investigations of L1 attrition, see Schmid & Köpke 2008) and allowed 60 seconds for each task. Bilinguals have been shown to be less productive on this task (Gollan, Montoya & Werner 2002), possibly due to the fact that they have overall more items to select from (Schmid & Köpke 2008). In our matched subset sample, the monolinguals outperformed the bilinguals with an average productivity of 26.11 (*SD* 5.1) compared to 22.6 (3.6) for the L1Aers and 22.9 (4.7) for the L2ers. The overall group difference was significant, $F(2, 57) = 3.579$, $p < .05$) with the RP being different from the L1Aers ($p < .05$) but not from the L2ers ($p = .081$) and the two bilingual populations not differing from each other ($p = .977$) in a post hoc test (Tukey HSD).

The last general measure of proficiency applied here was self-assessment. Participants were asked to indicate on a 5-point Likert scale how easy or difficult they found it to perform certain tasks in German. These Can Do scales were drawn from the ALTE descriptors used in the Common European Framework of Reference (http://www.coe.int/t/dg4/education/elp/elp-reg/CEFR_EN.asp), and the statements chosen were the ones describing the two most-advanced levels (C1 and C2). A total of 43 items were used across four skills (listening, speaking, reading, and writing), comprising statements such as: "I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided that I have some time to get familiar with the accent." (The full set of items can be found at http://www.let.rug.nl/languageattrition/CanDo). Items were coded on a scale of 1 (*I cannot do this*) to 5 (*I can do this without any difficulty*). An average score across all 43 items was calculated for each individual.

On this self-rating task, the two matched bilingual subpopulations again attained very similar results, with the attriters rating themselves at an average of 4.15 (*SD* .45) and the L2ers at 4.17 (.53). Interestingly, the monolingual controls (who outperformed the bilinguals on all formal tasks) rated their own proficiency significantly lower at an average of 3.66 (.62), $F(2, 57) =$

5.369, $p < .01$. Their self-rating differed from both bilingual groups at $p < .05$ in a post hoc test (Tukey HSD), while the two bilingual populations were again revealed to be very similar ($p = .992$). This finding underscores the point made previously that in self-assessment tasks, even when a very detailed scale is used and discriminators are clearly labeled, people tend to rate themselves not against an abstract, invariate ideal but against their own cohort. In the bilingual populations, the self-ratings correlate strongly with the C-Test score ($r = .56$, $p < .001$) and thus probably provide a good additional measure of proficiency across cohorts, while such a correlation is absent among the controls ($r = .020$, $p = .88$).

The scores attained by the two bilingual subpopulations on these three formal tasks thus suggest that their overall proficiency levels in German are very similar and that the L2ers have indeed attained a very-advanced command of this language. The question now is to what extent they are able to apply this proficiency in spontaneous speech and whether the two matched bilingual populations will perform differently under such circumstances.

4.2. Materials: Spontaneous Speech Samples

The production of spontaneous speech is arguably the most complex task for bilingual speakers, since information from all linguistic levels, as well as from other cognitive resources, has to be rapidly integrated in real time. Unlike most formal and experimental tasks, it also does not specifically tap into metalinguistic knowledge—speakers are simply invited to use their linguistic skills in the same way as they normally do.

Two speech samples were collected from each participant: a semistructured autobiographical interview (INT), lasting between 30 and 90 minutes per participant, and a retelling of a 10-minute excerpt from the silent Charlie Chaplin movie *Modern Times* (CC), lasting between 5 and 10 minutes.[3] For the three subpopulations, this resulted in two corpora of approximately 140,000 words (INT) and 45,000 words (CC). The full details on total and average length and range of the corpora are given in Table 4 for both the subcorpus used in the first part of the analysis and the full corpus used to establish the range effects of the individual bilingual speakers in the second part below.

TABLE 4
Total and Mean Length and Range (in Words) per Corpus and Population

		INT			*CC*		
Population	*n*	*Total words*	*Mean*	*Range*	*Total words*	*Mean*	*Range*
RP	53	93,029	1789.02	390–8,508	37,911	715.30	173–2,302
RP subset	20	28,876	1519.79	390–3,152	13,864	729.68	173–2,302
L1A	53	178,128	3425.54	1,025–7,469	38,362	737.73	376–1,617
L1A subset	20	72,718	3635.90	1,492–6,002	16,487	824.35	376–1,617
L2	20	40,872	2043.60	537–4,150	14,437	759.84	258–1,336

[3]Due to malfunctioning equipment for one of the L2 speakers, only the interview was recorded.

All recordings were transcribed and coded for errors in the eight morphosyntactic categories discussed previously by a native speaker of German. Transcription and error coding were checked against the original recordings by two other native speakers. A maximally lenient policy was adopted, and errors were only counted if the form was unacceptable not only in standard German but also in the regional varieties the speaker had been exposed to. In case of doubt (e.g., whether a certain form was a case or gender error), the most likely category was chosen based on indicators such as context or inflection of other items within the same phrase. The total number of errors in each category was assessed per speaker and then recalculated per 1,000 spoken words (excluding features such as hesitations, false starts, repetitions) in the relevant sample.

In the coding process it became evident that there were two further error categories not originally included in the overview presented in Schmid (2002), since errors of that kind are rare in monolingual or attrited native German, but were frequent in the data from the L2ers. These categories consisted of omissions of articles or pronouns. The high incidence of omissions of items in these word categories may indicate problems with either case or gender marking, with L2ers who were not certain of how to inflect a particular determiner or pronoun simply opting to omit it (or phonetically suppress it to the point where it was no longer audible).

5. RESULTS

5.1. Group Results

Group comparisons of all error categories were conducted by means of one-way ANOVAs with Tukey HSD post hoc tests. The full analyses, including descriptive statistics, are given in Table 5. Table 6 presents an overview of the mean error rate per 1,000 words for all three populations across the two corpora and summarizes the results of the post hoc tests of the error rates of the L1Aers compared with the controls, and of the L2ers compared with the L1Aers.

It is evident from these analyses that under the cognitive and communicative pressures of free-speech production, the late learners of German are considerably less accurate on morphological and syntactic features which are different in their two languages, than the L1Aers—despite the fact that their overall proficiency levels and language use habits are very similar. The L1Aers, on the other hand, despite their long-term immersion in an English-speaking setting, are not less accurate than the monolingual controls on any of the features investigated here. The differences between the populations are particularly marked in the area of NP morphology, as the descriptive statistics and effect sizes indicate. These discrepancies are illustrated in the graphical representation of the bilingual groups' performance on NP morphology (Figure 2a) and word order (Figure 2b).

Recall that in the debate on maturational constraints in late second-language acquisition there are two competing views: one which assumes that L2 learners of all ages can build targetlike representations but may fail to apply them consistently due to competition from their L1 and limitations of resources (the MSIH), and one which holds that learners retain a representational deficit for some grammatical features not instantiated by their L1, but that they can learn to approximate the linguistic behaviors of natives based on compensatory but nongrammatical strategies. Where L1 attrition is concerned, only the first of these explanations accounts for

TABLE 5
Accuracy on Morphological and Syntactic Features in Two Spontaneous Speech Samples: Descriptive and Inferential Statistics

		RP		*L1A*		*L2*		*ANOVA*			*Tukey*		
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*F(2, 57)*	*p*	η^2	*RP vs. L1A*	*RP vs. L2*	*L1A vs. L2*
INT													
	Case	.21	.47	.35	.38	6.11	5.13	24.851	<.001	.47	.987	<.001	<.001
	Gender	.03	.13	.11	.22	5.32	4.17	3.682	<.001	.52	.994	<.001	<.001
	Plural	.03	.14	.04	.15	1.48	1.53	16.960	<.001	.38	.999	<.001	<.001
	Article	.00	.00	.03	.09	1.42	1.40	19.448	<.001	.41	.993	<.001	<.001
	Pronoun	.00	.00	.06	.12	1.13	1.13	18.308	<.001	.40	.963	<.001	<.001
	Tense weak/strong	.00	.00	.01	.04	.00	.00	.974	.384	.03	.459	1.000	.450
	Auxiliary	.00	.00	.07	.13	.21	.44	3.208	<.05	.10	.708	<.05	.214
	Subject-verb agreement	.09	.22	.02	.07	.71	.69	16.121	<.001	.37	.844	<.001	<.001
	V2	.05	.15	.36	.45	2.36	2.74	11.861	<.001	.30	.816	<.001	<.001
	DWO	.05	.23	.16	.30	.75	.83	9.714	<.001	.26	.794	<.001	<.01
	Subordinate clause	.14	.33	.10	.26	.44	.71	2.979	.059	.10	.961	.136	.073
CC													
	Case	.59	.54	.36	.67	7.94	5.39	34.166	<.001	.55	.973	<.001	<.001
	Gender	.14	.43	.00	.00	1.34	7.81	33.706	<.001	.55	.995	<.001	<.001
	Plural	.07	.31	.03	.14	.14	.45	.600	.552	.02	.917	.777	.527
	Article	.00	.00	.50	.22	1.85	3.38	5.701	<.01	.17	.281	<.01	.144
	Pronoun	.00	.00	.00	.00	1.17	1.24	17.293	<.001	.39	.964	<.001	<.001
	Tense weak/strong	.12	.53	.30	.68	.31	.83	.446	.643	.02	.713	.674	.997
	Auxiliary	.00	.00	.16	.48	.47	1.05	2.485	.093	.08	.747	.083	.306
	Subject-verb agreement	.00	.00	.00	.00	.90	1.24	1.337	<.001	.27	1.000	<.01	<.01
	V2	.10	.43	.45	.98	3.14	3.39	12.729	<.001	.32	.852	<.001	<.001
	DWO	.17	.52	.22	.54	.37	1.01	.371	.692	.01	.978	.690	.802
	Subordinate clause	.03	.14	.35	.63	.12	.40	5.100	<.01	.16	.963	<.05	<.05

nontargetlike behavior in the L1 (due to competition from L1), as there is no maturationally determined representational deficit constraining L1 acquisition in these speakers (Schmid 2013a).

The fact that there is such a marked discrepancy between the two groups of bilinguals in the present data could thus either be attributed to an underlying difference in grammatical representation between those speakers who learned German in childhood on the one hand (controls and attriters) and those who learned it later in life (L2ers) on the other, or it might be the outcome of differential degrees of cross-linguistic competition and influence, with the L2ers more strongly affected by this factor than the attriters. Despite the long period of residence and immersion that both bilingual populations have, it is difficult to exclude the latter possibility based solely on group data, which might mask exceptionally low levels of cross-linguistic interference in the speech of individual L2ers or similarly high levels of inaccuracy among certain attriters. The

TABLE 6
Accuracy on Morphological and Syntactic Features in Two Spontaneous Speech Samples: Summary Statistics

	Mean errors per 1,000 words								*Tukey HSD*			
	RP		*L1A*		*L2*		*η2*		*L1A vs. RP*		*L2 vs. L1A*	
	INT	*CC*	*INT*	*CC*	*INT*	*CC*	*INT*	*CC*	*INT*	*CC*	*INT*	*CC*
NP												
Case	.21	.59	.35	.36	6.11	7.94	.47	.55	n.s.	n.s.	<.001	<.001
Gender	.03	.14	.11	.00	5.32	1.34	.52	.55	n.s.	n.s.	<.001	<.001
Plural	.03	.07	.04	.03	1.48	.14	.38	.02	n.s.	n.s.	<.001	n.s.
Art	.00	.00	.03	.50	1.42	1.85	.41	.17	n.s.	n.s.	<.001	n.s.
Pro	.00	.00	.05	.00	1.13	1.17	.40	.39	n.s.	n.s.	<.001	<.001
VP												
Weak/strong verbs	.00	.12	.01	.30	.00	.31	.03	.02	n.s.	n.s.	n.s.	n.s.
Auxiliaries	.00	.00	.07	.16	.21	.47	.10	.08	n.s.	n.s.	n.s.	n.s.
Subject-verb agreement	.09	.00	.02	.00	.71	.90	.37	.27	n.s.	n.s.	<.001	<.01
Word order												
V2	.05	.10	.36	.45	2.36	3.14	.30	.32	n.s.	n.s.	<.001	<.001
DWO	.05	.17	.16	.22	.75	.37	.26	.01	n.s.	n.s.	<.01	n.s.
sub	.14	.03	.10	.35	.44	.12	.10	.16	n.s.	n.s.	n.s.	<.05

following section will therefore consider the extent to which the individual attriters performed within the range of the controls and similarly whether individual L2ers can perform within the range of the attriters.

5.2. Individual Results

The second analysis focused on the performance of the 20 individual L2 learners and L1 attriters whose accuracy on morphological and syntactic features of German was investigated in the previous group comparison. For this analysis, a different baseline was adopted for each of the two groups: The error scores of the attriters were compared to those of the 53 predominantly monolingual controls from which data had originally been collected, while those of the L2ers were assessed against those of the 53 attriters. The decision to include the largest number of speakers available in the individual comparison was motivated by the desire to give each individual L2er and L1Aer the best possible chance to score within the range of the baseline population. For each error category and each corpus (CC and INT) it was determined whether an individual bilingual speaker fell within the reference population's range of error scores. For example, the highest number of gender errors per 1,000 words found for any of the 53 control speakers in the INT data was 1.11, while in the CC data, it was 1.89 (these error scores came from two different speakers). Any attriter who scored below 1.11 on the INT and below 1.89 on the CC was given 2 points, if they scored below one but above the other they received 1 point, and if they were above both,

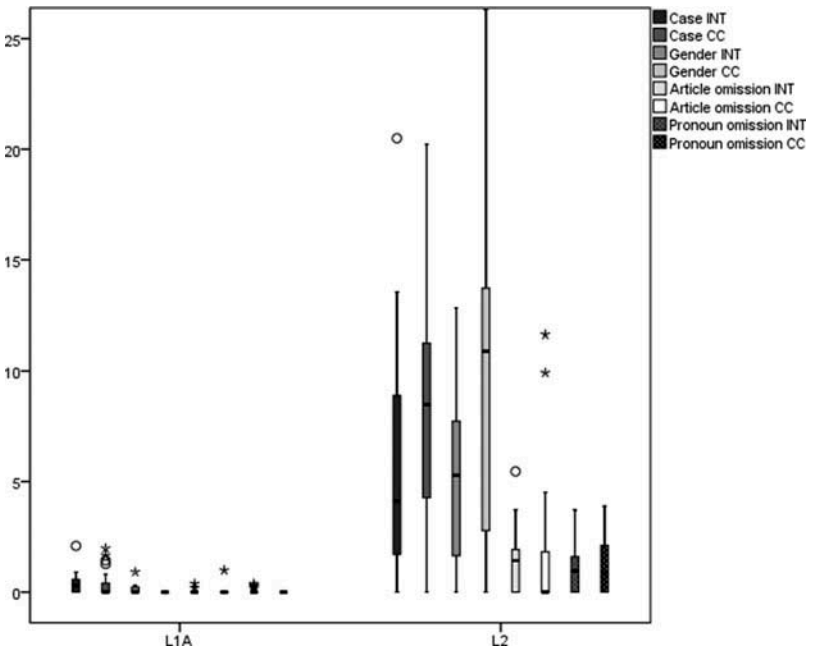

FIGURE 2a A comparison of errors per 1,000 words in the domain of NP morphology across samples and bilingual populations.

they received none. The same procedure, *mutatis mutandur*, was adopted for the comparison of the L2ers against the 53 attriters.

The outcome of this analysis is graphically represented in Figure 3. Results are presented in columns by participant from left to right in descending order of C-Test score as compared to the participant's cohort (i.e., the leftmost column represents the speaker with the highest score of their cohort, the rightmost one that with the lowest—recall that due to technical problems, only one speech sample was recorded from one of the L2ers, so there are only 19 individuals in this analysis). A lightly shaded box represents an error category in which a speaker fell within the range of the baseline population (the monolingual controls for the L1Aers, the attriters for the L2ers) in both speech samples, i.e., a score of 2. An intermediately shaded box symbolizes a score of 1 in the relevant category (i.e., this speaker had only one speech sample that fell within the baseline range), while a dark shade represents a score of 0 (neither sample was within the baseline range).

It is evident from this overview that there is considerable variability across the L2 population and that some of the speakers, particularly at the higher end of the proficiency range, have attained very high levels of accuracy. However, it is also evident that there is not a single L2er in this study who scores within the attriter range in both samples across all features. NP morphology again emerges as the most challenging grammatical aspect for the L2 learners. While for all other categories at least some of the speakers in the L2 population attained an accuracy score within the range of the attriters in both their speech samples, NP morphology and, in particular,

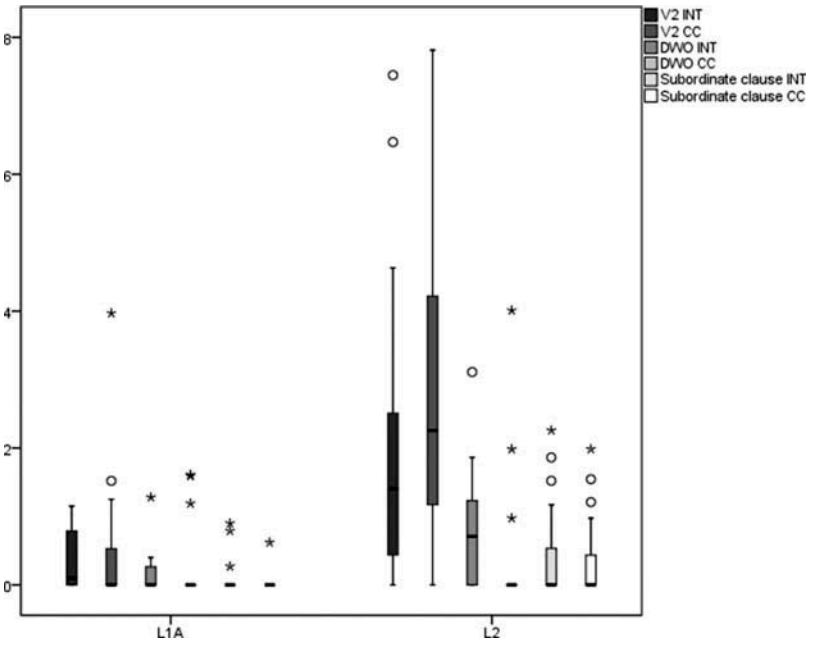

FIGURE 2b A comparison of errors per 1,000 words in the domain of word order across samples and bilingual populations.

grammatical gender appear to elude even the most proficient among these speakers. Grammatical case appears to be almost equally problematic (these observations are compounded by the large number of article and pronoun omissions, which could be an indication of either case or gender errors). Plural allomorphy and concord appear to be less challenging, and all L2ers have at least one speech sample in which they reach the attrition baseline. For all of these features the attriters, on the other hand, compare quite favorably with the unattrited controls.

Where VP morphology is concerned, attriters and L2ers appear similar, with the exception of subject-verb agreement, which poses more of a problem for L2ers than for L1Aers. Seven of the 19 L2ers, however, appear to be exempt from this and score consistently within the attriter range. Similarly, for word order, several L2 speakers perform within this range on all of the features investigated here.

6. DISCUSSION

The present investigation attempted to identify the areas of German morphology and grammar in which very advanced late L2 learners can or cannot match the performance and accuracy levels of L1 attriters speakers in spontaneous speech. In order to provide insight into the controversial question of whether deviances from the native norm result from cross-linguistic interference or from maturational constraints and representational deficits among L2ers, the study compared

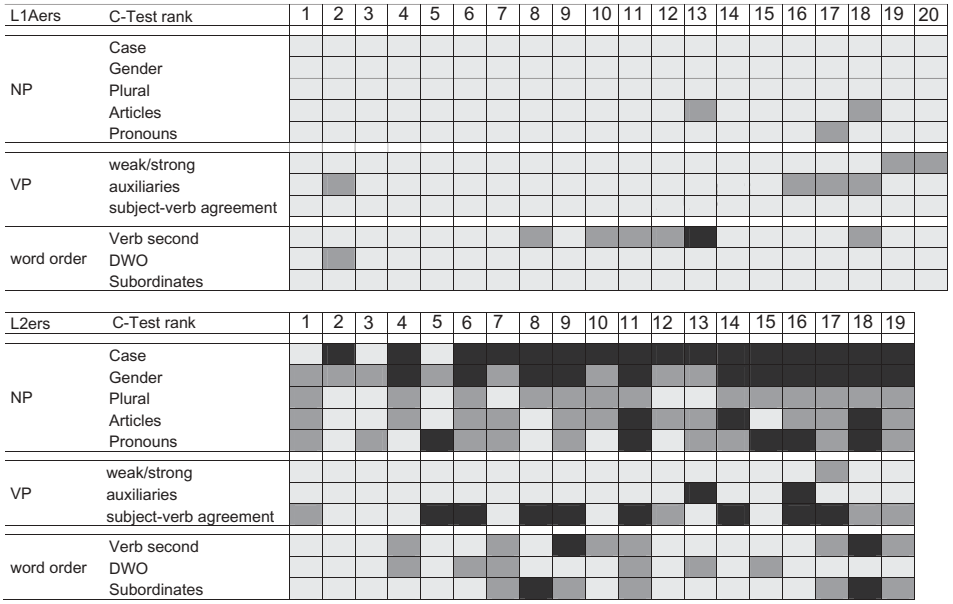

FIGURE 3 Speakers within the accuracy range of the reference population in both speech samples (light shading), in one speech sample (intermediate shading) or in none of the speech samples (dark shading).

advanced late L2 learners of German to late bilinguals who acquired German as their (only) native language in childhood but later moved to an L2 environment and experienced L1 attrition. The rationale for this approach is that in a comparison with a monolingual reference group, differences between natives and L2ers may be ascribed to either a bilingualism or a maturational deficit effect. A comparison of late and early bilinguals, as suggested by Singleton (2013), might confuse the picture due to the larger degrees of variability in ultimate attainment that are almost invariably present in such populations.

We therefore compared 20 late L2 learners and 20 late L1 attriters of German and contrasted their accuracy on 11 features of German morphology and word order in two spontaneous speech samples (one narrative and one descriptive task). The native language of the L2 learners and the second language of the attriters was English. The L2ers had been immersed in a German-speaking environment for more than 26 years on average, while the attriters had an average length of residence of 37 years in an English-speaking setting. The average length of time that both populations had lived in Germany was 26 years. The two populations were matched for general proficiency levels on three formal tasks (C-Test, verbal-fluency task, and a detailed self-assessment).

An investigation of language behavior in everyday, informal situations revealed similar levels of the use of German with friends and family between the two bilingual populations. In addition, of course, the L2ers were immersed in a German-speaking environment and thus surrounded by targetlike written and spoken German, while the L1Aers did not receive this kind of input. Some of them used their first language only extremely rarely, while others had a strong German-speaking social network. For example, one speaker who migrated to Canada at age 22, 41 years

prior to his taking part in the present study, stated that he used his L1 at most two or three times a year when he might exchange a few short sentences (e.g., in chance encounters with tourists), and that he had never returned to Germany since his migration. On the other hand, several of the participants had come to Canada in the 1950s through the German Catholic network *Kolping*, lived in close proximity to each other, regularly attended a German-speaking church together, were all married to other German speakers from the same circle, and saw each other very frequently. Five speakers from this network participated in the investigation, and all report that they use German on a daily basis.

With the current investigation comprising attriters after more than four decades of residence in an L2 environment with both extremely high and extremely low levels of L1 use as well as intermediate levels, it can be assumed that for at least some of them the level of L2-to-L1 interference must be as high as it can reasonably get for any late bilingual. On the other hand, some of the late immersed L2ers in the present study reported that they used German very frequently in all contexts of life and thus presumably minimized the impact of their L1. It was therefore assumed that the degree of cross-linguistic influence that the two bilingual populations experienced should be similar—if not at the group level, there should at least be some individuals among the L1Aers who experience at least as much, and very likely considerably more, traffic from English to German than those L2ers who have the least amount of such traffic.

The analysis thus combined two approaches: a statistical group comparison and a comparison of the performance of individual speakers against the range of a baseline population in two spontaneous speech samples from each speaker (an autobiographical interview and a film retelling).

The group comparison presented in the first part of the study revealed a relatively straightforward picture. The L1Aers did not differ from the monolingual reference population in their accuracy on any of the 11 morphological and syntactic variables that were assessed in free speech. The L2ers achieved quite high levels of accuracy on VP morphology (with the exception of subject-verb agreement, where they differed consistently from both native German populations) and were also relatively successful at observing German word-order rules. For NP morphology, on the other hand, there were a number of differences between the late L2ers and the L1Aers. These were particularly noticeable in the areas of case and gender, and in the related phenomenon of the undersuppliance of determiners and pronouns, which are obligatorily marked for these two features.

Research Question 1 asked whether there were any morphological or word-order features where both bilingual populations differed from the reference group but not from each other. While such features were not found in the present study, it should be noted that population differences in overall accuracy were found in an earlier study in which the entire populations of 53 L1Aers and monolingual controls were investigated (Schmid & Dusseldorp 2010). The failure of the differences between L1Aers and the RP to reach significance in the present study may therefore be linked to the limited sample size and thus be a Type II error.

With respect to Research Question 2, the present study has shown that L2 speakers differ markedly from both bilingual and monolingual natives in those areas of morphology in which their L2 is more complex than their L1, i.e., morphological contrasts not encoded, or encoded in a more restricted manner, by their native language (in the present case, NP as opposed to VP morphology). This finding suggests that competition effects between a bilingual's L1 and L2 may impact differentially on the performance of attriters and late learners, as suggested in Schmid

(2009): For L2 learners, high levels of accuracy are most difficult to attain where features are different between their two languages, while L1 attriters struggle more with those features that are similar.

In order to obtain more insight into the question of the impact of a potential maturational constraint, the individual comparison then sought to determine for all of these features whether any one of the L2ers attained the levels of accuracy within the range of a larger population of 53 L1Aers, and similarly, how many of the L1Aers fell within the range of the same number of monolingual controls (Research Question 3). For each of the grammatical features under investigation, the highest number of errors per 1,000 words found among the reference population was used to define this range. Here, again, it was found that the attriters had largely remained within the native range, in particular for NP morphology. For case and article/pronoun omission, only three L2 speakers were within the attriting range in both their speech samples (and all of them were outside this range for at least two other features of NP morphology), and not a single one attained this level of accuracy for grammatical gender.

This finding does suggest that there may be a qualitative difference between speakers who learn a language in childhood and those who learn it later in life. This difference cannot be adequately accounted for merely on the basis of cross-linguistic interference (i.e., with the assumption of underlying targetlike knowledge that is not applied consistently due to a high cognitive load), since both the L2ers and the L1Aers tested here experience this. It is interesting that the two populations are most clearly differentiated in the area of grammatical gender, as this linguistic feature has often and controversially been investigated in the context of the debate on maturational constraints (e.g., Foucart & Frenck-Mestre 2011; Franceschina 2005; Hopp 2013; Sabourin 2003; Hawkins 2001; L. White et al. 2004). A full review of this debate is beyond the scope of this article, but the results presented here appear to be consistent with the view held by proponents of a representational deficit account claiming that late L2 learners are unable to fully incorporate gender features unless these are instantiated by their L1, and have to rely on compensatory strategies instead for gender concord processes (e.g., Franceschina 2005; Hawkins 2001).

This assumption is particularly interesting in light of studies that do suggest that German L2ers can reach (monolingual) native levels of performance on grammatical gender, but use controlled tasks to elicit their results, such as Hopp (2013). It is possible that the higher cognitive load incurred in the production of spontaneous speech may interfere with compensatory strategies but that those strategies remain available in controlled experiments such as elicited production. However, Hopp (2013) also reports on an eye-tracking experiment in which online sensitivity to gender marking is found among L2 speakers of German whose L1 does not encode gender, and it is more difficult to see how compensatory strategies such as the ones proposed by Hawkins (2001) would enable L2ers to use the gender feature in this manner. To date, there are no studies that investigate bilinguals' performance with respect to individual grammatical features both on controlled (in particular neurocognitive) tasks and in free speech.

In order to reconcile these conflicting findings, future studies should replicate previous experiments using controlled measures, such as off-line response to violations of gender concord (L. White et al. 2004) and online sensitivity to gender information (Foucart & Frenck Mestre 2011; Franceschina 2005; Hopp 2013; Sabourin 2003), and to combine these with investigations of spontaneous speech, as used in the present study. In order to allow a differential assessment of bilingualism effects versus potential maturational constraints, such studies should move beyond

traditional comparisons of L2 learners versus monolinguals and include L1 attriters. Such an investigation of grammatical gender in German and Dutch in L2 acquisition and L1 attrition is currently ongoing at the University of Groningen, and it is hoped that its findings may shed further light on this controversial issue.

ACKNOWLEDGMENTS

I am grateful to Susanne Brouwer, Cornelia Lahmann, Nienke Meulman, Bregtje Seton, and Rasmus Steinkrauss, as well as to two anonymous reviewers for their insightful comments and suggestions on a previous version of this text. All remaining shortcomings are my own.

REFERENCES

Abrahamsson, Niclas & Kenneth Hyltenstam. 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59(2). 249–306.

Bialystok, Ellen. 2001. *Bilingualism in development: Language, literacy, & cognition*. Cambridge: Cambridge University Press.

Bialystok, Ellen & Kenji Hakuta. 1994. *In other words: The science and psychology of second-language acquisition*. New York: Basic Books.

Birdsong, David. 2004. Second language acquisition and ultimate attainment. In Alan Davies & Catherine Elder (eds.), *The handbook of applied linguistics*, 82–105. Oxford: Blackwell.

Boxtel, Sonja van, Theo Bongaerts & Peter-Arno Coppen. 2005. Native-like attainment of dummy subjects in Dutch and the role of the L1. *IRAL International Review of Applied Linguistics in Language Teaching* 43(4). 355–380.

Chang, Charles B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40. 249–268.

Clahsen, Harald. 1982. *Spracherwerb in der Kindheit: Eine Untersuchung zur Entwicklung der Syntax bei Kleinkindern* [*Child language acquisition: An investigation of infant language development*]. Tübingen, Germany: Narr.

Clahsen, Harald. 1995. German plurals in adult second language development: Evidence for a dual model of inflection. In Lynn Eubank, Larry Selinker & Michael Sharwood Smith (eds.), *The current state of interlanguage*, 123–137. Amsterdam: John Benjamins.

Cook, Vivian. 1995. Multi-competence and effects of age. In David Birdsong & Zsol Lengyel (eds.), *The age factor in second language acquisition: A critical look at the critical period hypothesis*, 51–66. Clevedon, England: Multilingual Matters.

Cook, Vivian. 2002. Background to the L2 user. In Vivan Cook (ed.), *Portraits of the L2 user*, 1–28. Clevedon, England: Multilingual Matters.

DeKeyser, Robert. 2003. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22. 499–533.

DeKeyser, Robert, Iris Alfi-Shabtay & Dorit Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31(3). 413–438.

Dussias, Paola E. & Nuria Sagarra. 2007. The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition* 10(1). 101–116.

Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47–65.

Flege, James E., Grace H. Yeni-Komshian & Serena Liu. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language* 41. 78–104.

Foucart, Alice & Cheryl Frenck-Mestre. 2011. Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition* 14(3). 379–399.

Franceschina, Florencia. 2005. *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: John Benjamins.

Gollan, Tamer, Rosa I. Montoya & Grace A. Werner. 2002. Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16. 562–576.

Grosjean, François. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36. 3–15.

Grosjean, François & Bernard Py. 1991. La restructuration d'une première langue: l'intégration de variantes de contact dans la compétence de migrants bilingues [The restructuring of a first language: The integration of contact variants into the competence of bilingual migrants]. *La Linguistique* 27. 35–60.

Gürel, Ayşe & Gülsen Yılmaz. 2011. Restructuring in the L1 Turkish grammar: Effects of L2 English and L2 Dutch. *Language, Interaction and Acquisition* 2(2). 221–250.

Hakuta, Kenji, Ellen Bialystok & Edward Wiley. 2003. Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science* 14. 31–38.

Hawkins, Roger. 2001. The theoretical significance of Universal Grammar in second language acquisition. *Second Language Research* 17(4). 345–367.

Hawkins Roger & Cecilia Yuet-hung Chang. 1997. The partial availability of Universal Grammar in second language acquisition: The failed functional features hypothesis. *Second Language Research* 13(3). 187–226.

Hawkins Roger & Hajime Hattori. 2006. Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research* 22(3). 269–301.

Hawkins Roger & Ianthi Tsimpli. 2009. Explaining target- and non-target performance in the acquisition of Gender concord by L2 speakers. Paper presented at ISB7, July, Utrecht.

Hopp, Holger. 2007. *Ultimate attainment at the interfaces in second language acquisition: grammar and processing*. Groningen, The Netherlands: University of Groningen dissertation.

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

Hopp, Holger & Monika S. Schmid. 2013. Perceived foreign accent in L1 attrition and L2 acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 34(2). 361–394.

Jarvis, Scott & Aneta Pavlenko. 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.

Johnson, Jacqueline S. & Elissa L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Journal of Cognitive Psychology* 21. 60–99.

Jordens, Peter. 1992. The cognitive function of case marking in German as a native and a foreign language. In Sue M. Gass & Larry Selinker (eds.), *Language transfer in language learning*, 138–175. Amsterdam/Philadelphia: John Benjamins.

Montrul, Silvina. 2008. *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam/Philadelphia: John Benjamins.

Opitz, Conny. 2013. A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *International Journal of Bilingualism* 17(6). 701–715.

Paradis, Michel. 2004. *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Phillips, Betty S. & Lowell Bouma. 1980. The acquisition of German plurals in native children and non-native adults. *IRAL International Review of Applied Linguistics in Language Teaching* 18(1). 21–29.

Prévost, Philippe & Lydia White. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16(2). 103–133.

Rogers, Margaret. 1987. Learners' difficulties with grammatical gender in German as a foreign language. *Applied Linguistics* 8. 48–74.

Sabourin, Laura. 2003. *Grammatical gender and second language processing*. Groningen, The Netherlands: University of Groningen dissertation.

Schaner-Wolles, Chris. 1988. Plural vs. Komparativerwerb im Deutschen. Von der Diskrepanz zwischen konzeptueller und morphologischer Entwicklung [Acquisition of plurals vs. comparatives in German. On the discrepancy between conceptual and morphological development]. In Hartmut Günther (ed.). *Experimentelle Studien zur deutschen Flexionsmorphologie* [*Experimental studies of German inflectional morphology*], 155–186. Hamburg: Helmut Buske.

Schmid, Monika S. 2002. *First language attrition, use and maintenance: The case of German Jews in Anglophone countries*. Amsterdam/Philadelphia: John Benjamins.

Schmid, Monika S. 2004. First language attrition: The methodology revised. *International Journal of Bilingualism* 8(3). 239–255.

Schmid, Monika S. 2007. The role of L1 use for L1 attrition. In Barbara Köpke, Monika S. Schmid, Merel Keijzer & Susan Dostert (eds.), *Language attrition: Theoretical perspectives*, 135–153. Amsterdam/Philadelphia: John Benjamins.
Schmid, Monika S. 2009. On L1 attrition and the linguistic system. *EUROSLA Yearbook* 9. 212–244.
Schmid, Monika S. 2010. Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition* 13(1). 1–7.
Schmid, Monika S. 2011a. Contact x time: External factors and variability in L1 attrition. In Monika S. Schmid & Wander Lowie (eds.), *Modeling bilingualism: From structure to chaos*, 155–176. Amsterdam/Philadelphia: John Benjamins.
Schmid, Monika S. 2011b. *Language attrition*. Cambridge: Cambridge University Press.
Schmid, Monika S. 2013a. First language attrition. *Linguistic Approaches to Bilingualism* 3(1). 97–116.
Schmid, Monika S. 2013b. First language attrition. *WIRE's Cognitive Science* 4(2). 117–123.
Schmid, Monika S. & Elise Dusseldorp. 2010. Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research* 26(1). 125–160.
Schmid Monika S. & Barbara Köpke. 2008. L1 attrition and the mental lexicon. In Aneta Pavlenko (ed.), *The Bilingual Mental Lexicon*, 209–238. Clevedon, England: Multilingual Matters.
Seton, Bregtje. 2011. *"De appel valt niet ver van het boom": On native and non-native processing of grammatical gender violations: An ERP study*. Groningen, The Netherlands: University of Groningen master's thesis.
Singleton, David. 2013. Bilingualism and age. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*. Chichester, England: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal0097
Snape, Neal, Yan-kit Ingrid Leung & Michael Sharwood Smith (eds.). 2009. *Representational deficits in SLA: Studies in honour of Roger Hawkins*. Amsterdam/Philadelphia: John Benjamins.
Sorace, Antonella. 2005. Selective optionality in language development. In Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and variation. Reconciling the biological and the social*, 55–80. Amsterdam: John Benjamins
Stevens, Gillian. 2004. Using census data to test the critical-period hypothesis for second-language acquisition. *Psychological Science* 15. 215–216.
Stolberg, Doris & Alexandra Münch. 2010. Die Muttersprache vergisst man nicht—or do you? A case study in L1 attrition and its (partial) reversal. *Bilingualism: Language and Cognition* 13(1). 19–31.
Tracy, Rosemarie. 1984. Fallstudien: Überlegungen zum Erwerb von Kasuskategorie und Kasusmarkierung [Case studies: Reflections on the acquisition of case as a category and case marking]. In Hartmut Czepluch & Hero Janssen (eds). *Syntaktische Struktur und Kasusrelationen*, 271–313.Tübingen, Germany: Narr.
White, Erin J., Fred Genesee & Karsten Steinhauer. 2012. Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult learners. *PLoS One* 7(12). e52318.
White, Lydia, Elena Valenzuela, Martyna Kozlowska-MacGregor & Yan-kit Ingrid Leung. 2004. Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics* 25. 105–133.

Submitted 26 May 2013
Final version accepted 30 September 2013