

## The impact of language co-activation on L1 and L2 speech fluency<sup>☆</sup>

Christopher Bergmann <sup>a</sup>, Simone A. Sprenger <sup>a,\*</sup>, Monika S. Schmid <sup>a,b</sup>

<sup>a</sup> Department of English Linguistics, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands

<sup>b</sup> Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom



### ARTICLE INFO

#### Article history:

Received 13 January 2015

Received in revised form 11 May 2015

Accepted 20 July 2015

Available online 21 August 2015

#### Keywords:

Language competition

Speech fluency

Bilingualism

Language attrition

### ABSTRACT

Fluent speech depends on the availability of well-established linguistic knowledge and routines for speech planning and articulation. A lack of speech fluency in late second-language (L2) learners may point to a deficiency of these representations, due to incomplete acquisition. Experiments on bilingual language processing have shown, however, that there are strong reasons to believe that multilingual speakers experience co-activation of the languages they speak. We have studied to what degree language co-activation affects fluency in the speech of bilinguals, comparing a monolingual German control group with two bilingual groups: 1) first-language (L1) attriters, who have fully acquired German before emigrating to an L2 English environment, and 2) immersed L2 learners of German (L1: English). We have analysed the temporal fluency and the incidence of disfluency markers (pauses, repetitions and self-corrections) in spontaneous film retellings. Our findings show that learners to speak more slowly than controls and attriters. Also, on each count, the speech of at least one of the bilingual groups contains more disfluency markers than the retellings of the control group. Generally speaking, both bilingual groups—learners and attriters—are equally (dis)fluent and significantly more disfluent than the monolingual speakers. Given that the L1 attriters are unaffected by incomplete acquisition, we interpret these findings as evidence for language competition during speech production.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

For most of us, speaking a second language (L2) is a challenging task. Even if we call ourselves 'fluent' in the L2, we may at times find ourselves unsure about the grammar or at a loss for the right words. In fact, we are often not so fluent at all. The most obvious reason for this lack of fluency is a lack of proficiency, that is, a speaker has insufficient knowledge of the L2 and insufficient practise in using it. A second important source of interference, however, may be the representations and procedures that we employ in our first language, which are highly accessible and rely on automatic routines that are difficult to suppress (language co-activation). The phenomenon of L2 disfluency is therefore likely to reflect an interaction of L2 proficiency and L1 transfer and competition, rather than L2 proficiency alone. Here, we will try to isolate and quantify the effect of language co-activation on fluency. To this aim, we compare the free speech performance of three groups of speakers: monolingual L1 speakers of German in Germany (controls), bilingual speakers of L1 English/L2 German in Germany (learners) and bilingual speakers of L1 German/L2 English in North America (attriters).

The underlying rationale of this comparison is that both groups of bilinguals—learners and attriters—potentially experience interference between the two languages they speak. Yet, incomplete acquisition of the L1 can be ruled out as a factor impacting on the linguistic behaviour of the L1 attriters. Instead, the disfluencies that we observe in the speech of attriters reflect transfer and competition from a co-activated L2.

#### 1.1. Fluency and disfluency

Speech fluency, as we discuss it here, refers to what Segalowitz (2010) calls "utterance fluency", that is, the ability to produce meaningful strings of linguistic symbols in a largely uninterrupted fashion (Crystal, 1997; Götz, 2013). It is understood as an automatic procedural skill (Schmidt, 1992), where automaticity implies that in proficient speakers, little attention and effort are needed to produce fluent speech. A prerequisite for fluency is that the psycholinguistic processes underlying speech planning and speech production function easily and efficiently (Lennon, 1990). Fluent speech requires input from various domains, for instance the lexicon, and is dependent on this information being rapidly accessible (Chambers, 1997). Therefore, fluency additionally requires a great deal of automaticity in other procedural components, such as lexical access, feeding into this skill (Levett, 1989).

Full fluency is mainly attributed to native speakers of a language, with some researchers asserting that all (unimpaired) native speakers

☆ This research was supported by a grant (016.104.602) of the Netherlands Organisation for Scientific Research (NWO), awarded to Monika S. Schmid.

\* Corresponding author.

E-mail addresses: [c.bergmann@rug.nl](mailto:c.bergmann@rug.nl) (C. Bergmann), [s.a.sprenger@rug.nl](mailto:s.a.sprenger@rug.nl) (S.A. Sprenger), [mschmid@essex.ac.uk](mailto:mschmid@essex.ac.uk) (M.S. Schmid).

can be described as fluent in their L1 (Hilton, 2008). However, this conceptualization of fluency should not be mistaken for the complete absence of disfluencies in typical native speech. On the contrary, all spontaneous speech is characterised by frequent occurrences of disfluencies, and these present a window into underlying planning processes, both in L1 speakers and L2 learners. As Goldman-Eisler (1968, p. 31) noted, "spontaneous production in any speaker is a highly fragmented and discontinuous activity in which hesitations act as necessary and natural speech management strategies."

Fluency is typically defined by some key concepts that are time-related on the one hand and performance-related on the other. Temporal variables of fluency are generally understood as measurements of the speech/pause relationship. These include speech rate (in syllables per time unit) as well as phonation time (speaking time vs. pause time). However, fluency should not be reduced to the speed of delivery (Chambers, 1997). It is equally important to consider performance aspects: Here, fluency is rather defined *ex negativo*, that is as stretches of speech in which disfluency markers do not exceed a certain frequency. These markers are events that interrupt the stream of words without contributing propositional content to the utterance (Tree, 1995). Typical classifications (Götz, 2013; Kormos, 2006; Maclay & Osgood, 1959; Shriberg, 1994) divide the different types of disfluency markers into

- 1) unfilled pauses (phonetically empty interruptions);
- 2) filled pauses (interruptions filled with sounds like 'uh' or 'er');
- 3) repetitions (verbatim iterations of syllables or words); and
- 4) self-corrections (alterations of the original material before an interruption).

Disfluencies, while occasionally serving 'semantic' functions, are in most instances involuntary reflections of cognitive processes without any signalling quality (Levelt, 1989). Hesitant speech can be indicative of the speaker's attentional preoccupation with macroplanning, whereas fluent passages would be contingent upon automated microplanning (Schmidt, 1992). Changes in time- or performance-related aspects of L1 fluency can often be traced back to lexical or informational access difficulties.

Even for monolingual L1 speakers, it is a challenge to bring together all that is required for fluent speech production, and the presence of disfluencies in L1 speech proves that speakers sometimes do not succeed. For L2 speakers, the challenge seems to be greater still.

Studies in which L2 speech is compared to L1 speech of the same language show that L2 speakers are considerably more disfluent. In an analysis of L1 and L2 English spontaneous production data, for example, it has been shown that L2 speakers make more self-corrections (Hieke, 1981). In L1 and L2 speakers of German and English, L2 speech in both languages was found to contain two to three times as many hesitation markers as L1 speech of the same languages (Wiese, 1984). Pausing behaviour has also been found to differ: L1 Russian speakers, for example, produced not more, but longer pauses in their L2 English than L1 speakers (Riazantseva, 2001). The data from studies that contrast L1 and L2 speakers of the same language are confirmed by intra-speaker comparisons for which the same individuals were tested in their L1 and L2. In the speech of L1 Dutch/L2 English bilinguals, for example, important quantitative differences surfaced: L2 speech contained nearly twice as many self-corrections as L1 speech. Also, interruptions in the L2 came earlier than in the L1 and repair times were longer (Van Hest, 1996). Comparing oral presentations of L1 Swedish/L2 English speakers in both languages, Hincks (2008) found shorter phrase lengths and slower speech rates in the L2, but there was a strong effect of individual speaking style across languages. Taken together, these results suggest that on different measures, speakers do not reach a native-like level of speech fluency in their L2.

## 1.2. Incomplete acquisition in late bilinguals

Why are bilingual speakers less fluent in their L2 than L1 speakers of the same language? Of the factors distinguishing L2 from L1 speakers, incomplete acquisition is probably the most obvious. It is not only in specific domains, such as fluency, that L2 learners fail to attain the same level of proficiency as L1 speakers, but across the full range of linguistic processing levels.

Most L2 speakers fall short of native standards in both comprehension and production. For example, the L2 lexicon is characterised by smaller breadth and depth (Schmitt, 2010) and speakers tend to be slower in naming pictures in their L2 (Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Ivanova & Costa, 2008). Late bilinguals' morphological and syntactic proficiency has been found to be limited both in listening and speaking (Hinkel, 2004; Lardiere, 1998; Ortega, 2003; Sabourin & Stowe, 2008; Van Boxtel, 2005; White, 2003). Additionally, late L2 learners are less accurate in understanding speech (Graham, 2003; Vandergift, 1997; Weber & Cutler, 2004), and struggle with producing unaccented speech themselves (Flege, MacKay, & Meador, 1999; Flege & Schmidt, 1995). These differences between native speakers and L2 learners are not always temporary. Rather, fossilisation occurs in many learners, which means that individuals do not reach native-like proficiency even after prolonged exposure. Also, ultimately reaching that level is more likely in some areas than in others. For example, acquisition seems to be more successful in the areas of syntax or vocabulary than in phonetics and inflectional morphology. Pragmatic proficiency can lag even further behind (Bardovi-Harlig & Dörnyei, 1998).

There is probably no one single cause why late L2 learners do not reach the native proficiency level. Rather, L1 learning is unique in many respects: First, concepts and their linguistic representation are acquired at the same time, whereas in the L2, there is a network of representations already in place (Appel, 1996; Slobin, 1993). Second, L1 learning relies on socially supporting circumstances that are beneficial for fast and successful acquisition (Zhang & Wang, 2007). Third, brain plasticity has been shown to be higher in infants (e.g., MacWhinney, Feldman, Sacco, & Valdés-Pérez, 2000; Mechelli et al., 2004). These factors suggest that there might be a qualitative difference between the L1 and the L2 (Schmid & Köpke, 2007).

What is the impact of this difference? Incomplete acquisition means that in the L2, weaker representations than in the L1 have been created. These L2 representations might also be more difficult or time-consuming to access. It has been shown that incremental processes underlying speech planning and articulation are highly sensitive to the temporal availability of input representations (Timmermans, Schriefers, Sprenger, & Dijkstra, 2012). On the basis of these findings, we hypothesise that incomplete acquisition is associated with increased disfluency. The types of disfluencies that we expect to be more frequent in learners who have not fully acquired their L2 depend on the speech monitoring processes in these speakers: If errors are caught before articulation, more (or longer) pauses and repetitions should be found, whereas errors that are only noticed after articulation would result in more self-corrections. If we assume that L2 representations are weak and difficult to access, speech monitoring might indeed be less efficient in catching errors. It can therefore be predicted that incomplete acquisition will be reflected most strongly in an increase of self-corrections in L2 learners, relative to L1 speakers.

## 1.3. Parallel language activation in late bilinguals

While it is conceivable that incomplete acquisition is a major source of L2 disfluency, one must also consider the possibility that the mere presence of an additional language in the mind of the speaker can change the underlying processing dynamics and that it probably does so for all languages involved. Empirical support for such fundamental changes comes from experimental studies on bilingual language processing. There is a growing body of evidence that bilinguals continually

access both of their languages, even if the task does not involve the other language at all. Such language co-activation has been found on virtually every level of processing, from word forms to rhetorical patterns, in comprehension and production (see e.g., Dijkstra, 2005; Kroll & Tokowicz, 2005, and Starreveld, de Groot, Rosmark, & van Hell, 2014, for a recent review).

In the present context, the evidence for language co-activation during production is of particular importance, as it may provide us with an alternative explanation for the lack of fluency in L2 speech. However, the majority of studies have focused on paradigms that involve the production of single words. It has been found, for example, that in L2 picture naming, auditory distractors that are phonologically related to the L1 translation of the picture name significantly slow down the response, indicating that the L1 picture name is highly active (Hermans, Bongaerts, de Bot, & Schreuder, 1998). In Spanish-Catalan bilinguals, shorter naming latencies were found for picture names that were Spanish-Catalan cognates, compared to non-cognate items (Costa, Sebastián-Gallés, Miozzo, & Caramazza, 1999). The effect was larger when subjects performed the task in their non-dominant language, indicating that the cross-linguistic activation effects are asymmetric, with the L1 affecting the L2 more strongly than vice versa. Cross-linguistic activation has also been found in phoneme monitoring (Colomé, 2001) and word naming (Jared & Kroll, 2001). In a phoneme decision task, mismatches between the initial sound of target words and their (irrelevant) L1 or L2 translations were shown to induce errors as well as additional processing costs, reflected in an increased late negativity in ERPs (Rodríguez-Fornells et al., 2005). Delayed cross-language priming was found in a Chinese-English rhyming task when the task was performed in the L2, but not in the L1 (Thierry & Wu, 2007), leading the authors to conclude that L1 production might be less strongly affected by competition from the other language. Taken together, these results indicate that in bilinguals, lexical representations of the non-target language are active during production, up to the level of the word form, even if this is not task-relevant. While most studies report bidirectional effects, it seems that the strength of the effect on the target language is modulated by the strength of the non-target representations. Appropriate language cues are another modulating factor that can make it easier for bilinguals to reduce the impact of parallel activation and restrict their selection to the target language (Miller & Kroll, 2002).

One might wonder in how far the results observed in single-word production studies can be generalized to real-world language production. In a recent study, Starreveld et al. (2014) have broadened the paradigm, studying the effects of sentence context, sentence constraint and target language on language co-activation in production. Testing L1 Dutch/L2 English bilinguals, they found reliable, but asymmetric cognate effects for picture naming in isolation: When the targets were named in the L1, the effect was smaller than when they were named in the L2. When the pictures were named in sentence context, the cognate effect varied both with sentence constraint and target language. In the L1, only responses in low-constraint sentences showed a cognate effect. In the L2, by contrast, the size of the cognate effect was unaffected by sentence predictability. The authors conclude that the bilingual language-processing system operates in a fundamentally language-non-selective way, but that the language of the sentence context and the strength with which the representations of a language are connected to each other affect the activation in a target language system. In other words, with sufficient context, speakers can suppress language co-activation when speaking their L1, but not when speaking their (weaker) L2.

The picture naming paradigm used in Starreveld et al. (2014) provides an interesting solution for the context problem, but it still excludes many of the processes involved in free speech production. For example, it is important to note that language co-activation might also affect the procedural knowledge involved in sentence formulation, up to the encoding of the corresponding gestures. Evidence for such effects comes for example from Hartsuiker, Pickering, and Veltkamp (2004) who found evidence for cross-linguistic syntactic priming and

conclude that syntactic representations are integrated between languages. An elegant study by Brown and Gullberg (2008) has demonstrated the bidirectional nature of cross-linguistic activation of rhetorical patterns using a speech-gesture paradigm. Studying the way in which monolingual and bilingual speakers of English and Japanese encode manner of motion in a narrative retelling, they found that the Japanese learners encode manner in speech in a way that is consistent with monolingual Japanese speakers, but that their corresponding gestures match those of monolingual English speakers, even while speaking Japanese. This means that the effects of one language on another are not only bidirectional, but that they also clearly exceed the boundaries of the mental lexicon.

In sum, the available experimental evidence indicates that bilingual speakers are constantly facing co-activation of the non-target language. This effect is best attested for the influence of an early-learned L1 on a late-learned L2, but an effect of the L2 on the L1 is clearly evident as well. In line with these findings, the notion that languages can be more or less active, on various levels of processing, also plays a central role in the language mode hypothesis (Grosjean, 1982, 1998) as well as in state-of-the art models of bilingual processing (e.g., the bilingual interactive activation (BIA) model; Dijkstra & van Heuven, 1998) and acquisition (e.g., MacWhinney, 2005).

We have seen that during experimentally elicited speech production, language co-activation leads to increased competition between candidate representations. In addition, cognitive control mechanisms have to be exerted in order to control language selection (Kroll, Bobb, & Wodniecka, 2006). This is bound to slow down processing or lead to erroneous selection. Consequently, language co-activation—similar to incomplete acquisition—predicts more (or longer) pauses as well as more repetitions and self-corrections in spontaneous speech.

#### 1.4. L1 attritors as the missing link

When we hypothesise that language competition will lead to increased disfluencies in the spontaneous speech of bilinguals, two caveats have to be considered: First, the data cited above were mostly collected in carefully controlled experiments. This does not take away that, from a theoretical viewpoint, longer reaction times or additional processing costs can be expected to translate to a higher incidence of disfluencies in free speech. However, bilinguals might be able to compensate for these disadvantages, rendering the effects too small to be detectable in spontaneous production. Second, it has to be kept in mind that many late L2 learners, even if highly proficient in their L2, have not reached the level of full native competence of a typical L1 speaker. The challenge therefore lies in distinguishing the role of language competition from that of incomplete acquisition in the context of free speech performance.

We suggest that L1 attritors can serve as the missing link between monolingual L1 speakers, who have no acquisition deficits and do not experience language competition, and L2 speakers, whose linguistic behaviour might be affected by both factors. Attritors are speakers whose exposure to and use of a certain language has been strongly reduced after leaving the environment in which this language is widely used. L1 attritors typically have grown up in a monolingual context and permanently emigrated at a later point in time to a country where their native tongue is not spoken. This change of language context can eventually result in a non-pathological loss of proficiency (Schmid, 2004).

When we consider late attritors, who have emigrated after the age of 16, there is no reason to doubt that they have acquired their L1 completely and in the same way as any other monolingual native.<sup>1</sup> Yet, with their L1 no longer being the dominant language, attritors

<sup>1</sup> Late attritors must not be confused with speakers who have emigrated at an age at which L1 acquisition has not been completed yet. In these so-called 'incomplete acquirers', it is not always clear if they have reached full native L1 proficiency at any point. In post-puberty attritors, this is not a concern.

face similar language competition as L2 speakers. The reversal of language dominance can be a gateway for language competition, resulting in raised levels of L1 inhibition (Hulsen, 2000) or a decrease of L1 activation (Paradis, 2004, 2007). L1 attrition has been shown to potentially influence all aspects of a native speaker's competence, ranging from phonetics (De Leeuw, 2008) to morphosyntax (Gürel, 2008). The lexicon in particular has frequently been found to be affected by L1 attrition (Köpke & Schmid, 2004; Schmid & Jarvis, 2014), although claims that this domain is affected first or most strongly are probably unfounded (Schmid & Köpke, 2009). It is true, however, that reduced L1 use can lead to lower accessibility of specific lexical items (Olshtain & Barzilay, 1991) with factors like word frequency and phonological L1/L2 similarity playing a role in the retrieval of L1 concepts and lemmas, especially in production (Ammerlaan, 1996; Hulsen, 2000). In spontaneous speech, difficulties in retrieving L1 items can be reflected in tip-of-the-tongue states (Ecke & Hall, 2012) or less diverse and sophisticated vocabulary in spontaneous speech (Schmid & Jarvis, 2014; Yilmaz & Schmid, 2012). Hansen (2001) concluded that not actual loss, but longer retrieval times for lexical items have to be considered the first sign of language attrition. In spontaneous speech, these longer retrieval times have to be bridged, which might result in disfluencies.

The direct impact of word retrieval difficulties on speech fluency has recently been demonstrated in monolingual speakers: Hartsuiker and Notebaert (2010) asked L1 speakers of Dutch to describe networks of drawings. Based on a pre-test, two sets of drawings were used: One consisted of pictures that pre-test participants agreed on how to name; the other set was made up of pictures with low naming agreement in the pre-test (i.e. different names were given for the same picture). In the condition with low pre-test agreement, two of three types of disfluency—pauses and self-corrections—were more frequent in the network descriptions; no difference was found for repetitions.

The finding that lexical access is a major stumbling block for fluency is in line with what has been found for attriters: In a study on German emigrants to the Netherlands as well as German and Dutch emigrants to Canada, Schmid and Beers Fägersten (2010) found higher incidences of most types of disfluencies in the spontaneous speech of all attriter groups, relative to monolingual native speakers. Most interestingly, the analysis of the placement of the disfluency markers revealed that L1 attriters overuse them particularly before nouns and verbs, pointing to difficulties in lexical access. In this study, however, only monolingual and bilingual L1 speakers, but no L2 speakers were included. The same is true for the above-mentioned study by Yilmaz and Schmid (2012) on L1 attriters of Turkish in a Dutch L2 context; here, the frequency of hesitation phenomena was also established to be higher in attriters than in monolinguals.

Given that L1 attriters in these studies were fully proficient native speakers, we have to interpret these findings as evidence for the impact of language competition on spontaneous speech. This is what we also expect to find in our L1 attriters: They, we predict, will be less fluent than monolingual L1 speakers, but more fluent than late L2 learners whose speech performance might additionally be influenced by effects of incomplete acquisition.

### 1.5. Summary and research questions

In the present study, we want to quantify to what degree language competition affects the number of disfluencies in the speech of bilinguals. Our aims are twofold. First, we want to extend the experimental literature on language competition to linguistic behaviour in more natural contexts of use. Second, we want to tease apart to which extent reduced fluency can be attributed to incomplete acquisition on the one hand or to language competition on the other. It has proven difficult to disentangle these two factors when only considering the data of L2 learners. We will compare the spontaneous production of monolingual L1 speakers to two groups of bilingual speakers: 1) Highly proficient late L1 attriters of German, whose L2 (English) is strongly active, but

who do not differ from monolinguals in terms of age of acquisition of the target language (German); 2) Late L2 learners of German (with L1 English). We will analyse the incidence of three general disfluency markers: pauses, repetitions and self-corrections. We expect to find that language co-activation, resulting in language competition, has an impact on both bilingual groups, rendering them more disfluent than monolingual controls. Yet, we also expect them to differ in specific ways: L1 attriters should be less disfluent than late L2 learners. The degree of disfluency that we observe in the attriters, relative to the native monolingual speakers, will serve as an estimate of the disfluency that arises from language competition. L2 learners, by contrast, should be the least fluent of the three groups of speakers, as they potentially experience additive effects of language competition and incomplete acquisition. Consequently, a possible difference between L1 attriters and L2 learners may serve to illustrate the degree to which disfluencies in the free speech of late L2 learners are the result of incomplete acquisition.

## 2. Method

### 2.1. Participants

Spontaneous speech data of three groups of participants (each  $n = 20$ ) with different linguistic backgrounds were analysed: first, predominantly monolingual native speakers of German residing in Germany (henceforth: controls); second, native speakers of English with German as their second language residing in Germany (learners); third, native speakers of German with English as their second language residing in North America (attriters).

Detailed descriptive and inferential statistics can be found in Tables 1–3. Some differences will be discussed here.

Learners had a slightly younger age of emigration to Germany than attriters to North America, but this difference was only marginally significant. Length of residence was almost twice as long in learners than in attriters—a significant difference.

Education was coded on the following scale: 1 = lower secondary education (German 'Volksschule/Hauptschule' or equivalents); 2 = intermediate secondary education (German 'Realschule' or equivalents like British O-level); 3 = higher secondary education (German '(Fach-) Abitur' or equivalents like British A-level); 4 = university degree. Educational differences were significantly different on the group level with learners being most highly educated, attriters coming second and controls third. Controls differed significantly from learners and attriters; the difference between learners and attriters was not statistically significant.

Participants completed a German pen-and-paper cloze test, constructed by Schmid (2011), as a measure of general language proficiency. They had to fill in two texts in which parts of every other word were missing. Answers were coded on a scale from 1 (incorrect lexical stem and word class) to 9 (same word as in original text). The percentage of points reached by each participant was calculated. The group of attriters performed best on this test, followed by the learners and the control group. The attriters did significantly better than both the controls and the learners. The difference between the latter two groups did not reach significance, suggesting that the L2 learners had indeed reached native-like proficiency.

### 2.2. Materials and procedure

Each participant was tested individually. The data were collected as parts of different larger research projects in two-hour testing sessions that included a variety of tasks. All participants were tested by L1 speakers of German. The free speech task was not the first in the testing session, which gave participants some time to get into the language mode of the experimental language (Grosjean, 1982, 1998).

**Table 1**  
Participant descriptives.

		Sex (% male)	Age	Age of emigration	Length of residence	Education	German cloze test
Monolingual controls	Mean (range)	15 (39–58)	48.55 (39–58)	–	–	2.60	82.06 (70.7–88.8)
	SD	–	5.83	–	–	1.05	5.09
Bilingual L2 learners	Mean (range)	45 (25–73)	45.25 (25–73)	25.05 (20–39)	26.05 (7–53)	3.60	85.67 (66.0–99.2)
	SD	–	12.03	4.08	11.47	0.60	9.34
Bilingual L1 attriters	Mean (range)	20 (29–62)	42.60 (23–40)	27.95 (23–40)	14.40 (7–34)	3.50	95.04 (79.1–99.1)
	SD	–	9.18	5.33	7.94	0.69	4.27

Participants watched a ten-minute excerpt from the silent film 'Modern Times' (1936), starring Charlie Chaplin and Paulette Goddard. It runs from a scene showing Charlie Chaplin's failed attempt at working at a shipyard (about 33 min into the film) to a scene in which a police officer chases the characters played by Chaplin and Goddard from a lawn where they had been sitting and daydreaming. This sequence has been used in L2 research for at least twenty years and is also a common tool in L1 attrition research (Perdue, 1993; Schmid & Köpke, 2007; Schmid, 2011).

Immediately after the end of the film, participants were asked to retell in German what they had seen. The experimenter prompted participants to start their retellings by using a question such as *Was ist da passiert?* 'What happened in the film?' or *Was haben Sie gesehen?* 'What have you seen?' and refrained from interrupting participants during their retellings to avoid influencing their linguistic behaviour. If participants failed to mention important parts of the film, the experimenter reminded them of the scenes at the end of their retelling. In order to avoid priming certain elements of the omitted scenes, this was done by referring to previous or subsequent scenes and asking participants what the main characters did before or after the scene mentioned by the experimenter.

The retellings were taped using different types of digital audio recorders and microphones, all producing 16-bit WAV files with a sampling frequency of 44,100 Hz.

### 2.3. Data analysis

Most participants summarised the film without omitting major plot points, so that recordings consisted of one continuous and uninterrupted retelling that was analysed in its entirety. In a minority of cases, reminding participants of scenes they had left out led to another longer speech stretch, so these data were included in the analysis as well. However, we discarded all answers to the experimenter's questions that were shorter than three consecutive sentences. Disfluencies in such short answers, which typically did not deal with the content of the film, are more likely to be caused by incomplete recollection of some scenes than by linguistic factors under investigation here. Apart from single-word code switching and occasional literal citations of text displayed in the film, German was the only language used in the retellings.

All data were transcribed by native speakers of German following the standards of the CHAT transcription format and were processed

using the CLAN software (MacWhinney, 2014a,b). Sentence examples, adapted from the transcriptions, are given in the CHAT format.

We extracted two variables from the recordings: First, speech rate in syllables per minute was extrapolated from a random sample. Syllables were manually counted in six ten-second fragments at different points of the retelling, typically starting at 0, 30, 60 s etc. into the recording. Second, lexical diversity was assessed through two different measures: type-token ratio (TTR) and D. Type-token ratio is the result of a simple calculation: The number of distinct word forms (types) is divided by the total number of words (tokens) in the transcript. TTR was used to assess the size of the vocabulary used in the retellings. The disadvantage of TTR is its negative correlation with sample size. D avoids this unwanted sensitivity by compensating for the relatively less steep increase in the number of types in longer retellings, caused by the repetition of highly frequent, closed-class items. This is achieved by drawing random token samples of different sizes from the text, calculating the empirical TTR of these samples and fitting a theoretical curve that best aligns with the empirical TTR curve. This theoretical curve is based on a coefficient D, which, for the best-fitting curve, is called optimal D. This value was calculated using the CLAN software (MacWhinney, 2014b; McCarthy & Jarvis, 2010).

In our analysis, we distinguished three different types of hesitation phenomena: 1) pauses (filled and empty); 2) repetitions; 3) self-corrections.

Empty pauses are interruptions devoid of phonetic material; filled pauses are interruptions between words that are bridged using non-lexical material, usually written as 'uh(m)' or, in German, 'äh(m)'. In the CHAT format, filled pauses are transcribed as 'ah@fp' ('fp' for 'filled pause'), regardless of their actual phonetic realisation. Duplications of words were counted as repetitions unless it became clear from the context that the repetition was semantically motivated. The latter can be assumed in 1a, as opposed to 1b in which angled brackets enclose the repeated material and [/] codes the repetition itself. Translations into English are not always fully literal, but meant to exemplify the disfluency type in question:

#### 1a) Repetition as rhetoric device:

Nein, nein, mach das nicht.

'No, no, don't do this.'

#### 1b) Repetition as disfluency:

Und dann rief er < die > [/] < die > [/] die Polizei.

**Table 2**  
Group comparisons (Kruskal-Wallis test).

	Sex (% male)	Age	Age of emigration	Length of residence	Education	German cloze test
$\chi^2$	5.196	5.4923	–	–	12.5615	27.6888
df	2	2	–	–	2	2
p	0.074	0.064	–	–	0.002	<0.001

**Table 3**  
Pairwise comparisons (Wilcoxon rank-sum test).

		Age of emigration	Length of residence	Education	German cloze test
Controls vs. attriters?	W	–	–	102	15
	p	–	–	0.005	<0.001
Controls vs. learners?	W	–	–	91	136.5
	p	–	–	0.002	0.088
Attriters vs. learners?	W	263	70.5	187	334.5
	p	0.089	<0.001	0.692	<0.001

'And then he called < the > [/] < the > [/] the police.'

In both filled and empty pauses as well as in repetitions, the listener has too little information to determine what may have gone wrong during speech production. This is why this type of disfluency has also been termed 'covert repair' (Levett, 1983).

Self-corrections, the third major type of disfluency, are instances of interruption and modification that may, but do not have to include exact repetition of sentence material. Self-corrections are sometimes called 'retracings' or 'retractions' because the speaker decides to 'take back' one or several words he has said that run counter to his current utterance plan. As in repetitions, we counted only instances in which prosody and other phonetic cues give away that the retraction was due to failed utterance planning or similar factors and not used as a rhetoric device as in 2a below. 2b and 2c are two examples in which the speaker got tangled up in his construction and started over. The retracted words are enclosed in angle brackets and [//] marks the correction:

2a) Self-correction as rhetoric device:

Denn er wollte, ja, musste sich ihr ergeben.

'Because he wanted to, yes, even had to surrender to her.'

2b) Self-correction as disfluency:

Er fragte sie, <dass> [//] < wie > [//] ob sie das wollte.

'He asked her < that > [//] < how > [//] if she wanted that.'

2c) Da war < eine hungrige > [//] < hungriges > [//] ein hungriges Mädchen.

There was a < hungry-FEM > [//] < hungry-NEU > [//] a hungry-NEU girl-NEU.

In self-corrections, also called 'overt repairs', we can make use of the words that were said, but eventually rejected, to deduce what seems to have gone wrong.

Building on an existing taxonomy (Levett, 1983), we divided self-corrections in two categories: error and appropriateness repairs. Error repairs, on the one hand, occur when speakers produce utterances that are incompatible with the rules of the language they speak. When the error is detected, speakers will pause and try to correct themselves. Appropriateness repairs, on the other hand, are self-corrections in which the sentence material that is modified was grammatically correct, but, in the eyes of the speaker, needed specification or clarification.

Both error and appropriateness repairs were further divided in subtypes, again following Levett (1983). Error repairs were subdivided in phonetic, syntactic, lexical and morphological repairs (the latter category was not part of the original taxonomy). Appropriateness repairs were subdivided into coherence-related and lexically motivated repairs: In coherence-related repairs, the speaker slightly changes the sentence structure or selects a different tense in order to shift the semantic focus. In lexical repairs, speakers replace correct terms by equally correct, but more precise wording or supplement nouns by descriptive adjectives or adverbs.

In addition to error and appropriateness repairs, there are two categories: Difference repairs are self-corrections in which the original sentence structure is discarded altogether and replaced by a fresh attempt at the same or closely related content. This, however, did not occur very frequently. Examples of all types of self-corrections are given in Table 4. There is an additional category (residuals) for self-corrections that defy classification. In most cases, this was due to insufficient lexical material being produced before the repair, leaving the reason of the interruption unclear.

### 3. Results

The transcripts were first analysed with respect to general sample characteristics, such as the length of the retelling, lexical diversity, and speech rate. A summary of these statistics is shown in Table 5. In the

Table 4

Examples of appropriateness and error repairs in our corpus of retellings, categorised according to Levett's (1983) repair taxonomy. The English versions are not literal translations, but equivalent structures that illustrate the repair in question.

Repair type	Subtype	Example
Difference	–	und < er hat dann > [//] da kam (ei)ne Kuh vorbei "and < then he has > [//] then a cow went past"
Appropriateness	Coherence	Dann < geht er > [//] ist er zu dieser Werft gegangen.
		'Then < he goes > [//] he went to that shipyard.'
		Und < die Dam > [//] die junge Dame ist bewusstlos.
Error	Phonetic	'And < the lad > [//] the young lady is unconscious.'
		Und < Farlie chag > [//] ah@fp Charlie fragte sie, ob ...
		'And < Warlie chanted > [//] ah@fp Charlie wanted to know if ...'
Lexical	Lexical	Die < Orangenblüt > [//] ah@fp der Orangenbaum steht vor der Tür.
		'The orange < blossom > [//] ah@fp the orange tree is in front of the door.'
		Dann beschließen die beiden < zu ab > [//] abzuhauen.
Morphological	Morphological	'Then the two decide to < lam > [//] go on the lam.'
		Er sagt ihr, <sich > [//] sie soll sich hinsetzen.
		'He tells her < sit > [//] to sit down.'
Syntactic	Syntactic	

following section, all test statistics result from Kruskal–Wallis tests of group differences on the subject means, which—if found significant—are followed by three Wilcoxon rank sum tests for individual group comparisons (control group vs. attriters, control group vs. learners, attriters vs. learners). For the latter, the significance levels have been adjusted to account for the number of tests (Bonferroni correction,  $.05/3 = .017$ ).

With respect to the length of the retelling, we found that the attriters, as the most educated of the three groups, delivered about 1.5 times longer retellings than both the monolingual controls and the L2 speakers. These differences are statistically reliable ( $\chi^2 = 15.87$ ,  $p < .001$ ; controls vs. attriters,  $z = -2.576$ ,  $p = .009$ , controls vs. learners,  $z = 0.019$ ,  $p = .985$ , attriters vs. learners,  $z = 2.912$ ,  $p = .004$ ). Accordingly, we chose two measures of lexical diversity: the frequently used type-token ratio (TTR) and the more elaborate D.

With respect to the type-token ratio, the groups differed significantly from each other ( $\chi^2 = 16.01$ ,  $p < .001$ ; controls vs. attriters,  $z = -2.576$ ,  $p = .009$ , controls vs. learners,  $z = 0.019$ ,  $p = .985$ , attriters vs. learners,  $z = 2.912$ ,  $p = .004$ ), suggesting less lexically diverse speech in the attriters than in the two other groups. However, the D measure showed no significant differences between the three groups ( $\chi^2 = 1.386$ ,  $p = .5$ ). We conclude that the differences in the type-token ratios are largely due to the differences in sample sizes and that the levels of lexical diversity are comparable in all three groups.

With respect to the speech rate, the table shows that the learners were slightly slower than the two groups of native speakers. This difference was found to be reliable ( $\chi^2 = 7.22$ ,  $p < .03$ ; controls vs. attriters,  $z = 0.299$ ,  $p = .765$ , controls vs. learners,  $z = 2.912$ ,  $p = .004$ , attriters vs. learners,  $z = 2.427$ ,  $p = .015$ ).

Table 5

General sample characteristics for the three groups of speakers.

	Monolingual controls	Bilingual L1 attriters	Bilingual L2 learners
	Mean (SD)	Mean (SD)	Mean (SD)
Retelling length (tokens)	790.6 (450)	1220 (400)	753.6 (322)
Lexical diversity: TTR	0.39 (0.07)	0.31 (0.05)	0.37 (0.06)
Lexical diversity: D	75 (23)	66 (11)	64 (21)
Speech rate (syll./min.)	212 (39)	209 (33)	183 (37)

Taken together, the general speech sample characteristics indicate that both the attriters and the learners were sufficiently proficient in German to deliver a long stretch of speech that can be compared to the control group performance and to each other. Most importantly, the results show that the attriters do not underperform with respect to the monolingual controls. Apart from the longer retellings, which may partly reflect their relatively high level of education, their sample characteristics do not differ from those of the unattrited, monolingual speakers.

We now turn to an analysis of the speakers' (dis)fluency. The groups are first compared on three general measures: pauses, repetitions and self-corrections. Pauses are further subdivided into empty and filled pauses. The self-corrections will be further subdivided below. All disfluency counts have been normalised per 1000 words and, as before, all test statistics result from Kruskal–Wallis tests of group differences on the subject means, followed by Wilcoxon rank-sum tests for individual group comparisons and assuming a Bonferroni-corrected significance level of  $p = .017$ . The normalised numbers of pauses, repetitions and self-corrections per group are shown in Fig. 1. Absolute numbers and percentages per category and sub-category are shown in Table 6.

### 3.1. Pauses

On average, the controls produce the fewest pauses, the attriters the most and the learners are found in-between the two groups. The differences between the attriters and the other two groups are statistically reliable ( $\chi^2 = 22.046$ ,  $p < .001$ ; controls vs. attriters,  $z = -3.584$ ,  $p < .001$ , controls vs. learners,  $z = 1.755$ ,  $p = .079$ , attriters vs. learners,  $z = 2.464$ ,  $p = .014$ ).

When we split up the pauses into empty and filled pauses, we see that the very same pattern emerges for the empty pauses: controls produce the fewest empty pauses, attriters most, and the learners are similar to the controls. Again, the differences between the attriters and the other two groups are significant ( $\chi^2 = 27.160$ ,  $p < .001$ ; controls vs. attriters,  $z = -3.584$ ,  $p < .001$ , controls vs. learners,  $z = -1$ ,  $p = .317$ , attriters vs. learners,  $z = 3.472$ ,  $p < .001$ ). With respect to the filled pauses, we see that—descriptively—both attriters and learners make more filled pauses than the control group. However, taking into account the corrected p-value, the differences are only significant in the omnibus test and not on the level of pairwise comparisons ( $\chi^2 = 7.183$ ,  $p = .028$ ; controls vs. attriters,  $z = -2.165$ ,  $p = .03$ , controls vs. learners,  $z = -1.605$ ,  $p = .108$ , attriters vs. learners,  $z = 0.56$ ,  $p = .576$ ).

Taken together, the pattern of pausing behaviour shows that all three groups of speakers frequently pause, but that attriters are more disfluent in this domain than both the control group speakers and the learners. The effect is driven most strongly by the empty pauses.

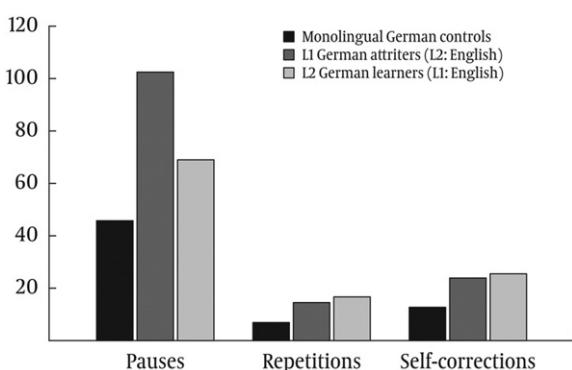


Fig. 1. Number of pauses, repetitions and self-corrections (normalised per 1000 words).

Table 6

Average counts of disfluencies per group, per category and subcategory (normalised per 1000 words). The percentages indicate a category's percentage of all disfluencies. All values above 1 have been rounded to the nearest full number.

	Monolingual controls		Bilingual L1 attriters		Bilingual L2 learners	
	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%
Pauses	46 (27)	70	103 (37)	73	69 (60)	63
Empty	10 (15)		45 (18)		12 (18)	
Filled	36 (29)		57 (27)		57 (52)	
Repetitions	7 (6)	11	15 (11)	10	17 (13)	15
Self-corrections	13 (4)	20	23 (13)	16	24 (10)	22
Difference repairs	1 (1)		3 (2)		4 (3)	
Appropriateness repairs	5 (3)		9 (5)		7 (3)	
Lexical	2 (2)		4 (2)		5 (3)	
Coherence	2 (2)		4 (3)		5 (3)	
Ambiguous	–		1		–	
Error repairs	6 (3)		11 (6)		13 (8)	
Phonetic	0.05 (0.89)		0.75 (0.80)		0.80 (0.95)	
Lexical	3 (2)		3 (2)		5 (4)	
Morphological	2 (2)		2 (3)		3 (3)	
Syntactic	2 (1)		5 (4)		4 (4)	
Residuals	0.03 (1)		0.75 (1)		1.3 (1)	

### 3.2. Repetitions

Descriptively, the control group makes the fewest repetitions. The learners make more than twice as many repetitions and the attriters follow closely. The Kruskal–Wallis test indicates the presence of significant group differences, which shows in the paired comparison of the control group and the learners, but does not reach significance for the attriters ( $\chi^2 = 9.977$ ,  $p = .007$ ; controls vs. attriters,  $z = -2.053$ ,  $p = .040$ ; controls vs. learners,  $z = 2.576$ ,  $p = .010$ ; learners vs. attriters,  $z = 0.075$ ,  $p = .941$ ). Yet, as the tests also do not reveal a significant difference between attriters and learners, we feel that it is safe to conclude that the attriters' repetition behaviour lies in-between that of the control group and the learners.

### 3.3. Self-corrections

With respect to self-corrections, the groups are more distinct: The control group makes fewer self-corrections than both the attriters and the learners, who do not differ from each other. The tests show that the differences are reliable ( $\chi^2 = 14.449$ ,  $p < .001$ ; controls vs. attriters,  $z = -3.136$ ,  $p = .002$ ; controls vs. learners,  $z = -3.080$ ,  $p = .002$ , attriters vs. learners,  $z = -0.467$ ,  $p = .641$ ).

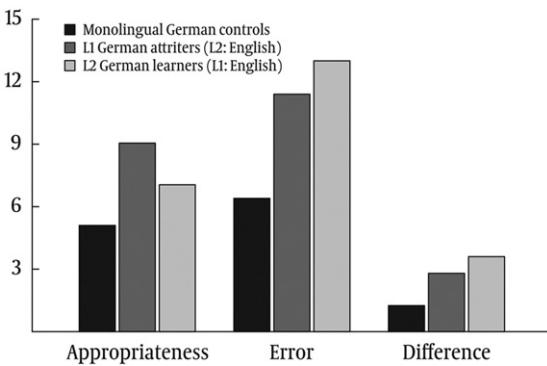
Taken together, the combined pattern of pauses, repetitions and self-corrections shows that both attriters and learners are more disfluent than the control group. A reliable difference between the attriters and the learners only surfaces in the number of pauses, with attriters making more pauses than learners. With respect to the number of repetitions and self-corrections, the attriters are indistinguishable from the learners.

To learn more about the levels of processing at which these disfluencies arise, we analysed the category of self-corrections in more detail, largely following Levelt's (1983) repair taxonomy (see Table 3 for examples in various repair categories).

Fig. 2 shows the average numbers of self-corrections per category and group.

#### 3.3.1. Difference repairs

Overall, speakers made very few difference repairs, which is in line with Levelt's (1983) observation for his corpus of spontaneous self-corrections in monolingual native speakers. The difference between the control group and the learners is reliable, but that between the control group and the attriters does not reach the corrected significance level of .017 and neither does the difference between attriters and



**Fig. 2.** Subdivision of self-corrections in appropriateness, error and difference repairs according to Levelt's (1983) taxonomy (normalised per 1000 words).

learners ( $\chi^2 = 8.476$ ,  $p = .014$ ; controls vs. attriters,  $z = 2.274$ ,  $p = .022$ ; controls vs. learners,  $z = -2.895$ ,  $p = .004$ ; attriters vs. learners,  $z = -0.731$ ,  $p = .465$ ). As in the case of the repetitions, we conclude that, with respect to the number of difference repairs, the attriters find themselves in-between the controls and the learners.

### 3.3.2. Appropriateness repairs

First, we collapsed across the three sub-categories of appropriateness repairs (level of terminology repairs, coherence repairs, and ambiguous repairs). The attriters made the largest number of appropriateness repairs and the control group the fewest, with the learners this time ranking in between. Yet, while the omnibus test for these differences is significant, the paired comparisons do not reach the corrected significance level of .017 ( $\chi^2 = 7.913$ ,  $p = .019$ ; controls vs. attriters,  $z = -2.341$ ,  $p = .019$ ; controls vs. learners,  $z = -2.174$ ,  $p = .030$ ; attriters vs. learners,  $z = 1.047$ ,  $p = .295$ ).

Next, we analysed each sub-category of appropriateness repairs separately, to see whether there are group differences for specific categories that do not surface in a global analysis. We find that, with respect to the level of terminology, control group speakers make the fewest repairs, followed by attriters and learners. However, these differences are not reliable ( $\chi^2 = 4.534$ ,  $p = .104$ ). In contrast, with respect to coherence repairs, we see a highly similar but reliable pattern of averages ( $\chi^2 = 8.010$ ,  $p = .018$ ). The paired comparisons show that the control group significantly differs from the learners (controls vs. attriters,  $z = -1.537$ ,  $p = .124$ ; controls vs. learners,  $z = -2.717$ ,  $p = .007$ ; attriters vs. learners,  $z = -0.694$ ,  $p = .488$ ), with the attriters again being placed in-between the two groups. Finally, the only group that produces the occasional ambiguous appropriateness repair (one on average) are the attriters. Taken together, the results for the appropriateness errors show that this type of repair is relatively rare and that the three groups of speakers are similar in this respect. Yet, it is also apparent from the data that the control group always makes the fewest repairs, usually followed by the attriters and the learners.

### 3.3.3. Error repairs

A global analysis of error repairs shows the highest incidence in learners and the lowest in controls; attriters again take an intermediate position, but are more similar to the learners. The statistical tests reveal that both the attriters and the learners differ significantly from the control group, but not from each other ( $\chi^2 = 10.719$ ,  $p = .005$ ; controls vs. attriters,  $z = -3.031$ ,  $p = .002$ ; controls vs. learners,  $z = -3.158$ ,  $p = .002$ ; attriters vs. learners,  $z = -0.504$ ,  $p = .614$ ).

To see whether the group differences were driven by a particular type of error repair, we further distinguished between phonetic, lexical, morphological and syntactic error repairs. Phonetic error repairs turned out to be very rare, occurring less than once per group. Lexical error repairs are slightly more frequent, but do not differ significantly between groups ( $\chi^2 = 5.077$ ,  $p = .079$ ). Neither do the morphological

error repairs ( $\chi^2 = 4.436$ ,  $p = .109$ ). It turns out that the difference between the groups with respect to error repairs is driven by the category of syntactic repairs, which is also the most frequently occurring type of repair ( $\chi^2 = 9.149$ ,  $p = .010$ ). The results of the paired comparisons follow the same pattern as in the overall analysis, with the attriters and the learners each differing significantly from the control group, but not from each other (controls vs. attriters,  $z = -2.886$ ,  $p = .004$ ; controls vs. learners,  $z = 2.776$ ,  $p = .006$ ; attriters vs. learners,  $z = 0.356$ ,  $p = .722$ ).

### 3.4. Correlations

To test whether the disfluencies that we observed in the attriters and the learners are a function of German language proficiency, we correlated the various types of disfluencies with the scores on the German C-test. For reasons of comparison, we also correlated the scores for the attriters. The results for both groups of speakers are shown in Table 7. First, we see that, in the case of attriters, there are no reliable correlations between disfluencies and proficiency, which is probably due to the fact that their C-test scores are at ceiling. In the group of learners, who show considerably more variation on the C-test scores, we see a similar lack of correlations, with the notable exception of the number of appropriateness repairs. Here, we find a relatively strong and reliable positive correlation, indicating that the more proficient learners are, the more likely they are to produce appropriateness repairs.

## 4. Discussion

In spontaneous speech, late L2 learners tend to be less fluent than monolingual L1 speakers. This disadvantage is to be expected, especially if we interpret the lack of fluency as the consequence of the incomplete acquisition of the L2. However, there is also a large body of experimental results showing that automatic co-activation of both the L1 and the L2 can influence the behaviour of bilinguals. Especially late learners may therefore experience competition between their two languages, in addition to potentially not having fully acquired their L2. Here, we have included a group of L1 attriters in our research design, so that we could quantify the impact of language competition on bilingual spontaneous speech.

By comparing L1 attriters to L2 learners and monolingual controls, we can assess the difference between effects of incomplete language acquisition in the learners and effects of language competition on utterance fluency in bilingual speakers in general.

We have looked at the incidence of time-related and performance-related indicators of fluency in spontaneous speech. On the temporal side, we find that learners speak significantly more slowly than both controls and attriters. On the performance-related side, our analyses of the frequency of pauses, repetitions and self-corrections reveal one dominant pattern: On each count, at least one of the bilingual groups differs from the monolingual control group by showing a higher incidence of disfluencies. Attriters produce more pauses than learners and controls, mainly empty ones. The frequency of repetitions in the learners is higher than in the other two groups. Both attriters and learners produce significantly more self-corrections than controls do.

**Table 7**  
Correlations between the number of disfluencies and the results of the German C-Test for attriters and learners.

	Bilingual L1 attriters		Bilingual L2 learners	
	Spearman's $\rho$	p-Value	Spearman's $\rho$	p-Value
Pauses	-0.19	0.433	0.13	0.583
Repetitions	-0.18	0.444	0.23	0.329
Self-corrections	0.03	0.913	0.21	0.364
Difference repairs	0.12	0.626	-0.01	0.972
Appropriateness repairs	-0.09	0.707	0.50	0.024
Error repairs	0.04	0.878	0.03	0.909

In our analyses, we have proceeded to splitting up self-corrections in appropriateness repairs and error repairs. This step is important because the two types of self-corrections are likely to arise at different stages of speech planning: Appropriateness repairs are intimately linked to message generation, with speakers occasionally noticing that they have not chosen the most suitable and precise linguistic form for what they intended to say. However, this type of repair requires speakers to have the necessary repertoire of alternative lexical items or expressions at their disposal. Error repairs, by contrast, typically correct mishaps at the level of grammatical, lexical and phonological encoding, rather than at the message level. This type of repairs is contingent upon a sufficient proficiency that allows speakers, who are monitoring their speech, to catch mistakes and correct them.

With respect to appropriateness repairs, we globally see no significant differences between the groups. However, learners produce more coherence-related repairs than both other groups. In lexical error repairs, there is a similar trend, but no significant difference.

In the error repairs, we find a recurrence of the pattern that we see in the sum of all types of self-corrections: Learners (who produce most of these repairs) and attritors do not differ reliably from one another, but both bilingual groups produce significantly more error repairs than controls. A more detailed analysis shows that the significant difference between the bilingual groups and the monolingual one is mostly driven by syntactic error repairs. In all other types of error repairs, there are no reliable differences.

In summary, the L1 attritors are as disfluent as the L2 learners and the speakers in both bilingual groups are significantly more disfluent than the monolingual L1 speakers. From these observations, we conclude that the disfluencies that are typical for the speech of high proficient L2 speakers are—to a considerable extent—the result of competition between the L1 and the L2, rather than incomplete L2 acquisition.

The results of the attritors, who have acquired their L1 the same way as any monolingual native, are the result of prolonged and intensive L2 exposure (and reduced L1 use) rather than a lack of proficiency. This is supported by our analyses: Our measures of lexical diversity and subtypes of self-corrections revealed no trace of specific competence deficits in the lexicon, an area that has frequently been associated with L1 attrition (Schmid & Köpke, 2009). Also, we do not find more morphological error repairs or lexical appropriateness corrections in attritors than in monolingual controls.

The fact that attritors are more disfluent than equally proficient monolinguals emphasises that speech fluency does not depend on knowledge alone, but also crucially on the ability to make use of that knowledge in real time. In line with our expectations, our results suggest that the parallel activation of two languages in attritors results in access problems that slow down processing. Especially our finding that the attritors produce the highest number of pauses is reminiscent of the results of an experiment by Hartsuiker and Notebaert (2010). To demonstrate the nexus of lexical access difficulties and increased disfluencies, they manipulated the naming agreement of items in a network of drawings. Monolingual L1 speakers of Dutch produced more pauses (and self-corrections) when trying to name items on which inter-speaker agreement was low rather than high. In our attritors, the high number of pauses also possibly reflects underlying access problems. To show even more clearly in how far lexical access (difficulties) and speech fluency are related, it would be interesting to analyse disfluency markers with respect to the lexical frequency of immediately adjacent parts of speech. However, such an analysis is beyond the scope of present paper.

Our data of the L1 attritors point in a different direction than those of Starreveld et al. (2014), who found competition effects in the L1 that were generally weaker than in the L2 and modulated by semantic constraint. We, by contrast, see that for bilinguals, it can be difficult to speak fluently, even in the L1 and in continuous production, where sentence constraints are by definition strong. Also, our findings are not in line with the experimental results of Miller and Kroll (2002), who showed

that appropriate language cues can ease the impact of language competition. Based on the observed pattern, we suggest that in bilinguals with full native proficiency, utterance fluency is crucially determined by cognitive control, that is the ability to inhibit irrelevant targets and to selectively increase the activation of the target language.

In addition to effects of language competition, it cannot be excluded that the remarkably non-native-like behaviour of the attriter group might be affected by *language mode*. According to Grosjean (1982, 1998), mode determines where a speaker stands on the bilingual continuum with a fully monolingual mode at one end and a fully bilingual mode on the other. It has been assumed—for instance in the bilingual interactive activation (BIA) model (Dijkstra & van Heuven, 1998)—that language mode has an influence on the resting level activation of lexical items, which in turn could impact on fluency. The importance of language mode seemed to have been refuted by experimental results like those by Paulmann, Elston-Güttler, Gunter, and Kotz (2006). In their experiment, lexical priming exerted a facilitatory effect regardless of the language mode their L1 German/L2 English speakers were supposed to be in. Yet, it might still be the case that the type of prolonged L2 immersion, which the L1 attritors and L2 learners in our study experience, has much more impact on language mode than a short experimental manipulation. The attritors' speech samples were collected within a prolonged testing session that required them to interact in their L1 for more than an hour before delivering their retelling. However, testing language (L1) and ambient language (L2) were not identical. This difference in language context might have made it more challenging for attritors to perform at a native-like level in their L1.

While the results for the L1 attritors are generally in line with our expectations, the comparison of attritors and L2 learners is not. The remarkable similarity between attritors and learners that we see in our data is at odds with our prediction of additional effects of incomplete acquisition on fluency that we expected to find regardless of the long length of residence in the L2 environment and the high L2 proficiency of these speakers. Our data provide no clear evidence for incomplete acquisition, at least with respect to lexicon and syntax. First, we find no group differences with respect to lexical diversity in our analysis of the D score. Second, the learners performed as well as the controls on the German cloze test. The cloze test is, of course, an offline pen-and-paper task, which does not allow far-reaching conclusions about language processing. It does, however, indicate that there is no lack of German offline proficiency in the L2 learners and that, in the absence of time pressure, they can perform at a native-like level. This is not entirely surprising, given the high level of education in these speakers and their length of residence in the L2 environment, which amounts to several decades on average.

If competence deficits cannot explain the increased number of disfluencies in both bilingual groups, the question imposes itself if we can find traces of the effects of language co-activation that are specific to the language combination these bilinguals speak. One example would be the differences in syntactic structure that exist between German and English. For example, German has at least two notable features that English lacks: First, the presence of gender and case marking in German allows a relatively free ordering of constituents. Second, certain constraints on word order, applying in German main and subordinate clauses, lead to VSO or SOV structures, rather than the SVO typically found in English. The automatic activation of English syntactic procedures should therefore interfere with the production of fluent German, especially with respect to syntax. Indeed, we see that both English-speaking groups in our study show a higher incidence of self-corrections related to the repair of syntactic errors, relative to the control group. No such difference is found for lexical and morphological errors. This latter aspect is surprising, given that German surpasses English in terms of morphological richness (e.g., with respect to gender and case marking), which might cause co-activation that is detrimental to fluency. It could be the case that this lack of difference is caused by a high number of

errors that are left uncorrected. The advanced language proficiency of both bilingual groups makes this explanation less plausible, but we cannot ultimately refute it on the basis of our current data. It will be left to future studies to investigate the nexus of self-corrections on the one hand and uncorrected errors on the other.

Finally, two limiting factors call us to interpret our data with caution: First, we have not conducted analyses of utterance complexity that could tell us if both bilingual groups produced descriptions at the same level of linguistic sophistication with respect to, for instance, semantic detail, syntactic structure or use of fixed expressions. Our general impression of the retellings is that neither of the groups produced considerably less complex utterances or detailed description than any of the others. However, future research will be required to quantify the relationship between utterance complexity and the presence of speech disfluencies. Second, it is important to keep in mind that language co-activation is a double edged-sword: On the one hand, it has convincingly been shown that co-activation leads to competition in areas where the languages differ from one another. On the other hand, though, speakers can sometimes benefit from the fact that the languages they speak are similar (e.g. Costa, Santesteban, & Caño, 2005). In our participants, this is, for instance, true with respect to the lexicon and some aspects of morphology. German, as another West-German language, is closely related to English, encouraging the transfer of words and structures. The similarity between the languages and the possibility of positive transfer might explain why both groups of bilinguals do not produce more lexical and morphological errors than monolingual controls.

## 5. Conclusion

Summing up, we can say that particularly the high incidence of disfluencies in the L1 attritors—as speakers with a deeply entrenched L1, but also a highly active L2—highlights the role of language competition on speech production. The fact that L2 learners and L1 attritors perform similarly to one another, but differently from the monolinguals indicates that the crucial factor for ultimate attainment in spontaneous speech is cognitive control. It is of foremost importance for bilingual speakers to be able to select the relevant lexical items or morphological and syntactic structures for the language that they want to use. Thus, they must actively keep the influence of the non-target language at bay. Once language selection has successfully taken place, fluency is contingent on the presence of processing routines that make it possible to effortlessly combine different sources of information. This emphasises that fluent speech is not primarily a matter of competence. Above all, it is an automatic skill that, with extensive language use, can reach a remarkable level even in L2 learners, but that can also decay when language use is reduced, as we see in the L1 attritors.

## References

Ammerlaan, T. (1996). *"You get a bit wobbly ...": Exploring bilingual lexical retrieval processes in the context of first language attrition*. Katholieke Universiteit Nijmegen.

Appel, R. (1996). The lexicon in second language acquisition. In P. Jordens, & J. A. Lalleman (Eds.), *Investigating second language acquisition* (pp. 381–404). Berlin: Mouton de Gruyter.

Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly*, 32(2), 233–262.

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1–L2 encoding of manner in speech and gesture. *Studies in Second Language Acquisition*, 30(02), 225–251. <http://dx.doi.org/10.1017/S0272263108080327>.

Chambers, F. (1997). What do we mean by fluency? *System*, 25(4), 535–544. [http://dx.doi.org/10.1016/S0346-251X\(97\)00046-8](http://dx.doi.org/10.1016/S0346-251X(97)00046-8).

Colomé, Á. (2001). Lexical activation in bilinguals' speech production: language-specific or language-independent? *Journal of Memory and Language*, 45(4), 721–736. <http://dx.doi.org/10.1006/jmla.2001.2793>.

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94(1), 94–103. <http://dx.doi.org/10.1016/j.bandl.2004.12.002>.

Costa, A., Sebastián-Gallés, N., Miozzo, M., & Caramazza, A. (1999). The gender congruity effect: Evidence from Spanish and Catalan. *Language & Cognitive Processes*, 14(4), 381–391. <http://dx.doi.org/10.1080/016909699386275>.

Crystal, D. (1997). *The Cambridge encyclopedia of language*. Cambridge: University Press.

De Leeuw, E. (2008). *When your native language sounds foreign: A phonetic investigation into first language attrition*. (PhD dissertation) Edinburgh: Queen Margaret University.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). Oxford: Oxford University Press.

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger, & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–226). Mahwah, NJ: Lawrence Erlbaum Associates.

Ecke, P., & Hall, C. J. (2012). Tracking tip-of-the-tongue states in a multilingual speaker: Evidence of attrition or instability in lexical systems? *International Journal of Bilingualism*, 17(6), 734–751. <http://dx.doi.org/10.1177/1367006912454623>.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106(5), 2973–2987.

Flege, J. E., & Schmidt, A. M. (1995). Effects of speaking rate changes on native and non-native speech production. *Phonetica*, 52(1), 41–54.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, 33(7), 1220–1234.

Götz, S. (2013). *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins.

Graham, S. (2003). Learner strategies and advanced level listening comprehension. *Language Learning Journal*, 28(1), 64–69. <http://dx.doi.org/10.1080/09571730385200221>.

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press.

Grosjean, F. (1988). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2), 131–149.

Gürel, A. (2008). Research on first language attrition of morphosyntax in adult bilinguals. *Second Language Research*, 24(3), 431–449. <http://dx.doi.org/10.1177/0267658308093611>.

Hansen, L. (2001). Language attrition: The fate of the start. *Annual Review of Applied Linguistics*, 21, 60–73. <http://dx.doi.org/10.1017/S0267190501000046>.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, 57(3), 169–177. <http://dx.doi.org/10.1027/1618-3169/a000021>.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science*, 15(6), 409–414. <http://dx.doi.org/10.1111/j.0956-7976.2004.00693.x>.

Hermans, D., Bongaerts, T., de Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(3), 213–229. <http://dx.doi.org/10.1017/S1366728998000364>.

Hieke, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech*, 24, 147–160. <http://dx.doi.org/10.1177/002383098102400203>.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36(2), 153–166. <http://dx.doi.org/10.1080/09571730802389983>.

Hincks, R. (2008). Presenting in English or Swedish: Differences in speaking rate. *Proceedings of FONETIK 2008* (pp. 21–24). Gothenburg: Department of Linguistics, Gothenburg University.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hulsen, M. (2000). *Language loss and language processing: Three generations of Dutch migrants in New Zealand*. Katholieke Universiteit Nijmegen.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127(2), 277–288. <http://dx.doi.org/10.1016/j.actpsy.2007.06.003>.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, 44(1), 2–31. <http://dx.doi.org/10.1006/jmla.2000.2747>.

Köpke, B., & Schmid, M. S. (2004). First language attrition – The next phase. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First language attrition – Interdisciplinary perspectives on methodological issues* (pp. 1–43). Amsterdam: John Benjamins.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(2), 119–135. <http://dx.doi.org/10.1017/S1366728906002483>.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308–325). Oxford: University Press.

Lardière, D. (1998). Case and tense in the "fossilized" steady state. *Second Language Research*, 14(1), 1–26. <http://dx.doi.org/10.1191/02676589874105303>.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language learning*, 40(3), 387–417.

Levett, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104.

Levett, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19–44.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford: Oxford University Press.

MacWhinney, B. (2014a). *The CHILDES project: Tools for analyzing talk – Electronic edition. Part 1: The CHAT transcription format* (Retrieved from <http://childepsy.cmu.edu/manuals/CHAT.pdf>).

MacWhinney, B. (2014b). *The CHILDES project: Tools for analyzing talk – Electronic edition. Part 2: The CLAN programs* (Retrieved from <http://childepsy.cmu.edu/manuals/clan.pdf>).

MacWhinney, B., Feldman, H., Sacco, K., & Valdés-Pérez, R. (2000). Online measures of basic language skills in children with early focal brain lesions. *Brain & Language*, 71(3), 400–431. <http://dx.doi.org/10.1006/brln.1999.2273>.

McCarthy, P. M., & Jarvis, S. (2010). MTL, vocD, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., et al. (2004). Structural plasticity in the bilingual brain: Proficiency in a second language and age at acquisition affect grey-matter density. *Nature*, 431(7010), 757. <http://dx.doi.org/10.1038/nature03016>.

Miller, N. A., & Kroll, J. F. (2002). Stroop effects in bilingual translation. *Memory & Cognition*, 30(4), 614–628. <http://dx.doi.org/10.3758/BF03194963>.

Olshtain, E., & Barzilay, M. (1991). Lexical retrieval difficulties in adult language attrition. In H. W. Seliger, & R. M. Vago (Eds.), *First language attrition* (pp. 139–150). Cambridge: University Press.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 121–133). Amsterdam: John Benjamins.

Paulmann, S., Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2006). Is bilingual lexical access influenced by language context? *Neuroreport*, 17(7), 727–731. <http://dx.doi.org/10.1097/01.wnr.0000214400.88845fa>.

Perdue, C. (Ed.). (1993). *Adult Language Acquisition. Field Methods*, Vol. 1. Cambridge: University Press.

Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23, 497–526.

Rodríguez-Fornells, A., van der Ligt, A., Rotte, M., Britti, B., Heinze, H.-J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, 17(3), 422–433. <http://dx.doi.org/10.1162/0898929053279559>.

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

Schmid, M. S. (2004). First language attrition – The methodology revised. *International Journal of Bilingualism*, 8(3), 239–255. <http://dx.doi.org/10.1177/13670069040080030501>.

Schmid, M. S. (2011). *Language attrition*. Cambridge: University Press.

Schmid, M. S., & Beers Fägersten, K. (2010). Disfluency markers in L1 attrition. *Language Learning*, 60(4), 753–791. <http://dx.doi.org/10.1111/j.1467-9922.2010.00575.x>.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, 17(4), 729–748. <http://dx.doi.org/10.1017/S1366728913000771>.

Schmid, M. S., & Köpke, B. (2007). Bilingualism and attrition. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 1–7). Amsterdam: John Benjamins.

Schmid, M. S., & Köpke, B. (2009). L1 attrition and the mental lexicon. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol: Multilingual Matters.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357–385.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Berkeley: University of California.

Slobin, D. I. (1993). Adult language acquisition: a view from child language study. In C. Perdue (Ed.), *Adult language acquisition: cross-linguistic perspectives. The results*, Vol. II. (pp. 239–252). Cambridge: University Press.

Starreveld, P. A., de Groot, A. M. B., Rosmark, B. M. M., & van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, 17(2), 258–276. <http://dx.doi.org/10.1017/S1366728913000308>.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), 12530–12535. <http://dx.doi.org/10.1073/pnas.0609927104>.

Timmermans, M., Schriefers, H., Sprenger, S. A., & Dijkstra, T. (2012). Describing simple events: The dynamics of incremental grammatical encoding. *Journal of Cognitive Psychology*, 24(4), 441–456. <http://dx.doi.org/10.1080/20445911.2011.649254>.

Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of memory and language*, 34(6), 709–738.

Van Boxtel, S. (2005). *Can the late bird catch the worm? Ultimate attainment in L2 syntax*. Radboud Universiteit Nijmegen.

Van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg: University Press.

Vandergrift, L. (1997). The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals*, 30(3), 387–409.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1–25. [http://dx.doi.org/10.1016/S0749-596X\(03\)00105-0](http://dx.doi.org/10.1016/S0749-596X(03)00105-0).

White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, 6(2), 129–141. <http://dx.doi.org/10.1017/S1366728903001081>.

Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second Language Productions* (pp. 11–25). Tübingen: Gunter Narr Verlag.

Yılmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon*, 7(3), 249–274. <http://dx.doi.org/10.1075/ml.7.3.01yil>.

Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10(2), 147–160. <http://dx.doi.org/10.1017/S1366728907002908>.