General discussion
7.1 Research questions and hypotheses

Does long-term immersion in a second-language environment result in changes to linguistic skills in the first language? This is the overarching question that I tried to answer in this dissertation, using data from native speakers of German who have emigrated to Anglophone North America as adults. In this chapter, I will discuss the results of our studies in the light of this and other questions. Specifically, I will review the results with reference to two questions: Are different linguistic skills differentially affected by attrition? And do L1 speakers, when they are affected by attrition, become more similar to proficient L2 speakers?

On a more detailed level, I will provide a discussion of three aspects of attrition that figure prominently in the theoretical approaches that were presented in the introductory chapter. The first concerns the relationship between the L2 and the L1 of bilingual speakers. The L2 always plays some role in the attrition of the L1, but the influence of the L2 comes in different guises, ranging from direct transfer to more indirect interference. I will determine which of these manifestations is predominant in the population investigated here. Second, I will consider data patterns within the monolingual and bilingual groups: It has been suggested that attrition – or multilingualism in general – is associated with increased variability, probably due to the fact that each speaker handles the challenge of simultaneously dealing with more than one language differently. A synoptic review of the results will show if this is indeed the case in the present data. Third, I will try to identify behavioural variables that are significantly correlated with variability in bilingual speakers (irrespective of whether there is more variability in a bilingual than in a monolingual group). One of the aspects that may be interesting in this respect is the use of the L1, which we have studied on the basis of self reports from our participants.

7.2 Results of the empirical studies

The data that have been presented in the main part of this dissertation come from four linguistic areas: phonetics and native-likeness ratings; speech fluency; use of formulaic sequences; and morphosyntactic processing. The most important results of these empirical studies will be briefly summarised in the following.
7.2.1 Phonetics and native-likeness ratings (Chapter 3)
In this study, we examined the L1 speech of monolingual and bilingual speakers of German from two perspectives: native-likeness ratings and acoustical measurements. In the native-likeness ratings, we found that the way bilinguals spoke was perceived to be significantly different from monolingual speech: As a group, the attriters were rated as less native-like than monolinguals. 40% of the attriters even sounded less native-like to raters than the least native-like monolingual. Also, the ratings of the attriters were spread over a wider range than those of the monolinguals. Among the attriters, native-likeness was negatively associated with length of residence in the L2 environment and positively associated with L1 use. In the acoustical analysis, we tried to pinpoint the source of these differences: The attriters were divided into a native-like group (i.e., speakers who were rated at least as good as the least native-like monolingual) and a non-native-like group (i.e., speakers who were rated worse than the least native-like monolingual). Both attriter groups were compared to a monolingual group on the formants of three vowels and one consonant: /aː/, /ɛ/, /ɔ/ and /l/. On two sounds, /ɛ/ and /ɔ/, there were no group differences. On the other two sounds, at least one of the attriter groups differed significantly from the monolingual group. These differences went into the direction of similar sounds in the attriters’ L2 (English). However, no wider variability was evident in either of the attriter groups: On five out of eight formants, both groups of bilinguals even showed smaller standard deviations than the monolingual group. In the other cases, the variability was similar in all three groups. Significant correlations of the acoustic measurements with length of residence abroad or L1 use were few and far between. Also, native-likeness ratings and formant measurements were not found to be consistently associated with one another, which illustrates the difficulty of tracing back the changes in pronunciation that we found to discrete sources.

7.2.2 Speech fluency (Chapter 4)
How does language attrition affect speech fluency? Our investigation included two groups of bilingual speakers – L1 attriters and L2 speakers – as well as monolingual speakers. We analysed recordings of film retellings for temporal fluency (speech rate) and the incidence of disfluency markers (pauses, repetitions and self-corrections). Speech rate was found to be similar in monolinguals and attriters, whereas L2 speakers spoke significantly more slowly. There was, however, no evidence for
wider variability in either of the two bilingual groups. In the disfluency markers, two different patterns emerged: When it comes to pauses, attriters stood out as the group with the highest incidence, differing from both monolinguals and L2 speakers, who did not differ from one another. With respect to repetitions and self-corrections, the attriters descriptively fell in between the other two groups with the monolinguals having a lower and the L2 speakers having a higher number of both types of disfluency markers. Self-corrections made to rectify an error (as opposed to self-corrections for stylistic reasons) were significantly more frequent in both attriters and L2 speakers, compared to monolinguals. More variability, as indicated by higher standard deviations, was evident in both bilingual groups for most disfluency markers (with the exception of empty pauses in the attriters). Additional analyses for the attriters revealed only weak negative correlations of disfluency with L1 use (i.e., fewer disfluencies were associated with higher L1 use) and weak positive correlations with length of residence abroad (i.e., more disfluencies were associated with longer residence). None of these correlations was significant (see Table 22).

### 7.2.3 Formulaic language (Chapter 5)

Like the disfluency study, our study of the use of formulaic sequences in spontaneous speech included two groups of bilinguals – L1 attriters and L2 speakers – and monolingual speakers. The analysis, based on Granger and Paquot’s (2008) taxonomy,
Table 23: Correlations of the number of formulaic sequences with L1 use and length of residence in the L2 environment for attriters. p-values were FDR-adjusted for multiple testing.

<table>
<thead>
<tr>
<th></th>
<th>L1 use</th>
<th>L2 residence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rho</td>
<td>p</td>
</tr>
<tr>
<td>Grammatical collocations</td>
<td>0.123</td>
<td>.606</td>
</tr>
<tr>
<td>Lexical collocations</td>
<td>0.173</td>
<td>.606</td>
</tr>
<tr>
<td>Attitudinal formulae</td>
<td>0.154</td>
<td>.606</td>
</tr>
<tr>
<td>Irreversible binomials</td>
<td>-0.172</td>
<td>.606</td>
</tr>
</tbody>
</table>

revealed that speakers mainly used four types of formulaic sequences: grammatical collocations (e.g., *interessiert an* ‘interested in’), lexical collocations (*scharfe Bremsung* ‘sharp braking’), attitudinal formulae (*will mal sagen* ‘I would say’) and irreversible binomials (*schließlich und endlich* ‘eventually’). For all types, a similar pattern emerged: Monolinguals used the highest number of sequences, attriters used fewer and L2 speakers fewest. However, significant group differences were only found for lexical collocations between monolinguals and L2 speakers. Attriters did not differ markedly from other monolingual speakers in their use of formulaic language. Also, there was no evidence for a wider variability in the bilingual groups. Among the attriters, no significant correlations between the incidence of the use of formulaic sequences and L1 use or L2 residence were found in an additional analysis (see Table 23). However, increased use of lexical collocations was positively correlated with lexical diversity in all three groups. Only in the bilingual groups, the use of lexical collocations also showed significant positive correlations with two types of fluency measures (number of repetitions and number of self-corrections): The more such collocations bilingual speakers used, the more fluently they spoke. The use of grammatical collocations, by contrast, did not correlate significantly with any proficiency or fluency measures in the attriters.

7.2.4 **Morphosyntactic processing (Chapter 6)**

The event-related potential (ERP) experiment was concerned with differences in morphosyntactic processing between monolingual and bilingual L1 speakers. Two
types of structures were investigated: verb form combinations (e.g., hat geholfen ‘has helped’) and gender-marked determiner-noun combinations (e.g., der Baum ‘the-MAS tree’). Correct examples and violations of both structures were presented auditorily. In response to violations of gender agreement (e.g., das Baum ‘the-NEU tree’), monolinguals and bilinguals showed the same response, namely a significant P600 effect strongest over posterior electrodes. This effect was found irrespective of the presence of an adjective intervening between the gender-marked determiner and the noun (e.g., der alte Baum ‘the-MAS old tree’). No differences in effect size between the groups and no larger variability in the individual results of the attriters were evident. Incorrect verb form combinations (e.g., kann geholfen ‘can helped’) also elicited a significant posterior P600 effect in both groups. In attriters, however, there was an additional posterior N400 effect for this type of violation. This biphasic pattern was present in the majority of the attriters.

7.3 Overarching findings
The most basic question one can ask with respect to bilingual native speakers of a given language – here: German – is whether their linguistic performance in production and comprehension is different in comparison to monolingual native speakers. Based on the hypothesis that permanent mental activation of more than one language will have consequences for the use of all languages involved (Kroll & Bialystok, 2013), we expect to find that the linguistic behaviour of bilinguals differs from that of monolinguals.

The clearest results with respect to this question came from the analyses of speech disfluencies (Chapter 4) and the use of formulaic language (Chapter 5). The attriters were found to speak more disfluently than the monolingual speakers in almost all categories. These results are in line with earlier research on spontaneous speech: In a study on German and Dutch natives in English immersion and German natives in Dutch immersion, Schmid and Beers Fägersten (2010) found an overuse of several types of disfluency markers. An analysis of the speech of L1 speakers of Turkish in Dutch immersion produced similar results (Yilmaz & Schmid, 2012). Both in that study and in our own investigation, reduced fluency was attributed to decreased automaticity in accessing linguistic information and rapidly integrating it in the continuous speech flow.
While yielding equally clear results, our study of the use of formulaic language painted a different picture. Here, attriters as a group were indistinguishable from monolingual natives on all measures, including lexical diversity. However, a correlation between speech fluency and the use of lexical collocations was found in the attriters; this correlation was not present in the monolingual group. This suggests that the use of formulaic language makes processing resources available that can then be used for speech planning, with bilingual speakers, who were found to be more prone to disfluency, benefiting most substantially. As this is the first study of the use of formulaic language among attriters, further research is necessary to verify this interpretation.

Lexical diversity, by contrast, has previously been studied in the spontaneous speech of attriters: In L1 speakers of German in English immersion, Opitz (2011) found no differences between monolinguals and bilinguals, but Schmid (2007) did. For other language combinations, results tend to be mixed as well (see section 1.2.3 for more detail). It has been suggested that sample size is the crucial factor in explaining these conflicting findings (Schmid & Jarvis, 2014): Several analyses of small populations (25 speakers or fewer) showed descriptive group differences that did not reach significance in statistical analyses; in studies of larger samples, by contrast, group differences were found to be statistically significant (Keijzer, 2007; Schmid & Dusseldorp, 2010; Yilmaz & Schmid, 2012). As we have also focused on a small group of speakers, it is possible that a lack of statistical power has played a role, but descriptive differences between groups were also small or absent in our study. Another methodological caveat with respect to the interpretation of the null results concerns the use of the D score, which many studies, including ours, have calculated in order to quantify lexical diversity. While there is no doubt about the technical validity of the score (McCarthy & Jarvis, 2010), it may be the case that lexical diversity is too complex a phenomenon to be captured in one single figure. In-depth investigations of the present sample, as for example suggested by Schmid and Jarvis (2014), may provide more insight in future investigations.

The studies on phonetics (Chapter 3) and grammatical processing (Chapter 6) have produced mixed results: In the phonetics study, attriters were rated as sounding less native-like than monolinguals, but group differences with respect to acoustical measurements were only found for few of the formants. Also, no robust correlational pattern between the rating results and the acoustic data could be established.
There is hardly any research that is directly comparable with our study, so the question of whether weak L2 influence on L1 vowels is typical in bilingual populations has to remain largely unanswered for now. Our results on consonant production, however, tie in with the findings of an analysis of the pronunciation of German /l/ by native speakers in Anglophone immersion (de Leeuw, 2008; de Leeuw et al., 2013), confirming that this speech sound can be seen as an indicator of L1 attrition in speakers with English as their L2.

Similarly to the findings in the phonetic study, the electrophysiological investigation of morphosyntactic processing showed that the L1 of bilinguals is both robust against and susceptible to L2 influence. On the one hand, attriters were indistinguishable from monolingual speakers in their processing of violations of grammatical gender agreement, which was an unexpected finding within our study, but confirmed the results of previous research (Scherag et al., 2004). On the other hand, attriters differed from monolinguals in their processing of verb form combinations, that is, of structures that are identical in the L1 and L2. Attriters showed a biphasic N400–P600 pattern, where monolinguals only showed a P600 effect. We will return to possible sources of these differences below when considering the role of the L2 in L1 attrition.

In summary, systematic differences between the bilingual and monolingual groups were found in three out of four studies, providing evidence for clear attrition effects in the bilingual group and supporting our general hypothesis. There are, however, two important qualifications that have to be made with respect to these findings: First, not every statistically significant group difference can be seen as indicative of dramatic attrition in the bilinguals. Attrition often rather comes in subtle guises – for instance in the articulatory deviations that we detected in some speech sounds. Also, there was at least one variable in each of the studies where attriters did not differ from monolinguals. But we also saw instances of attrition in our investigations that were not so subtle: In the disfluency study, for instance, the attriters produced twice as many self-corrections and repetitions on average as the controls. Second, not every statistically significant group difference translates to a change in the linguistic surface behaviour of a bilingual that can be detected with the naked eye or ear. The differences in the processing of verb form combinations, for instance, are unlikely to be noticeable at a surface level or to have an effect on everyday language use. By contrast, the results of the native-likeness rating, in which
40% of the bilinguals did not sound native-like to monolingual listeners, suggest real-life consequences of attrition with potential ramifications for the linguistic identity of some speakers. In other areas, we can only guess at whether the effects that we found would be noticeable for other speakers. More detailed investigations into the perception of attriters’ speech by unattrited listeners are therefore desirable.

7.4 Differential susceptibility of linguistic skills to L1 attrition
Ranging from clear evidence for attrition in one area to clear evidence against attrition in other domains, the present results suggest that attrition is a multifaceted phenomenon, which is best captured in panoramic investigations like the present one. However, the results also prompt the question of whether – and why – different (sub)skills in the L1 are differentially affected by attrition. The only serious theoretical predictions on the topic of differential susceptibility to attrition have been formulated within the framework of psycholinguistic usage-based approaches, such as the Activation Threshold Hypothesis (ATH). The predictions of the ATH run along the lines of two distinctions: The first is the one between production and comprehension. It is hypothesised that the L1 is more strongly affected in production than in comprehension. Producing language requires the self-activation of mental representations and is therefore more effortful than comprehension, which is mainly dependent on external stimulation. The second distinction concerns the representation of language in different memory systems: The ATH predicts that linguistic skills that are based on procedural memory will be more resistant to attrition than skills based on declarative memory, the latter not being modality-specific and hence more susceptible to L2 interference (see section 1.3.2 for details).

When we consider the present data in the light of these two hypotheses, we find the first prediction – stronger effects on production than on comprehension – not to be supported: The results of the EEG study, as the only experiment in this dissertation focusing on comprehension, show that attriters differed from monolingual native speakers in the processing of one of two structures. By contrast, the analysis of the use of formulaic language – one of three studies based on the same production task – revealed that attriters and monolinguals as groups were indistinguishable from one another. Given these findings, there seems to be no convincing evidence for differences between comprehension and production tasks in the present dataset.
When we try to apply the second hypothesis of the ATH, concerning the involvement of different memory systems, to the present dataset, this is not without problems: At first sight, the problems primarily seem to lie in the dataset. It does not include any study investigating linguistic skills that exclusively rely on declarative memory. Also, there is only one single study focusing on linguistic skills that exclusively rely on procedural resources, namely phonetic production. At second sight, however, the theory itself seems to be as much part of the problem as the data: For three out of the four studies in this dissertation, it seems impossible to make a neat distinction between skills relying on either declarative or procedural memory systems. Formulaic sequences, for instance, as a part of the lexicon, have a strong declarative component, but their use also depends on automatic procedures for integrating them in the continuous speech stream. For disfluencies, a similar reasoning applies: On the one hand, speech fluency has been described as a skill dependent on the highly automated micro-planning of utterances (R. Schmidt, 1992), that is, on procedural memory. On the other hand, detailed analyses of disfluency markers in German attriters have shown that the incidence of such markers is increased preceding lexical items, such as nouns and verbs. This suggests that access problems in the lexicon, which is represented in declarative memory, account for a sizeable share of disfluencies (Schmid & Beers Fägersten, 2010). In other linguistic areas as well, it is difficult to draw a sharp line between skills that rely on declarative memory and skills that rely on procedural resources with many requiring an involvement of both memory systems (see also Clahsen & Felser, 2006, for a criticism of the declarative/procedural dichotomy). These observations notwithstanding, it is certainly possible to devise experiments or conduct analyses that isolate specific skills that are (mostly) reliant on one of the two memory systems. However, the way the prediction is formulated makes it difficult to apply to studies that are not specifically tailored to testing it. As we face such difficulties with the present dataset, we cannot determine whether the attrition patterns in our participant group follow the hierarchy of susceptibility postulated by the ATH.

A third question of a general nature concerns similarities between attriters and L2 speakers, that is, other bilinguals. We have discussed in how far attriters differ from monolingual speakers, with both groups sharing that they were tested in their native language. A commonality between attriters and L2 speakers, however, lies
in the fact that they both experience the permanent mental presence of another language, which is assumed to have an influence on the production and comprehension of both of a bilingual’s languages. This may be reflected in a similar performance of attriters, who were tested in their L1, and L2 speakers, who were tested in their L2. When making this comparison, it has to be kept in mind that potentially non-native-like behaviour in attriters can be attributed with some certainty to the one factor in which attriters systematically differ from monolingual natives, namely potential language interference from the L2. In L2 speakers, proficiency differences may additionally come into play.

Two of the studies in our dataset included L2 speakers alongside attriters and monolingual natives. In the disfluency analysis, we saw that the attriter group took an intermediate position between monolinguals, who used fewer disfluency markers in many categories, and L2 speakers, who used more disfluency markers. In other categories, for instance repetitions and self-corrections, both bilingual groups were found to differ significantly from the monolingual control group, but not from one another. On one count, filled and empty pauses, attriters were even more disfluent than L2 speakers. Descriptively, a similar picture emerges from the study of the use of formulaic language: In most of the analyses, attriters fell in between the two other groups, with monolinguals using more and L2 speakers using fewer formulaic sequences. However, it has to be pointed out that there were no significant differences between attriters and either of the two other groups. Learners, however, differed significantly from monolinguals.

While acknowledging the limitations of these data, we can conclude that there is evidence for similarities between L1 attriters and L2 speakers. What is more, it seems to be the case that attriters, who share some aspects of their linguistic background with both monolingual L1 speakers and bilingual L2 speakers, also share part of their linguistic behaviour with both of the other groups. Still, L2 speakers have only rarely been included in attrition studies. Given that including these speakers can contribute to putting the attriters’ results into perspective and help quantify the effects of language interference (as opposed to a lack of linguistic proficiency), studies comparing monolinguals to speakers from several bilingual groups can only be strongly encouraged.
7.5 The role of the L2 in L1 attrition

The L2 plays an instrumental role in shaping the attrition of a bilingual’s L1 in at least two ways: through direct transfer or indirect interference. The former category has been discussed most extensively within generative approaches to attrition (see section 1.3.1). While transfer is not the only attrition mechanism generative theories are concerned with, it is the one that we can most readily test with the data at hand. Indirect interference features most prominently in usage-based and dynamic frameworks (see sections 1.3.2 and 1.3.3).

Direct transfer refers to L1 speakers producing structures or applying comprehension strategies that are (more or less) identical to those encountered in the L2. This type of influence is most easily attested in the use of L2 lexical items in L1 speech, but can also be found in morphology or syntax (Sharwood Smith, 1989). Indirect influence refers to non-native-like behaviour in the L1 cannot directly be ascribed to any structures or strategies of the L2, but that in all likelihood can only have been caused by the mental presence of the L2 (Paradis, 1993). It has to be pointed out, though, that this distinction is more blurry than it seems at first sight, as influence that one would classify as indirect interference can in fact have two sources: On the one hand, it can result from direct transfer in areas that have not been investigated explicitly in the study in question or that cannot be investigated given current possibilities. In this case, putatively indirect interference would actually be covert direct transfer. ‘True’ indirect interference, on the other hand, results from the specific challenges that (late) bilingual speakers face, unlike monolinguals. Selecting and using only one of the two languages requires bilinguals to permanently inhibit the non-target language. This inhibition takes up part of the limited processing capacity of the speaker, which then cannot be used for other linguistic tasks. This lack of processing capacity may be noticeable and measurable in production and comprehension, irrespective of the combination of languages.

Within generative theories, two mechanisms for direct transfer have been discussed, both concerning the effect of L2 exposure on language-specific parameters of the L1. The possibility of marking unmarked features has been considered as the most likely option: When a feature has the default value in the L1, but is marked in the L2, there may be transfer of the marked L2 value onto the unmarked L1 value (Sharwood Smith & van Buren, 1991). Alternatively, there may be unmarking of marked features: When a feature is marked in the L1, but unmarked in the L2,
the unmarked value may be transferred onto the marked L1 value. With few exceptions (Håkansson, 1995), this possibility is deemed unlikely (Sharwood Smith, 1989). Is there evidence for one or both types of influence in the result from our attriter group? The EEG experiment is the only study for which hypotheses based on generative theories can be easily formulated.

With respect to grammatical gender, the parametrisations of the L1 and L2 differ, but the L1 is marked for this feature. Excluding the possibility of unmarking, we predicted no changes in the attriters’ L1 and therefore no non-native-like results. This is indeed what we find for violations of grammatical gender agreement: There is no evidence of transfer from the L2 to the L1. However, the absence of a difference is a null result, which is difficult to interpret. When it comes to verb forms, by contrast, there is no structural difference between the L1 and L2. Still, we find that attriters expend additional processing effort – as evidenced by an N400 effect for violations of verb form selection – that is not found in monolinguals. Given the structural identity between the L1 and L2, this is an unexpected result within a generative framework of attrition. However, we have seen that processing strategies involving biphasic patterns have been found in monolingual speakers of English and languages that are in close contact with English, such as Dutch (see section 6.3.4). In view of the fact that even identical structures (or structures based on identical parameters) need not be processed the same way across different languages, this means that there may be some kind of transfer after all.

Given the null effect for grammatical gender violations in attriters, the question arises whether other bilinguals experience linguistic interference in this domain. The (elicited) production and comprehension of gender-marked elements have been investigated in late L2 speakers, for whom the L1 may be a source of direct transfer. L2 gender acquisition is often discussed within two generative frameworks: the Failed Functional Features Hypothesis (FFFH; e.g., Franceschina, 2005; Hawkins & Chan, 1997) or Full Transfer Full Access models (FTFA; e.g., De Garavito & White, 2002; White, Valenzuela, Kozlowska-MacGregor, & Leung, 2004). While the theories differ with respect to predictions of which features of an L2 can be acquired, these differences are not immediately relevant for the current question. Suffice it to say that both the FFFH and FTFA models assume that transfer from the L1 to the L2 in this domain is basically possible.
Experimental studies suggest that the question of transfer in this domain is a layered one: A gender assignment study with late learners of Dutch included L1 speakers of German and French, which both mark grammatical gender, and English, which does not. German and French speakers outperformed English learners, suggesting that the mere presence of a feature in the L1 made a difference (Sabourin, Stowe, & de Haan, 2006). This seems to be irrespective of whether the manifestation of the feature is identical between the L1 and L2, so this type of transfer has been called ‘deep transfer’. A follow-up study showed that this is not the whole story, though: Of the German and French grammatical gender systems, only the German one is highly similar to that of Dutch. In an EEG experiment, it was shown that L1 speakers of German performed even better than L1 speakers of French and were able to process grammatical gender violations in an almost native-like way (Sabourin & Stowe, 2008). This result cannot be explained with reference to deep transfer only and suggests that surface transfer – as this type of transfer has been called – plays a role in L2 acquisition as well. Similar effects of surface transfer have been found in the L2 Dutch (Lemhöfer, Schriefers, & Hanique, 2010; Lemhöfer, Spalek, & Schriefers, 2008) or L2 French (Weber & Paris, 2004) of L1 speakers of German, respectively. Additional evidence for a lack of successful transfer in the case of incompatible gender systems comes from an EEG study of the processing of L2 Dutch by L1 speakers of Polish (Loerts, 2012). Taken together, these findings show that transfer is possible and that it can be a beneficial factor in the L2 acquisition of grammatical gender marking, but that successful surface transfer requires gender systems to be similar.

These findings also suggest that transfer plays a larger role in acquisition than in attrition. This is not entirely surprising, given that L2 speakers quickly need to establish a (provisional) gender system in their L2. In order to do so, they will draw on any resources they have available, including their L1. For learners of a language, a more or less compatible gender system has to be considered a ‘signal’ that they can make use of. For attriters, by contrast, an L2 gender system is ‘noise’ that potentially interferes with their L1 gender system, the noise being heaviest when the two languages and their gender systems are highly similar. In any case, attriters will try to keep an involvement of the L2 in L1 gender production and comprehension at bay, that is, they have to inhibit the L2. Hence, indirect interference, rather than direct transfer, may be expected for language combinations where the L1 and L2 gender
systems are similar. For our attriters, however, their L2 does not play a large role, as it does not even have grammatical gender. In that case, there is no need for inhibition on the part of the attriters.

In other domains, inhibition is required in every language combination. Phonetics is one of them. In our study of acoustical differences between monolingual and bilingual native speakers, transfer from the L2 was predicted, though not in generative terms. Our hypotheses were based on the Speech Learning Model (SLM; Flege, 1995, 2002), which assumes that all speech sounds of a bilingual’s languages are contained in one common phonological space and may mutually influence one another. We identified English speech sounds that are similar, but not identical to the German sounds we investigated, predicting that the German sounds would come to resemble their English counterparts in the speech of the attriters. This is a common approach to acoustic studies in bilinguals, as testable hypotheses can be formulated that way. In our study, the predictions of the SLM were confirmed: In all cases in which attriters significantly differed from monolinguals, their formant values had changed towards those of English speech sounds, providing clear evidence of direct transfer from the L2 to the L1. These results are not only in line with theoretical predictions, but also with the results of earlier experimental research (de Leeuw, 2008; Flege, 1987; Ulbrich & Ordin, 2014). However, it has to be kept in mind that attraction of L1 speech sounds to L2 counterparts is not the only type of influence that has been found in bilinguals. There can also be repulsion, that is, sounds drifting away from one another, which can lead to a sharper differentiation between phonemes. This, however, has mainly been observed in settings where code-switching and code-mixing regularly happen (e.g., Guion, 2003).

When we think of crosslinguistic influence in broader terms than direct transfer, we can also recognise the role of interference in the results of the disfluency study. Fluent speech is a complex process, reliant on unimpeded access to declarative resources and on procedural routines that require some processing capacity. Both lexical access and processing routines may well be influenced by the mental presence of the L2. Some speakers in our attriter group indeed seemed to experience this kind of interference, which results in less fluent speech than in typical monolingual speakers. However, there may also be instances of direct transfer with respect to disfluency markers: It has been suggested that speakers of different languages
favour different ‘disfluency styles’, for instance by using empty rather than filled pauses or vice versa (Hincks, 2008; Riazantseva, 2001). Our analysis does not allow us to tease apart the sources of potential influence, so we do not know if any direct transfer – or ‘only’ indirect interference – took place. Direct comparisons of L1 and L2 speech samples of both monolingual and bilingual speakers would be a worthwhile addition to follow-up studies and could shed more light on the interaction between the languages in the production of spontaneous speech.

The only study that did not provide any evidence of L2 influence is the one on the use of formulaic language. In the coding process, we noted whether the expressions were used in a native-like way, labelling the use of L2 idioms or other sequences – verbatim or in translation – in L1 speech as ‘non-native-like’. However, the share of incorrectly used sequences was vanishingly low, so no formal analysis was conducted. In all measurements that were included in the analysis, attriters performed in a native-like manner.

Summing up, our findings suggest that both direct transfer, clearly identifiable in the acoustic analyses, and indirect interference, likely at the source of non-native-like results in the EEG and fluency studies, shape the linguistic behaviour of the attriters. But our results also show that a full understanding of the properties of the L2 is an important factor in teasing apart the two types of L2 influence on the L1. Adding to the point that we have made at the end of the previous section, we can say here that including not only L2 speakers in attrition studies, but also, where possible, monolingual native speakers of all languages involved seems crucial for gaining more profound knowledge of the relationship of these languages in the mind of a bilingual.

7.6 Increased variability in bilinguals
The discussion so far has shown that in the majority of the analyses, the attriter group differed from the monolingual control group. However, these results were all based on comparisons of group averages. Considering individual data points may afford additional insights into the attrition effects in the bilingual group. The relevance of looking beyond group averages is most strongly emphasised by dynamic theories of attrition (see section 1.3.3), which suggest that the presence of a second language creates a state of flux in a bilingual’s linguistic system. The balance
between the two languages is not fixed once and for all, but assumed to fluctuate as a function of input, use, context and other factors. Hence, the linguistic behaviour of bilinguals is predicted to fluctuate as well.

This hypothesis can be best tested in longitudinal, rather than cross-sectional research. In longitudinal data, one would see performance differences over time in the same speakers. All investigations in this dissertation, however, are cross-sectional. How can the hypotheses be applied to the present data? When we assume that bilingualism leads to a state of flux, this suggests that every bilingual speaker should be at a different point in terms of language balance at the time of testing: Some speakers will be at a point where their L1 is particularly weak, while others will be at a point where their L1 is particularly strong. When we combine these bilingual speakers into one group, we expect to find large differences between the speakers, that is, a relatively varied group performance. In a monolingual group, language balance does not play any role, so the performance within this group should be more homogeneous. Monolinguals are also expected to perform more homogeneously over time, but this will not be investigated here.

Based on the hypotheses formulated in dynamic approaches to attrition, we will consider two questions: Do we find larger standard deviations in the attriter group? And how many of the attriters fall into the range defined by the monolingual controls?

When we consider the data from our analyses of speech fluency, we find that the standard deviations in the attriter group were indeed considerably larger in each of the three major categories of disfluency markers (pauses, repetitions and self-corrections), compared to the monolingual control group. The smallest difference between monolinguals and attriters was found in the pauses; the largest was found in the self-corrections, where the standard deviation in the attriters was almost three times as large as in the monolinguals. The share of attriters whose results were in the monolingual range was 35% in the self-corrections, 75% in the repetitions and 80% in the pauses, respectively.

It has to be noted, though, that the large share of native-like attriters with respect to pauses is primarily due to an extreme outlier in the monolingual group. When we exclude this speaker, only 25% of the attriters are within the range defined by the remaining monolingual speakers. The small share of native-like attriters in the self-corrections probably has two causes: First, the monolingual group was remarkably homogeneous in this analysis, more so than in the other categories. Second,
this category includes three different counts (error corrections, stylistic corrections and content corrections). In each of these categories, 70% or more of the bilingual speakers are within the monolingual range, but these are not necessarily the same speakers in each subcategory. When combining the results from all subcategories, the non-native-likeness therefore adds up. Irrespective of this, we can conclude that the hypothesis is supported by these data: In each category, the data points of the attriters were scattered across a wider range than those of the monolinguals. Also, in each category, some attriters performed in a native-like way, but there was also always a minority – or in some isolated cases even a majority – of speakers who have not remained native-like in L2 immersion.

Similar results come from the study in which we investigated in how far the spontaneous speech of attriters and monolinguals was perceived as native-like by native raters. In these ratings, the standard deviation in the attriter group was about two times as large as in the monolingual group. 60% of the attriters had a rating within the range defined by the monolingual speakers or, conversely, 40% of the attriters did not sound native-like any more to native raters. In the phonetic domain, it would be particularly interesting and feasible to investigate speaker-internal variability as a function of context, input and similar factors. The data used in our analyses were collected after 45–65 min of monolingual L1 interaction with the participants. The results may have been different after even longer L1 interaction or no L1 interaction at all prior to testing.

The results of the detailed acoustic analysis of German speech sounds, by contrast, point in a different direction: In this study, we have analysed the first two formants of three vowels and one consonant. In five out of eight formants, the standard deviations in the attriter group were smaller than in the monolingual group. In these cases, all attriters were within the native range. In the other three cases, the standard deviations in the attriters were only slightly larger than in the controls with the majority of speakers falling into the monolingual range. The results from both the native-like attriters and non-native-like attriters (the group division was based on their results in the native-likeness ratings) are at odds with the hypotheses of dynamic theories of attrition.

We are not aware of any studies that have produced similar results, so the interpretation of the smaller standard deviations in the attriters has to remain somewhat speculative: In both monolingual and bilingual speakers, articulatory variability
arises from the fact that none of the sounds of which the formants were measured was produced in isolation. In all cases, some coarticulation – that is, influence from preceding or following sounds – is involved. However, coarticulation does not work exactly the same in all languages (Beddor, Harnsberger, & Lindemann, 2002; Magen, 1997; Strange & Bohn, 1998; West, 1999). A speaker has to be aware of the language-specific constraints to avoid producing a sound that is so strongly assimilated that it is not recognisable any more as an example of the phoneme in question. Adult monolingual speakers usually know how far they can go in assimilating a sound to its environment. Attriters, by contrast, may have become less certain about the constraints, similarly to what has been found in L2 speakers (Levy & Law, 2010; Oh, 2008). Smaller standard deviations in attriters may be seen as a ‘precautionary measure’: In order to avoid the effect of speech sounds becoming difficult to recognise as a result of excessive or non-native-like coarticulation, these speakers produce sounds that most closely resemble the prototype of the phoneme. In the absence of comparable data, we can only emphasise that further research is needed to verify this account of the results from this analysis. The validity of the interpretation, however, does not affect the observation that the data patterns found in the attriters do not agree with the predictions of the abovementioned hypotheses.

This is also true of the results from our analysis of the use of formulaic language. In this study, we have investigated the incidence of formulaic sequences in spontaneous speech, finding no significant differences between the control and the attriter group in any of the categories. As in the acoustic study, the standard deviations of the attriter group were smaller than those of the monolingual group for all four types of sequences. For each type of formulaic sequence, between 5% (attitudinal formulae) and 25% (lexical collocations) of the monolingual speakers were found to use a higher – or, very rarely, lower – number of expressions than the attriters. This demonstrates a high degree of overlap between monolingual and bilingual groups, with data patterns in the monolinguals being almost identical to those of the attriters in three out of four cases. In one aspect, however, the groups differed: The monolingual group included a few outliers in each category who used a particularly high number of formulaic sequences, thereby increasing the standard deviation. In the attriter group, no such outliers were found, that is, the group was more homogeneous than the monolingual control group. Grammatical collocations were an exception to this pattern: No outliers were found in either group.
This may be due to the fact that this type of sequence is used for syntactic rather than stylistic reasons, so there is less room for individual variability in this category.

Summing up, there were as many results that support the prediction of higher variability in the attriters as results that argue against it. Also, we saw some cases in which the majority of the attriters were within the monolingual range, but also several analyses in which a sizeable share of the attriters did not perform in a native-like way.

When applying the predictions of dynamic theories to real-life attrition datasets, one notices aspects of these hypotheses that may need to be specified or modified:

First, we have worked with the prediction that the presence of more than one language in a bilingual’s mind creates a state of flux, which may be evident from increased variability or optionality in the linguistic behaviour of these speakers. However, the domain-specific mechanisms and consequences of a balance shift towards the L2 have not been detailed yet. The higher variability that we have observed in our study on fluent speech need not have the same causes as the similar results that were found in the native-likeness ratings. When discussing the role of the L2 in L1 attrition, we have distinguished two types of potential L2 influence on the L1: One of them is direct transfer. Its effects were reflected in the L1 speech of the attriters, which monolingual listeners did not perceive to be fully native-like, possibly due to the articulation of L1 speech sounds having changed towards the L2 in some speakers. For the higher incidence of disfluency markers, by contrast, indirect interference seems to have played a more important part: Mental resources that are needed for inhibiting the L2 are assumed to be lacking for speech planning and production in the L1 (with factors such as lexical access difficulties creating additional disfluency). Through the lens of dynamic theories of attrition, the phonetics and disfluency studies basically show the same effect in bilinguals, namely higher variability, thereby supporting the theories’ predictions. Irrespective of this common ground, however, there may be a need for additional differentiation with respect to the underlying causes of the same surface effects.

Second, the predictions of dynamic theories do not seem to apply equally well to all types of data. The source of increased variability in bilinguals, it is suggested, lies in the inability of these speakers to perform a linguistic task the way monolinguals do. For the disfluency analysis, this seems a convincing explanation: The larger variability in the bilinguals can be traced back to differences between the two groups, resulting from the effects of indirect interference and other concomitants
of bilingualism. Here, the variability at least in part reflects the inability of some bilinguals to produce speech that is as fluent as that of most monolinguals. For our analysis of the use of formulaic language, however, the same predictions seem to make much less sense. The inability to use the L1 in a native-like way is not reflected in a higher, but a lower number of formulaic sequences, so we see less rather than more variability than in monolinguals. As some monolingual speakers already used few formulaic sequences in their spontaneous speech, there was not much room for additional variability at the low end of the scale either. Our data indeed show that the range of the bilingual speakers was a subset of the range of the monolingual speakers in most of the categories. In this case, this means that we have to interpret the data the other way around as usual, asking how much more variability there is in the monolinguals. Such cases, in which a non-native-like performance is reflected in less rather than more variability, should be, but have not been explicitly discussed in the context of dynamic approaches to attrition yet.

7.7 Factors associated with variability in bilinguals

In the present dataset, increased variability was present in the attriter group in several cases, presumably due to fluctuation in the individual language balance of bilingual speakers. But which are the factors that cause the language balance in a speaker to fluctuate? Or, speaking of the effect rather than the cause, why do we sometimes find native-like behaviour in attriters and sometimes not? In an attempt to offer cognitively plausible explanations of the linguistic behaviour of bilinguals, psycholinguistic usage-based theories – such as the Activation Threshold Hypothesis (ATH) – have suggested a relationship between native-likeness on the one hand and frequency or recency of language use on the other.

In the ATH, each language a bilingual speaks is seen as one ‘language system’. These language systems compete with one another during language comprehension and production. It is assumed that successfully activating mental representations that belong to one of the systems requires a certain amount of neural energy. This amount is referred to as the activation threshold (which is assumed to differ between comprehension and production; see section 7.4). The more recently and frequently a representation has been activated, the lower the threshold will be and the more easily and quickly the representation can be activated again. When L1 representations have been activated less frequently and recently in attriters, the
threshold rises. It is hypothesised that this is reflected in longer access times or, in the least favourable case, interference, that is, the unsuccessful selection of a language system or its representations. For the present dataset, we test these predictions by analysing the relationship between various linguistic variables and two proxy variables of mental activation of the L1: the share of L1 use, as reported by the speakers themselves, and the length of residence in the L2 environment. A higher share of L1 use is assumed to have a beneficial effect on native-likeness, whereas longer length of residence would be associated with less native-like results.

In the pronunciation native-likeness ratings, a significant linear relationship between the rating results and both variables was evident. The correlation was positive for L1 use, that is, speakers who reported using the L1 more frequently were perceived to sound more native-like by monolingual raters. For L2 residence, the correlation was negative, that is, speakers who had been living longer in North America were rated as less native-like. These findings are in accordance with the hypotheses of the ATH. By contrast, no significant relationship between linguistic variables and L1 use or L2 residence could be established in the acoustic analysis: Both were not found to make a relevant contribution to the statistical model that was used for the analysis. In our analyses of disfluency markers and the use of formulaic language, no significant correlations were found either. In summary, there seems to be no significant linear relationship between the linguistic behaviour of the attriters and factors like L1 use or L2 residence in most of the analyses. There are at least two possible explanations for these findings.

The most straightforward explanation would be that there is indeed no significant relationship between the factors or that there is a significant, but non-linear relationship. The ATH makes predictions about language systems in general, but none about the differential impact of reduced L1 use on linguistic subskills – except for a brief reference to a version of the highly questionable regression hypothesis. It is not entirely surprising to find that some subskills correlate less strongly with decreased L1 use, but usage-based theories do not provide detailed hypotheses. Alternatively, it has been suggested that the absence of a relationship may be related to the fact that a stable L1 system – in the terminology of dynamic theories – is a powerful attractor state, even in the presence of a highly active L2. For some subskills, a saturation point of rehearsal may have been reached (Schmid, 2007), that is, these skills have been used so frequently in the past and have become so deeply
entrenched that no recent use is needed for them to remain accessible. The only relevant factor would then be the ability to inhibit the L2.

It is equally important to keep in mind a point that has been made by proponents of dynamic theories of bilingualism: They suggest that factors influencing linguistic behaviour must not be considered in isolation because they will interact with one another and have a joint influence on linguistic behaviour. In the present case, some speakers may have been influenced by additional factors that are able to offset the effects of decreased language use, thereby leading to a non-linear relationship in the attriter group as a whole. We will discuss possible additional factors of influence below.

Even if there is indeed no significant relationship between most linguistic variables in our dataset and L1 use or L2 residence, the question arises why a significant correlation for both factors was found in the native-likeness ratings. One difference between the linguistic variables in question lies in their scope: The analyses in which we do not find a significant relationship all deal with well-delineated, specific phenomena, such as the use of disfluency markers or the articulation of particular speech sounds. The native-likeness rating, however, in which we do find a significant correlation, is based on the holistic, impressionistic judgement of naïve monolingual listeners. The ratings combine the perceived effects of attrition processes in several subskills, for instance segmental and suprasegmental phonetic phenomena (rather than just one of the two). We tried to eliminate the influence of additional, non-phonetic factors – such as vocabulary use or disfluency – by carefully selecting the speech samples used in the ratings. However, there is no direct proof that we were completely successful in doing so. The ratings may therefore reveal more about the spontaneous speech of the attriters and its perception by monolingual speakers than any other of our analyses. If at least some of the attrition processes detectable in the ratings are in phase with one another, we expect the effects of decreased activation to accumulate and to make it more likely to find a significant linear relationship with variables such as L1 use. This comprehensive quality is an advantage of holistic data. In isolated variables, it can be more difficult to identify a dominant factor of influence, which may or may not be obscured or governed by additional factors. Meanwhile, it is a disadvantage of holistic rating data that they do not enable us to pinpoint the aspects in the speech of attriters
that are perceived as non-native-like and that they reveal little about the underlying production processes.

A second possible explanation of the findings is that there is a significant relationship between linguistic variables and L1 use, but that we have failed to detect it, due to one of the variables, either the linguistic data or the use data, not having sufficient validity. We have seen that the type of linguistic data – holistic or specific – may have an influence on whether or not a significant linear correlation is found. Another example of an association between data types and results relates to the prediction of the ATH that longer access times are one of the hallmarks of the early phases of attrition (Paradis, 2007). In our studies on disfluencies and on formulaic language, we only analysed the incidence, but not the timing of the disfluency markers and the formulaic sequences. It may be the case that the attriters produced these sequences less quickly than the monolinguals and that the production speed was related to L1 use. However, this conjecture cannot be substantiated on the basis of the present data. While there is no reason to assume any systematic deficiencies in the linguistic data, we can conclude that some types of data seem less suitable or valid than others for studying the effects of L1 use.

In addition to that, we have to consider potential weaknesses of the variables we have used to quantify the level of L1 activation in the attriters. It has been pointed out that unsuitable ways of assessing and quantifying L1 use may have influenced the results of early attrition studies (Schmid, 2007). In our case, the share of L1 use, assessed in a questionnaire interview, combined self reports about three settings (home, work and elsewhere). Self-reported length of residence in the L2 environment (in years) was used as an additional proxy variable. Although self reports are the only way to learn something about the L1 use of bilinguals, one may generally question the speakers’ ability to make reliable statements about the way they use or have used language. While using more elaborate questionnaires and gradual scales rather than dichotomous factors certainly contributes to researchers getting a clearer picture of the L1 use of their participants, these methodological advances may still not be enough to close the gap between the actual and reported language use of bilingual speakers. Also, even in the unlikely case that self reports accurately capture actual language use, it seems to be the case that not all types of L1 use are equal. The relationship between linguistic performance and L1 use was found to be
strongest when language use in professional rather than home or other settings was considered (Schmid, 2007; Schmid & Dusseldorp, 2010). Unfortunately, the present data were not ideally suited to investigating whether this was the case in our participant group: A majority of the attriters did not use their L1 at work at all, so there were too many missing values for a meaningful analysis.

The special role of L1 use at work points to a third way to read the results of our analyses, relating to language mode and cognitive control. Following Grosjean (2001), bilingual language use can be classified along a continuum of language modes that differ with respect to the degree of activation of each of a bilingual’s languages: In settings where everyone speaks both languages and code-mixing or code-switching is acceptable, a bilingual mode is used. At the other extreme, Grosjean places situations in which everyone speaks only one of the two languages and code-mixing or-switching would be communicatively disruptive. Here, a monolingual mode is used. In between these two extremes, we find situations where both languages are used, with some interlocutors only speaking one of them, so code-mixing or code-switching have to be avoided. In such situations, which are often found at work or in other formal contexts, bilinguals are assumed to be in an intermediate mode. Both languages are activated to some degree, but the situation calls for language separation and, consequently, inhibition of the non-target language. Using language in an intermediate mode on a regular basis may act as an informal training of cognitive control mechanisms (alternatively called executive functions). Bilinguals who are experienced at inhibiting the influence of irrelevant language systems will do so more easily in situations in which language separation may not be strictly necessary and will therefore perform more similarly to monolingual speakers. But while being learnable and trainable to a certain extent (Hussey & Novick, 2012; Janus, Lee, Moreno, & Bialystok, 2016; Karbach, 2008), the individual degree of cognitive control – including the ability to inhibit a non-target language – is substantially reliant on genetic factors (Diamond, 2013; Friedman et al., 2008; Miyake & Friedman, 2012).

From this, we can draw two tentative conclusions: First, the partly genetic source of cognitive control mechanisms suggests that a high share of L1 use at work need not be the cause of a higher likelihood of native-like behaviour in the L1. It may just as well be the case that many speakers whose work requires frequent use of both languages – likely in an intermediate mode – already had a high degree of cognitive
control before pursuing that career and that they chose their professional environment in accordance with their innate abilities. Second, it stands to reason that the degree of cognitive control has an equally strong influence on the linguistic behaviour of bilingual speakers as L1 use. While this assumption does not directly contribute to answering the question whether there is a significant relationship between linguistic behaviour and language use, it reminds us not to overestimate the relevance of the latter.

Research on other individual variables, which may or may not be genetically determined, with respect to non-native-like behaviour or increased variability in attriters has only just begun. One of the factors that have recently been discussed is language aptitude, defined as a “generalized capacity to handle language structure” (DeKeyser, 2000, p. 506), which can be quantified using one of various tests (see Abrahamsson & Hyltenstam, 2008, for a brief comparison). Investigations so far have produced contradictory findings from pre-puberty (Bylund, Abrahamsson, & Hyltenstam, 2010) and post-puberty attriters (Bylund & Ramírez-Galan, 2014). Although late attriters were found to perform less accurately than monolinguals, no relationship between language aptitude – assessed using the LLAMA Language Aptitude Test – and performance on a grammaticality judgement test (GJT) was found. These initial findings bring us back to a point that was mentioned above: Finding a relationship between two variables crucially depends on their quality and validity. On the one hand, most language aptitude tests, such as LLAMA, have not been fully validated yet, so it is not entirely clear what exactly is being measured (Skehan, 2015). On the other hand, GJTs elicit binary offline responses to grammatical constructions – a particularly artificial task the results of which seem to be as much influenced by non-linguistic processing constraints as by actual language proficiency (McDonald, 2006). Such a task may therefore produce data that are not very well-suited to substantiating a relationship between native-likeness in attriters and language aptitude. It is also important to keep in mind that language aptitude is seen as a predisposition that is distinct from the ability to switch between languages or to inhibit non-target languages, which is rooted in cognitive control. Under the heading of bilingual advantages, this factor has been widely studied with respect to the L2 of bilingual speakers (see Costa & Sebastián-Gallés, 2014, and Kroll & Bialystok, 2013, for recent reviews), but similar investigations in attriter groups have not been published yet.