LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Factors Affecting Grammatical and Lexical Complexity of Long-Term L2 Speakers' Oral Proficiency

Cornelia Lahmann,[a,b] Rasmus Steinkrauss,[a] and Monika S. Schmid[a,c]

[a]University of Groningen, [b]University of Cologne, and [c]University of Essex

There remains considerable disagreement about which factors drive second language (L2) ultimate attainment. Age of onset (AO) appears to be a robust factor, lending support to theories of maturational constraints on L2 acquisition. The present study is an investigation of factors that influence grammatical and lexical complexity at the stage of L2 ultimate attainment. Grammatical and lexical complexity were assessed in 102 spontaneous oral interviews. Interviewees' AOs ranged from 7 to 17 years old.

Cornelia Lahmann is now at the Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache, University of Cologne. This research was financed by the Dutch national research foundation (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), Grant number 360-70-420. We are greatly indebted to all the student assistants involved in the transcription and rating process and we are grateful to Martijn Wieling for helping us with the statistics. We have also benefited from the fruitful discussions with Alex Housen and Bram Bulté. Thanks to Scott Crossley and Kristopher Kyle we were able to generate additional lexical measures. Finally, we would also like to thank the following archives for granting us the permission to use their oral history testimony data: University of Southern California Shoah Foundation Institute (Los Angeles), Werkstatt der Erinnerungen (Hamburg), Alte Synagoge (Essen), Fortunoff Video Archive for Holocaust Testimonies (Yale University Library), Tauber Holocaust Library and Education Program (San Francisco), United States Holocaust Memorial Museum (Washington, DC), and the Association of Jewish Refugees (London). The manuscript is based on data previously published in a doctoral dissertation.

This article has been awarded an Open Materials badge. All materials are publicly accessible in the IRIS digital repository at http://www.iris-database.org. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Cornelia Lahmann, University of Cologne, Mercator Institut für Sprachförderung und Deutsch als Zweitsprache, Triforum, 3rd floor, Innere Kanalstraße 15, 50823 Cologne, Germany. E-mail: cornelia.lahmann@gmx.de

DOI: 10.1111/lang.12151

Multifactorial analyses yielded consistently significant effects of gender and level of education for grammatical and lexical complexity. Additionally, native language use at work was a significant predictor for lexical complexity; conversely, AO did not emerge as a significant factor. We conclude that grammatical and lexical complexity at the stage of L2 ultimate attainment is the result of a complex interplay of variables that are general to language learning and performance rather than L2 specific.

Keywords naturalistic L2 attainment; spontaneous speech; grammatical complexity; lexical complexity

Introduction

Fifty-eight years after escaping from Nazi Germany to England as a child, ND[1] recalled, "I think within 6 months I could speak English." Fellow German-Jewish survivors—all of whom were between 7 and 17 years old at the time of emigration—recount similar experiences of how they came to acquire English as their second language (L2), such as "English came to me easily" or "as children you learn it much quicker." The advantage of learning a (second) language during childhood is a well-established fact known to the lay person and found by many scholars in the field of L2 acquisition (e.g., Granena & Long, 2013; Luk, De Sa, & Bialystok, 2011), but there is disagreement about the underlying mechanisms (see, e.g., DeKeyser & Larson-Hall, 2005).

To explore whether child-onset learners do indeed obtain higher L2 proficiency than adult-onset learners, researchers have often assessed advanced stages of L2 acquisition in naturalistic settings (Abrahamsson & Hyltenstam, 2009; Hopp & Schmid, 2013). In such studies, participants were generally long-term immigrants with varying ages of onset (AOs). The L2 attainment of linguistic aspects, such as grammar and the lexicon, has frequently been investigated by means of controlled, experimental designs. Most research has found that, despite adult-onset learners' extensive L2 exposure within naturalistic settings[2] and their often very advanced L2 knowledge, child-onset L2 learners are nevertheless more likely to score within native-speaker ranges, compared to adult-onset L2 learners. Altogether, AO appears to be a robust factor that explains L2 attainment across multiple linguistic domains.

Because age corresponds to distinct life stages, resulting in different experiences, researchers have argued against maturational constraints by pointing to a range of experiential factors (Birdsong, 2006). Among those are the amount and quality of (L2) input and bilingualism effects, that is, the simultaneous use and practice of two languages, which prevents bilinguals from behaving like monolinguals in either of the two languages (Schmid, 2014). Others have highlighted

the role of sociopsychological factors, such as motivation and attitude toward L2 acquisition (see Muñoz & Singleton, 2011, for an overview). What emerges is a complex interplay of factors, challenging purely biological explanations of L2 acquisition. In the present study, we assessed highly advanced L2 speakers with AOs ranging from childhood to adolescence on the basis of spontaneous oral productions. The speakers used the L2 almost exclusively for most of their lives. Our goal was to assess which factors affected the speakers' oral production, in particular their levels of grammatical and lexical complexity.

Theoretical Background

The present study applied the Complexity, Accuracy, and Fluency (CAF) approach, most frequently used in the assessment of beginning to advanced stages of L2 acquisition in classroom-based settings (Housen, Kuiken, & Vedder, 2012), to investigate advanced L2 proficiency. It thereby contributes to the small body of existing CAF research targeting productive L2 grammatical and lexical complexity in naturalistic settings beyond the advanced stage of L2 development (e.g., Forsberg Lundell et al., 2013; Forsberg Lundell & Lindqvist, 2012).

The majority of research that has addressed late stages of L2 development generally falls under the umbrella terms of ultimate attainment and fossilization (Birdsong, 2005, 2006; Birdsong & Molis, 2001; DeKeyser & Larson-Hall, 2005; Abrahamsson & Hyltenstam, 2008; Long, 2005; Muñoz & Singleton, 2011; Rothman, 2008). Most studies have found that variability in L2 users' linguistic abilities grows with increasing AOs, resulting in the decreased likelihood of becoming nativelike or near-native in the L2 (see Abrahamsson & Hyltenstam, 2009; Hyltenstam & Abrahamsson, 2012, for discussions of these terms). Various maturational explanations have been proposed, the critical period hypothesis (CPH) probably being the most prominent one. First formulated by Lenneberg (1967), it states that maturational constraints guide (L2) language acquisition, which after a certain age no longer takes place automatically (implicitly) on the basis of mere exposure. Therefore, a discontinuity in the declining slope of linguistic performance would be expected for learners with AOs around the end of the critical period. As discussed and empirically demonstrated by Granena and Long (2013), the end of the critical period(s) may be reached at different points in time. They found a discontinuity in the declining slope first for L2 phonology between the ages of 4 to 5, next for L2 lexis and collocations between the ages of 9 to 10, and finally for L2 morphosyntax around the age of 12.

However, a number of studies have not shown clear patterns of discontinuity in generally declining performance slopes, and others have described adult

learners with nativelike proficiency (Bialystok, 1997; Bialystok & Hakuta, 1995; Bialystok & Miller, 1999; but see White & Genesee, 1996). Therefore, sociopsychological explanations in place of maturational ones have been offered to account for differential outcomes in L2 attainment. Whereas such research has tapped into the entire spectrum of linguistic domains, we will concentrate on studies on the acquisition of L2 syntax, morphosyntax, and the lexicon. However, these domains have mostly been assessed receptively, while only a few studies investigated them in naturalistic, informal L2 use. To keep within the focus of the current study, we only review studies that have employed primarily linguistic, behavioural tests for assessing the acquisition of L2 morphosyntax and the lexicon, as opposed to reaction time experiments or brain imaging techniques.

Syntax and Morphosyntax

Studies in favor of the CPH or maturational constraints have frequently employed grammaticality judgment tests (GJTs) (DeKeyser, 2000; Johnson & Newport, 1989, 1991). Fewer studies have employed other measures tapping not only into receptive grammatical knowledge but also capturing syntax and morphosyntax at the productive level by asking participants to supply missing information, such as prepositions for prepositional verbs, grammatical gender, or morphological endings of nouns (e.g., Abrahamsson & Hyltenstam, 2009; Granena & Long, 2013). Many studies targeting the L2 attainment of syntax and morphosyntax have looked at L2 English speakers, focusing on learners from different first language (L1) backgrounds, including Hungarian (DeKeyser, 2000), Chinese and Korean (Bley-Vroman, Felix, & Ioup, 1988; Johnson & Newport, 1989, 1991) and Spanish (Birdsong & Molis, 2001), as well as mixed L1 backgrounds (Patkowski, 1980). Focusing on a L2 other than English, Granena and Long (2013) looked at L2 Spanish, and Abrahamsson and Hyltenstam (2009) and Abrahamsson (2012) at L2 Swedish. Across these studies AO ranged from birth (into a migrant family) up to adulthood. Participants' minimum length of residence (LoR) was usually around 5 years and their age at testing (AaT) ranged from 20 to 50 years old. All these studies point to significant age effects, as most adult-onset L2 learners' performance did not fall within the (monolingual) native speaker range. Abrahamsson and Hyltenstam (2009) found only a few cases of child-onset learners who actually performed to this yardstick. Overall, a vast amount of evidence supports AO effects on long-term speakers' L2 proficiency: Because older immigrants appear to be less successful in acquiring a L2, many scholars have argued in favor of maturational constraints.

In addition to the studies reviewed thus far, there is also some evidence from investigations of productive grammar (i.e., morphosyntactic characteristics of spontaneous or semi-spontaneous L2 production) that highlight the role of AO. In a series of studies, Bartning and colleagues (Bartning, Forsberg, & Hancock, 2009; Bartning, Forsberg Lundell, & Hancock, 2012; Forsberg Lundell et al., 2013) looked at L1 Swedish adult learners of L2 French with LoRs of 1–3 years (university students), 5–15 years, and 15–30 years. A comparison of these three groups with native speakers matched for AaT showed that all three L2 groups differed from their native counterparts with respect to morphosyntactic performance. Schmid (2014) also found a group of highly advanced late learners of L2 German to differ significantly from monolinguals and L1 attriters, particularly where noun phrase morphology in spontaneous speech was concerned. These findings from speech production strengthen the claim that nativelike performance no longer seems to be possible for adult-onset learners in that domain.

Evidence against strong claims for the CPH comes from a series of studies that share a number of methodological features. In these studies, L1 backgrounds of participants varied and included English, Dutch, and Russian (Hopp, 2010); Chinese (Bialystok, 1997); Korean (Flege, Yeni-Komshian, & Liu, 1999); and German and French (Bialystok, 1997; van Boxtel, Bongaerts, & Coppen, 2003; White & Genesee, 1996), in addition to other Germanic and Romance languages (White & Genesee, 1996). The L2s under investigation were again mostly English (Bialystok, 1997; Marinova-Todd, Marshall, & Snow, 2000; White & Genesee, 1996), but also German (Hopp, 2010) and Dutch (van Boxtel et al., 2003). The distribution of the participants in terms of their AOs and LoRs was comparable to those in studies arguing for maturational constraints. However, the findings pointed to the possibility of nativelike L2 grammar attainment (e.g., van Boxtel et al., 2003) and a generally continuous decline of L2 proficiency with increasing AOs. These studies have shown that the observed AO effects are confounded with factors such as L1–L2 similarity (Bialystok, 1997; Hopp, 2010), LoR effects (Bialystok, 1997; van Boxtel et al., 2003), and amount of current L2 use (Birdsong & Molis, 2001; Flege et al., 1999). Level of education was also proposed to be a viable factor in L2 ultimate attainment by Dąbrowska (2012), who reviewed studies on adult monolingual speakers' grammatical knowledge.

In sum, practically all studies report relatively robust age effects on L2 ultimate attainment. But given the fact that some studies have found nativelike performance accompanied by a continuous rather than a discontinuous decline in performance, alternative (possibly age-confounding) variables, such as L1

development and use, typological distance between the L1 and the L2, learners' motivation and attitude, should be considered. As for the data under investigation, grammatical complexity in spontaneous speech has rarely been assessed in studies of ultimate L2 attainment. To obtain a more accurate picture of factors that affect grammatical complexity at the level of advanced L2 proficiency, we therefore investigated migrant L2 learners in a naturalistic setting where age of emigration corresponded to AO.

Semantics and the Lexicon

Turning from morphosyntax to the lexicon, it appears that evidence for and against maturational constraints is limited. Studies that have clearly argued for a CPH in the lexical domain include Coppieters (1987), Abrahamsson and Hyltenstam (2009), and Granena and Long (2013); however, these investigations have focused mostly on grammatical or other formal aspects of the lexicon. For example, the approach employed by Coppieters lies at the crossroads of syntax and discourse semantics. L2 speakers of French, identified as near-native, with a mean LoR of 17.4 years in France were asked to judge the acceptability of sentences not only with respect to syntax but also discourse semantics. In another study, Abrahamsson and Hyltenstam employed several gap-filling tasks to test lexical knowledge and formulaic sequences. The testing battery administered by Granena and Long to L2 speakers of Spanish included lexical tasks on multiword correction and completion, a picture-guided narrative, and a two-word preference task, all tapping into lexical proficiency. Across these studies, significant differences between the L2 participants and native-speaking controls were found, and the lack of nativelike performance among the adult-onset learners was interpreted in favor of maturational constraints. Given these studies' focus on grammatical and other formal aspects of the lexicon, these findings are more in line with the hypothesis that critical period(s) are mostly relevant to grammatical and other formal aspects of language, compared to its lexical characteristics.

On the other hand, the study by Montrul and Slabakova (2003), which also focused on formal aspects of the lexicon, provides evidence against definitive age effects on the acquisition of the L2 lexicon. These researchers administered a truth-value judgment task and a sentence-conjunction task to investigate the preterite-imperfect distinction in L2 Spanish. They targeted a group of L2 learners of Spanish of differing proficiency levels (near-native, superior, or advanced) who began studying Spanish in high school around the age of 12. They found that approximately one third of their L2 participants performed within

the native speaker range in both tasks. More than half of these participants had been grouped as near-native.

Additional evidence against age effects in the acquisition of the L2 lexicon comes from studies focusing on vocabulary size and depth (Hellman, 2011; Marinova-Todd, 2003). Hellman administered several receptive vocabulary knowledge tests including the Peabody Picture Vocabulary Test (PPVT), a Self-Rated Vocabulary Test, and a word associates test to a group of 33 adult-onset learners of English with Hungarian as their L1, as well as to 30 monolingual and 30 bilingual native speakers of English.[3] Generally, the first group performed significantly worse on all measures, as compared to the monolingual and the bilingual native English group who defined the native range. Nevertheless, 76% of the adult-onset learners were judged as having achieved a native level of L2 vocabulary (in reference to the mean scores of the adult US native speaker population). This was partly in line with Marinova-Todd's PPVT results, showing some adult-onset learners to be indistinguishable from native controls. This finding led Hellman to argue that adult-onset L2 learners can achieve a native level in their L2 vocabulary. In addition, Hellman found five exceptional late-onset learners who outperformed the native speakers in all three administered tasks. All of them turned out to be highly educated and gifted, which suggests that they are outliers. All studies discussed here conclude that adult-onset learners who are sufficiently engaged with the L2 can attain nativelike lexical proficiency. Positive effects of education on the lexical knowledge of adult native speakers have also been found by Mulder and Hulstijn (2011). These findings support Dąbrowska's (2012) suggestion to consider level of education as a factor in research on L2 ultimate attainment with highly advanced L2 speakers.

Some evidence against maturational constraints on lexical L2 acquisition stems from studies on spontaneous speech, suggesting that adult-onset learners can attain nativelike levels in this domain. Bartning and colleagues looked at the collocational knowledge of L2 speakers of French with Swedish as their L1 (Bartning et al., 2009, 2012; Forsberg, 2010; Forsberg Lundell et al., 2013). No proportional differences in the use of collocations were found between adult onset learners with LoRs of 15 to 30 years and native speakers of French, while there were significant differences between the group with the lowest LoR (university students with 1–3 years of LoR) and all other groups with longer LoRs. These findings suggest that sufficiently long LoRs might lead to nativelike levels in the lexical domain, even among adult-onset learners. Taken together, studies based on the assessment of semi-spontaneous oral production

suggest that nativelike levels of L2 lexical knowledge can be achieved, though LoR and L1–L2 pairing appear to be moderating factors.

In sum, in the lexical domain, studies on advanced L2 proficiency are generally scarce, whether they test receptive or productive lexis. Evidence suggesting that AO might have an effect originates mostly from studies that have focused on grammatical or other formal aspects of the lexicon. Studies focusing on assessing vocabulary size based on receptive or productive lexical knowledge suggest otherwise: They indicate that other factors must be considered, including LoR and level of education.

The Current Study

Studies on the acquisition of syntactic and lexical proficiency have argued both for and against maturational constraints. One reason why the controversy has not yet been resolved may be related to methodological limitations. For example, as discussed by Long (2005), there are a number of issues concerning previous studies, above all their often narrow methodological scope with respect to the predominant use of GJTs for testing. Furthermore, as demonstrated above, very little is known about L2 grammatical and lexical complexity when it is assessed productively at the stages of advanced L2 proficiency reached after an extensive period in the L2 environment. Following Long's suggestions, DeKeyser (2013) presents a list of criteria toward better sampling for future CPH studies, among which are:

- LoR $>$ 10 years;
- sufficient variability in LoR;
- spread in socioeconomic status (SES);
- AaT below 50;
- participants who have almost exclusively used the L2.

The present study aimed at taking a step toward these suggestions by extending the scope of research to very advanced L2 speakers who have spent the majority of their lives in a L2 environment. Instead of GJTs, we analyzed speakers' spontaneous productions on the basis of oral history interviews to determine whether grammatical and lexical complexity in the L2 are related to different internal and external-experiential background variables, including AO, LoR, AaT, and level of education. We chose AOs ranging from 7 to 17 years, because a discontinuous decline is expected to be found within this age range for L2 lexis and morphosyntax according to recent findings by Granena and Long (2013).

Table 1 Overview of independent variables

Variable	*n*	*M* (*SD*)	Range/Category
Age of onset	101	12.15 (2.6)	7–17
Length of residence	100	61.33 (6.12)	41–73
Age at interview	101	73.59 (6.97)	57–87
Continued L1 exposure	98	4.34 (1.44)	1–7
Use of German at work (yes/no)	80		14/66
Gender (m/f)	102		42/60
Level of education (low/mid/high)	91		11/34/46

Given our discussion of the effects of AO and other factors on advanced L2 proficiency and nativelikeness regarding grammatical and lexical L2 attainment, we formulated the following research questions:

1. How does AO affect the complexity of L2 syntax in spontaneous oral production? Are there any other variables that might explain the acquisition of L2 grammatical complexity?
2. How does AO affect the complexity of L2 lexical use in spontaneous oral production? Are there any other variables that might explain the acquisition of L2 lexical complexity?

Method

Interviewees

The study was based on 102 oral history testimonies given by German-Jewish immigrants in the United States, the United Kingdom, and Australia between the 1970s and 2010. The L1, which they spoke exclusively before emigration, was German.[4] At the time of departure from Germany, the interviewees were between the ages of 7 and 17 years old, with a mean AO (which corresponded to their age at emigration) of 12.5 years. On average, the interviewees' LoR in the three English-speaking host countries was 61.3 years (41–73). Table 1 shows that the greater part of their lives was spent in the L2 environment. At the time of the interview (AaI), which corresponds to AaT, interviewees were on average 73.6 years old (57–87). The majority of interviewees were female ($n = 60$).

The interviewees' level of continued exposure to German (L1 Exp) after emigration was assessed by three independent raters. The raters were student assistants in the project who were also involved in the transcribing. All had

native to nativelike command of English. They were asked to rate L1 Exp on a scale from 1 (low) to 7 (high). These ratings were based on the occurrence of statements in the interview regarding (a) avoidance of speaking German, (b) manner of emigration (e.g., adoption into a foster family), (c) contact with family members, (d) origin of marital partner (native German, native English, or other L1), (e) continued use of German (during studies, work, or extracurricular engagements), and (f) integration into the English-speaking community (through studies, work, and/or extracurricular engagements). If German as the language of communication was not avoided (e.g., when emigrating with family members or having frequent contact with family members from Germany, when the partner had German heritage, or when there was a clear indication of continued L1 German use), then L1 Exp was rated as 7. The opposite was true for ratings of 1. Interrater agreement for all pairs based on Pearson correlations was $r > .70$. We considered the median values instead of the mean ratings as the final L1 Exp score to avoid the influence of outliers. Interviewees' average L1 Exp rating was 4.34, which suggests that most interviewees were still exposed to their L1 after emigration.

In addition to L1 Exp, we also included a categorical variable for the use of German at work (L1 at work) to capture professional exposure to German. This variable is more specific than L1 Exp which was based on subjective ratings. In addition, L1 at work was assumed to complement the L1 Exp variable in that the use of the L1 at work allows for a monolingual mode. Opportunities for switching may be reduced in a working environment, allowing for a more consistent exposure to and use of German (Schmid, 2007). On the other hand, with family and friends, a bilingual mode is more likely, with inconsistent exposure to the L1 (e.g., Schmid & Dusseldorp, 2010). Based on the interviewees' statements, it was established that 14 used German at work, 66 did not, and 22 did not provide any information on their use of the L1 at work. With respect to level of education (Edu), 11 of the interviewees did not obtain a high school diploma (low); 34 interviewees obtained a high school diploma or completed a vocational training if they had not finished high school (mid); 46 of the participants graduated from college or university (high). For 11 of the participants, we were unable to identify their education level due to lack of background information; these were excluded from the analyses. Table 1 gives an overview of the background variables. We refrained from using a control group. Because all data used in the analyses were from interviewees who emigrated under severe circumstances, it was not feasible to find a comparable monolingual control group (see also Schmid, 2002).

Oral History Testimonies

The data consisted of 102 oral history testimonies (autobiographical interviews) in which the participants narrated their lives before and after emigration. We acquired these testimonies from several sources, including libraries and archives in Germany, the United Kingdom, and the United States. With the rise in the oral history method over the past few decades (Pavlenko, 2007), generally accepted interviewing standards have been developed. This means that, despite their different origin with respect to the source archive, the interviews were usually quite similar in their scope and content. Each interview usually started with background questions regarding date and place of birth as well as the current age of the interviewee. Next, the interviewees talked about their childhood (and adolescence) in Germany, often with references to their parents and other family members. Most interviewees also reported their experiences of the pogrom on the night of November 9, 1938. For many, this date was a turning point after which they and their families knew that they would have to leave Germany as soon as possible. All of our interviewees left between the pogrom and the beginning of World War II on September 1, 1939. The testimonies usually proceeded with discussing the interviewees' process of emigration, their arrival in the country of destination, and the subsequent years. Given the focus of such testimonies on personal histories, they are primarily considered a historical source providing researchers with a window into people's personal past. On the other hand, these narratives also represent a rich source of spontaneous speech that may give us insights into patterns of discourse (Labov, 2013; Schiffrin, 2001) and into processes of language development, such as L1 loss (attrition) and L2 acquisition (Schmid, 2002, 2012).

Data Generation

The first step in the data analysis was to establish detailed transcripts for the first 30 minutes of each interview based on extensive guidelines. These guidelines (available from http://www.let.rug.nl/languageattrition/tools) adhere to the transcription standards of CHAT (MacWhinney, 2000) and specify what belongs in the main utterance line, how to deal with anonymization of personal information (e.g., names and places), the transcription of compound words, phrases, and collocations, as well as epistemic phrases, acronyms, incomplete, contracted, and dialectal forms. The transcripts also contained information on filled and empty pauses, repetitions, retracing and reformulation, false starts, stutters, and codeswitches. The guidelines furthermore specify to split up utterances according to the Analysis of Speech unit (AS-unit) suggested by Foster,

Tonkyn, and Wigglesworth (2000). An AS-unit must consist minimally of one or more clausal or subclausal units with the option of subordinated clauses associated with either. This approach allowed for the inclusion of fragment-like independent units common in naturalistic speech. These fragment-like units may be phrases that can be elaborated to a full clause based on content (e.g., "Five years" in response to the question, "For how long were you there?"), or they may simply be irregular phrases or nonsentences as identified by Quirk, Greenbaum, Leech, and Svartvik (1985). Intonation and pauses served as additional indicators of where to break an utterance. In most cases, utterances stretched across several lines (main tiers).

For generating syntactic measurements, the first 1,800 words of pruned speech (i.e., excluding repeated, retraced, reformulated, or incomplete material) were annotated based on the detailed guidelines described above. The syntactic annotation below each main tier included information on whether the tier was the start of an utterance and AS-unit, whether it was a main or subordinate clause or a fragment-like unit, the length of the tier according to the number of words in the pruned speech, the length of the subject, the number of finite and nonfinite verbs, and the length of noun phrases. Where applicable, additional tags were included to capture the type of relative clause (e.g., object vs. subject) and passive constructions. The following example illustrates some of the grammatical tagging:

*XYZ: so she was away (..) for quite some time.
%xcsy: UTT|AS|MC:8|SUBJ:1|V:1:0|NP:2:4|

Part-of-speech tagging to generate lexical measurements based on type-token ratios (TTRs) was carried out for the first 30 minutes of speech using the English morphological (MOR) grammar of CLAN (MacWhinney, 2000). To derive the hypernymy measure of lexical sophistication, the Coh-Metrix software (Graesser, McNamara, Louwerse, & Cai, 2004) was run on approximately the first 2,000 words per transcript (about 1,600 words of pruned speech).

Measures of Grammatical and Lexical Complexity

Based on current discussions focusing on defining and operationalizing complexity (Bulté & Housen, 2012; Ortega, 2003, 2012; Pallotti, 2014), complexity was approached here in structural terms. In that sense, the complexity of a phenomenon or entity (i.e., L2 grammar and lexicon) can be defined in terms of (a) the number and nature of its components and (b) the number and nature of connections between its components (Bulté & Housen, 2012). From this

definition follows a multidimensional construct of complexity that should be operationalized using multiple measurements.

To capture the multidimensionality of grammatical and lexical complexity (see Bulté & Housen, 2012, for an illustration), we selected several measures based on three criteria: (a) the measures should have little overlap (conceptually) and should capture distinct dimensions, such as diversity (i.e., the variation in grammatical and lexical choices) and sophistication (i.e., the use of advanced grammatical constructions and lexical items); (b) the measures should be applicable to spoken language, capturing characteristics, such as noun phrase length and diversity, which are generally assumed to be limited in conversations (Leech, 2000) and thus reveal differences in complexity; (c) finally, they should tap into distinct proficiency levels of L2 users who are beyond the advanced stage of L2 development, where especially morphosyntactic measures have been demonstrated to be indicative of nonnativeness (Bartning, 2012). Because English does not make use of extensive inflectional morphology, and given that we targeted highly advanced L2 speakers, we chose to consider voice morphology by looking at the frequency of passive use.[5] All in all, we generated several measures to assess grammatical and lexical complexity.

We used five measures of grammatical complexity. For syntactic complexity, we chose mean number of words per AS-units to capture the sentence level, mean number of dependent clauses and mean number of nonfinite adverbial dependent clauses (DCs) per AS-unit to capture the subclausal level, and mean number of words per noun phrase to capture the phrasal level. As Norris and Ortega (2009) point out, for advanced learners, complexification is to be expected especially at the phrasal level, where length of noun phrases, rather than number of subordinate clauses, has been found to increase (e.g., Michel, Kuiken, & Vedder, 2007). The number of passive constructions per clause served as a morphological measure.

Lexical complexity is again a multidimensional construct for which several dimensions have been distinguished: diversity, density, compositionality, and sophistication (see Bulté & Housen, 2012, for illustration). However, Šišková (2012) showed that various measures apparently capturing diversity and sophistication overlap. Based on her findings and recent investigations of lexical characteristics of L2 output (Crossley, Salsbury, McNamara, & Jarvis, 2011; Jarvis, 2013b), we selected four measures. For diversity, TTR was chosen; and for density, a ratio of content words (nouns, verbs, and adjectives) to the total number of words. Both measures capture the entire spectrum of word classes. For sophistication, we selected two measures: frequency bands to capture the level of infrequent content words (nouns, verbs, and adjectives) and mean

hypernymy to capture the level of abstraction of selected content words (i.e., nouns and verbs). The frequency band ratio is a corpus-internal, frequency-based measure (see Verspoor, Schmid, & Xu, 2012, for details), showing the relationship between infrequent lemmas used by each interviewee in comparison to the infrequent lemmas used across the overall corpus. Hypernymy refers to semantic relationships between words (for nouns and verbs only), capturing associations between hypernyms (superordinate words, such as *entity* or *furniture*) and hyponyms (subordinate words, such as *chair* or *stool*). In the WordNet scale (Crossley, Salsbury, & McNamara, 2009; Fellbaum, 2013), a word located at the lower end of the hierarchy (e.g., *entity*) has a higher hypernymy level and is thus more abstract, while a word located at the higher end (e.g., *chair*) is more concrete.[6] Thus, through the measure of hypernymy, we captured the degree of abstractness of nouns and verbs used in the interviewees' oral production. As Crossley et al. (2009) found, advanced learners have access to a wider range of hypernymy levels; hypernymy was also found to be a predictor of holistic human lexical proficiency ratings (Crossley et al., 2011).

Mixed-Effects Modeling

Because each interviewee received multiple complexity scores, we conducted linear mixed-effects regression modeling, using the *lme4* package (version 1.1-7) for the *R* statistical platform (R Core Team, 2012; version 3.1.1), with interviewee as a random-effect factor, in order to take the structural variation linked to each interviewee into account. We assessed if random intercepts and random slopes were necessary by comparing the Akaike Information Criterion (AIC; Akaike, 1974). The inclusion of random intercepts allows to take the variability associated with the interviewees into account (some interviewees tend to have higher complexity scores than others). The inclusion of random slopes allows for accounting the variability in the effect of a certain predictor. The AIC offsets the complexity of the model to the goodness of fit. An AIC difference of at least 2 (with the more complex model having a lower AIC) indicates that the higher complexity of the more complex model is warranted (Wieling, Nerbonne, & Baayen, 2011). AIC is related to the evidence ratio, which expresses the relative probability that the model with the lowest AIC is more likely to provide a more precise fit to the data (Blankevoort et al., 2013). The consideration of random intercepts and slopes avoids the possibility of being anti-conservative, such as reporting too high p values (Baayen, 2008; Baayen, Davidson, & Bates, 2008). To perform mixed-effects regression modeling using the different complexity scores, we first z-transformed all grammatical and lexical measures. Using the reshape package (version 0.8.5) in

Table 2 Overview of grammatical and lexical measures

Dimension	Level	Statistical construct	*M* (*SD*)	*Min*	*Max*
Syntactic diversity	Sentence/clause	Words per AS-unit	9.47 (1.34)	6.18	12.53
	Subclause	DCs per AS-unit	0.48 (0.13)	0.20	0.83
	Subclause	Non-finite adverbial DCs per AS-unit	0.09 (0.05)	0.01	0.26
	Phrase	Words per NP	1.79 (0.15)	1.51	2.12
Morphological sophistication		Passives per clause	0.05 (0.02)	0.01	0.14
Diversity	All words	Type-token ratio (TTR)	0.27 (0.02)	0.22	0.33
Density	All words	Content words/total words	0.32 (0.02)	0.25	0.38
Sophistication	Content words	Frequency bands	0.20 (0.05)	0.10	0.32
	Nouns and verbs	Hypernymy mean scores	1.42 (0.11)	1.16	1.71

Note. AS-unit = analysis of speech unit, DC = dependent clause, NP = noun phrase. Descriptive statistics represent raw values.

R, we merged the scores into a grammatical complexity and a lexical complexity score which were again *z*-transformed. In the process of model building, only significant predictors were kept that led to decreasing numbers of the interviewees contributing to each subsequent model.

Results

Grammatical Complexity

The statistical model for grammatical complexity was based on 91 interviews and 455 grammatical complexity scores. Table 2 shows descriptive statistics for each grammatical complexity measure.

Table 3 shows the coefficients and associated statistics for the fixed-effect factors and covariates in the final mixed-effects regression model. The explained variance of the complete model including all random intercepts and slopes was 74%; the fixed-effect predictors on their own accounted for 10% of the variance. The model summarized in Table 3 shows that male interviewees ($\beta = .528$, $t = 3.775$) and those with more education ($\beta = .228$, $t = 2.272$) obtained overall

Table 3 Linear mixed-effect model for grammatical complexity scores

Fixed effect	Estimate	*SE*	*t*	*p*
Intercept (gender female)	–0.793	0.254	–3.127	< .01
Gender male	0.528	0.140	3.775	< .001
Level of education	0.228	0.100	2.272	< .05

Note. *SE* = standard error.

Table 4 Fixed-effect goodness of fit for grammatical complexity

Additional fixed effects	Log-likelihood increase	AIC decrease	Evidence ratio	Likelihood ratio test	Additional *df*
Random intercept only					
Gender male (vs. female)	5.63	9.26	102.56	$p < .001$	1
Level of education	72.85	143.70	>1000.00	$p < .0001$	1

Note. Each row specifies the significant increase in goodness of fit obtained by adding the current predictor to the model including all preceding predictors. AIC = Akaike Information Criterion, df = degrees of freedom.

higher complexity scores. The other potentially confounding variables (i.e., AO, AaI, LoR, and L1 Exp) did not reach significance alone or through interactions with any other variable and were therefore excluded from the model. Eleven interviewees had to be excluded due to missing information on their education level.

Table 4 provides an overview of how log-likelihood and AIC values changed with the inclusion of the significant fixed-effect predictors while keeping the random-effects structure constant (see Wieling et al., 2011). The baseline model only consisted of the random intercepts for the interviewee. The AIC decreased by at least 2 with the stepwise addition of each fixed-effect factor, suggesting that the inclusion of gender and level of education in the final model was warranted. Data inspection revealed that the distribution of residuals was normal, as illustrated in Figure A1, available in Appendix S1 in the Supporting Information online.

We did not find a significant effect for AO. A graphical illustration of the data (depicted in Figure 1) shows the *z*-transformed scores for the individual grammatical complexity measures in relation to AO. It reveals that higher

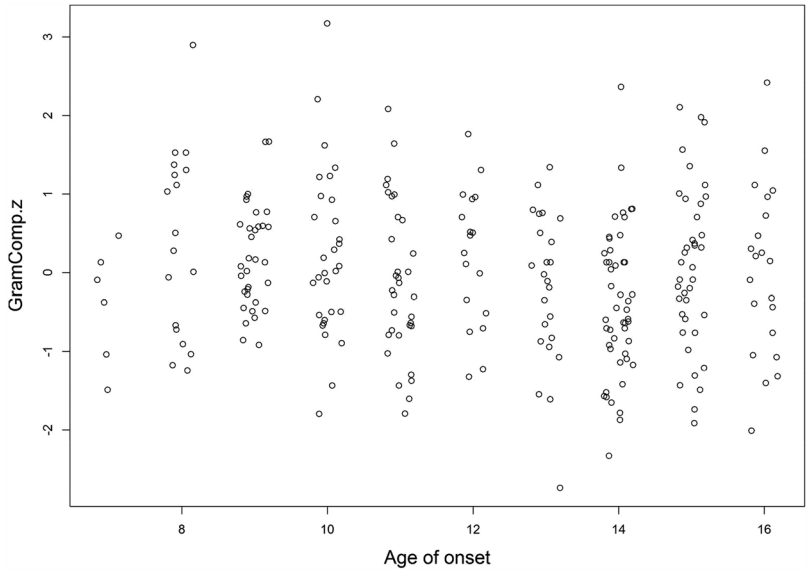

Figure 1 The relationship between age of onset and grammatical complexity, with individual grammatical complexity scores (GramComp.z) plotted against age of onset for all speakers.

grammatical complexity scores were found across all AOs. The highest scores were obtained by the interviewees with AOs around and below the age of 12, whereas lower scores were more likely to be obtained by the interviewees with AOs beyond 12.

A closer examination of several outstanding cases, that is, interviewees with the highest and lowest scores, revealed that the three interviewees with the highest scores (for the number of passives per clause and the number of nonfinite dependent clauses per AS-unit) were male. Their AO ranged from 9 to 12, their AaI ranged from 68 to 81, and their LoR from 59 to 71. All three had a medium to high level of education and were found to have a high degree of continued L1 exposure (with a rating between 5 and 7). They all had succeeded professionally in the areas of photography, business, and academia. The interviewees with the lowest scores (for the number of dependent clauses per AS-unit and the number of words per AS-unit) were female. They had AO ranging from 14 to 16, with AaI ranging from 68 to 78, and LoR from 52 to 64 years. Two of them had obtained a low level of education, whereas one had reached a high level. Their continued exposure to German ranged from medium

Table 5 Linear mixed-effect model for lexical complexity scores

Fixed effect	Estimate	*SE*	*t*	*p*
Intercept (gender female)	–0.925	0.260	–3.553	< .001
Gender male	0.394	0.145	2.714	< .01
Level of education	0.278	0.103	2.696	< .01
German at work	0.514	0.183	2.802	< .01

Note. SE = standard error.

(3) to high frequency (7); and in terms of professional occupations, they varied from one interviewee to another (i.e., homemaker, UN employee).

Lexical Complexity

The statistical model for lexical complexity was based on 75 interviews and 300 scores (see Table 2 for the descriptive statistics). Table 5 shows the coefficients and associated statistics for the fixed-effect factors and covariates in the final mixed-effects regression model. The explained variance of the complete model including all random intercepts and slopes was 51%; the fixed-effect predictors on their own accounted for 13% of the variance.

The model summarized in Table 5 shows that the interviewees who were male ($\beta = .394$, $t = 2.714$), highly educated ($\beta = .277$, $t = 2.696$), and used German at work ($\beta = .514$, $t = 2.802$) obtained higher lexical complexity scores. The other potentially confounding variables (i.e., AO, AaI, LoR, and L1 Exp) did not reach significance alone or through interactions with any other variable and were therefore not included in the model. Twenty-seven interviewees had to be excluded due to missing information on their education and missing data on their use of German at work. Data inspection again revealed that the distribution of residuals was normal, as illustrated in Figure A2, available in Appendix S1 in the Supporting Information online.

Table 6 provides an overview of how log-likelihood and AIC values changed with the stepwise inclusion of the significant fixed-effect predictors while keeping the random-effects structure constant (Wieling et al., 2011). The inclusion of each additional predictor resulted in an AIC decrease of at least 2, yielding the final model with the three fixed-effect factors (gender, level of education, and the use of German at work).

A qualitative analysis of the interviewees with the highest and lowest lexical complexity scores (shown in Figure 2) shows that the three interviewees with the highest scores (for the ratio of content words over the total words and TTR) were two male and one female. Their AO ranged from 12 to 16, and

Table 6 Fixed-effect goodness of fit for lexical complexity

Additional fixed effects	Log-likelihood increase	AIC decrease	Evidence ratio	Likelihood ratio test	Additional *df*
Random intercept only					
Gender male (vs. female)	3.96	5.91	19.25	$p < .01$	1
Level of education	59.83	117.66	>1000.00	$p < .0001$	1
German at work	94.71	187.42	>1000.00	$p < .0001$	1

Note. Each row specifies the significant increase in goodness of fit obtained by adding the current predictor to the model including all preceding predictors. AIC = Akaike Information Criterion, *df* = degrees of freedom.

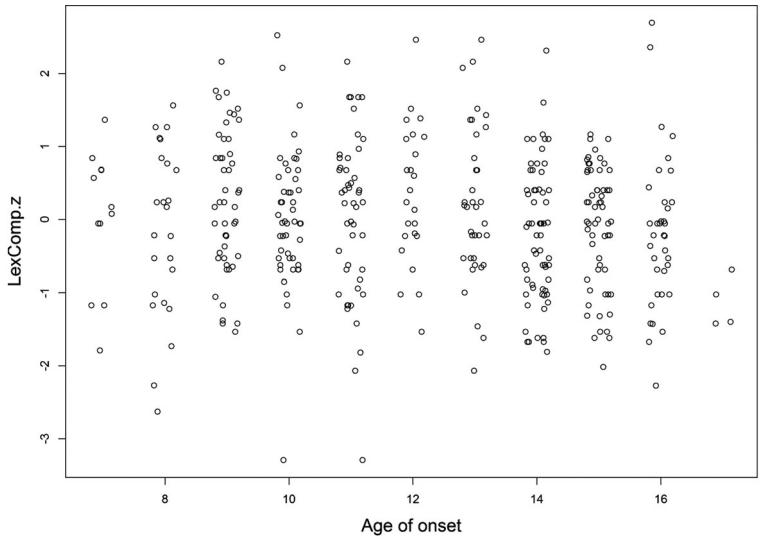

Figure 2 The relationship between age of onset and lexical complexity, with individual lexical complexity scores (LexComp.z) plotted against age of onset for all speakers.

their AaI from 71 to 81; their LoR ranged from 59 to 66 years. The three interviewees had had frequent continued L1 exposure (ratings of 5–6) and had obtained a medium to high level of education. Whereas two pursued academic careers as university lecturers, one worked as a professional photographer. The interviewees with the lowest score (for the ratio of content words over the total

words and TTR) were two male and one female with AO ranging from 8 to 11. For one of the interviewees we have no further background information; the other two were 65 and 66 years old at the time of the interview and had resided in the L2 environment for 55 and 57 years, respectively. Both had continued L1 exposure ranging from some (3) to a high degree (7) and had obtained a medium to high level of education (one worked as a nurse whereas the other was an engineer).

Discussion

Age of Onset

Our findings do not lend support to the CPH and maturational constraints. The analyses did not reveal a significant AO effect, neither for grammatical complexity nor for lexical complexity at the word level (see Figures 1 and 2). In fact, the relationship between AO and grammatical and lexical complexity scores was based on a similar distribution of scores across all AOs. However, when looking at the highest and lowest scores, we found different patterns for grammatical and lexical complexity.

Regarding grammatical complexity, we found that very low scores were obtained by several interviewees with an AO above 12. Very high scores, on the other hand, were obtained by several interviewees with AOs below 12. This qualitative analysis of the grammatical complexity scores supports previous research that has found AO effects, particularly when grammar and other formal aspects of the lexicon are investigated (e.g., Abrahamsson & Hyltenstam, 2009; Forsberg-Lundell et al., 2013). A different picture emerged for lexical complexity. Here several of the interviewees with AOs below 12 obtained the lowest scores, whereas interviewees across all AOs obtained high scores, suggesting the lack of AO effects on vocabulary knowledge. A detailed look at the data for the interviewees with the highest and lowest lexical complexity scores reveals differences in age at the time of the interview and the professions that the interviewees pursued. Interestingly, the interviewees who scored highest on the lexical measures were older at the time of the interview than those who scored lowest. This suggests that lexical growth takes place well into older age (e.g., Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). Qualitative inspection of the data further suggests that a profession that requires extensive use of, and exposure to, language (such as teaching) might benefit productive lexical complexity.

Altogether, the lack of evidence for age effects in our study, as compared to the findings of previous research (e.g., Abrahamsson, 2012; Abrahamsson & Hyltenstam, 2009; Schmid, 2014), might be due to a number of factors.

First, our interviewees spent most of their lives in the host country. Their LoR exceeded the usual time span found in similar studies by at least 10 to 20 years (Erman, Denke, Fant, & Forsberg Lundell, 2014; Forsberg Lundell et al., 2013). Even though it has often been suggested that not much linguistic change occurs beyond 10 years of LoR (Stevens, 2006), one cannot ignore that there are areas of language, such as the lexicon, which change continuously for L1 and L2 speakers alike. This would also explain why the qualitative analysis of the lexical complexity scores showed that the interviewees with higher ages at emigration obtained some of the highest scores. What may be at play here is that the L1 lexicon, which had already developed substantially by the time of emigration, might have served as a foundation for building and enriching the L2 lexicon. A larger pool of L1 lexical items could thus lead to a positive transfer effect from the L1 to the L2 lexicon.

A second reason for the lack of age effects may be that all interviewees emigrated from Germany under exceptional circumstances. Many of them had been directly exposed to anti-Semitic actions and to the violence by the Nazi regime. Especially those interviewees who were adolescents at the time would have experienced the increasing hostility, which was likely to trigger their motivation for emigration and for distancing themselves from Germany. Many of our interviewees explicitly talked about their wish of leaving the past behind and assimilating to the host community, which might partly explain why we did not find an AO effect. This hypothesis is in line with Schmid's (2002) findings of severe L1 attrition in a similar group of German-Jewish refugees who emigrated shortly before World War II. Finally, finding no AO effects may be due to the chosen age range. While we cannot exclude the possibility that the window of opportunity closes before the chosen age range, research targeting grammar (particularly morphosyntax) has repeatedly demonstrated that the offset of a critical period is to be expected around the age of 12 (e.g., Granena & Long, 2013; for reviews, see DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003).

Instead of AO effects, which would be specifically related to a L2 learning context, our findings of significant level of education effects on grammatical and lexical complexity appear to be consistent with patterns to be expected in a large group of native speakers varying in levels of education (e.g., Dąbrowska, 2012; Mulder & Hulstijn, 2011). In other words, our interviewees' L2 proficiency was not related to variables specific to their L2 learning context (e.g., AO) but was linked to variables, such as level of education and gender, which might affect language learning in general and might influence the performance of L1 and L2 speakers in a similar manner.

Gender

For grammatical and lexical complexity, we found a significant effect of gender. If interviewees were male, they were more likely to obtain higher scores in both domains. Qualitative analyses confirmed this difference for grammatical complexity, where the interviewee with the highest score was male and the one with the lowest score female. To our knowledge, gender effects regarding grammatical and lexical L2 proficiency have not received much attention. Potential reasons for the obtained gender effect might be specific to our data set and the population under investigation. It is conceivable that the men in our sample were more likely to be the providers of the family, as would have been common at the time. Being responsible for their families, men were likely to engage in the labor market more actively than women. In turn, they would have had more opportunities to converse and to connect to English-speaking coworkers and hence to integrate more easily than their female counterparts who often took over the role of being homemakers. Men's chances for conversation therefore included (to some extent), but also expanded beyond, the private realm. On the other hand, women's realm of communication, especially if they stayed at home, might have been more restricted to family, neighbors, and fellow women. However, there was no correlation between gender and level of education and between gender and the choice to pursue a professional career for the interviewees in our data set, suggesting that women were equally well educated and that they also pursued professional occupations. For now, there is no clear explanation for why men obtained significantly higher scores than women, based on their level of education and their career choice. Therefore, our findings require further investigation. Immigrant communities where men traditionally hold the role of the provider may help to shed light on our current interpretations.

Level of Education

We also found a significant effect of level of education for grammatical and lexical complexity. If interviewees were highly educated, they were more likely to obtain higher scores in both domains. The education effect can be regarded an external-experiential, sociopsychological factor according to Birdsong's (2006) classification. Level of education has been found to play an important role in L2 proficiency, particularly in the lexical domain (e.g., Hellman, 2011). Notably, level of education effects are not specific to L2 learning but have been demonstrated for L1 proficiency and L1 grammar (Dąbrowska, 2012; Dąbrowska & Street, 2006) as well as the L1 lexicon (Mulder & Hulstijn, 2011). Because we found this effect among L2 speakers, with no significant

effect of AO, we tentatively conclude that our interviewees show similar patterns as did the native speakers in the studies reviewed above; in essence, level of education appears to have a positive effect not only on various linguistic aspects of language users' L1 but also on our interviewees' L2 grammar and L2 lexical knowledge. Qualitative analyses further revealed that the interviewees with the highest scores for grammatical and lexical complexity not only had obtained a higher level of education but were also professionally accomplished, mainly as businessmen or academics. The choice of professional occupation might also have had an effect on their language input and thus their L2 proficiency. However, due to missing data, we were unable to pursue this question further.

Use of L1 German in the Workplace

The more general variable capturing continued exposure to German after emigration (L1 Exp) did not yield any effect, while the more specific factor, the use of German at work, did show a positive effect on lexical complexity. This finding is speculative as, to the best of our knowledge, there are no studies targeting complexity that have addressed this particular relationship. However, as Wolter (2006) suggests, the L1 lexicon can be both beneficial and detrimental to the acquisition of a L2 lexicon. In our case, for lexical complexity, there appears to be a beneficial effect of bilingualism (i.e., active and consistent use of both languages). This might partly be due to the high degree of abstract words with a Greco-Latin origin in both languages. The usage of such words should contribute to a more sophisticated lexical production, especially when including hypernymy as a measure of lexical proficiency, as in our analyses. Alternatively, a continuous use of the L1 at work might also encourage the L2 speaker to expand his/her L2 vocabulary. Furthermore, bilinguals' extensive linguistic experience probably enables them to use language much more creatively, allowing for a more diverse linguistic output. However, these findings require further investigation.

Limitations

There are several limitations to this study. First, there may be variation due to posttraumatic stress disorder (PTSD) from which some interviewees might have suffered given their life experiences. Previous studies on Holocaust survivors diagnosed with PTSD found poorer learning and memory performance (Golier, Yehuda, Lupien, & Harvey, 2003; Yehuda, Golier, Halligan, & Harvey, 2004). Such learning and memory deficits might be relevant to language acquisition and speech production, suggesting that trauma can affect language. Unfortunately, precise clinical details of individual participants were not available to us

and therefore could not be considered. Second, we also refrained from using a monolingual, native speaker control group given our participants' experience of emigration and the possibility of having suffered from PTSD. Future studies could attempt to find suitable control groups, possibly with similar traumatic experiences. For example, a group of native speakers, such as war veterans who have likely suffered from traumatic experiences, could be considered as a possible comparison group for war time emigrants. However, it is unclear to what extent trauma would be comparable across such groups of participants and to which degree it would affect language production. Furthermore, World War II veterans would have probably been slightly older at the time of conscription, as compared to our interviewees who were mostly children or young adolescents when they experienced anti-Semitism. The difficulty of finding a control group underlines the special characteristics of our participant group due to which we can only draw tentative conclusions regarding their L2 nativelikeness.

A third limitation concerns the rather high age of our interviewees at the time of the interview. As suggested by Long (2005), when looking at an aged population when testing the CPH, one should take into account general aging effects and cognitive decline. We therefore checked whether age at the time of the interview would affect our measurements, given the range of interview ages in our sample. While our analyses did not yield a significant effect, we do agree with Long that, at such an advanced age, a general cognitive decline might affect language performance. For instance, it has been shown for native speakers that syntactic complexity in diary entries decreases with increasing age (e.g., Kemper, 1987); on the other hand, it has been argued that lexical growth continues with age (e.g., Ramscar et al., 2014).

A fourth limitation concerns the grammatical and lexical complexity measures we used. The selection of measures was motivated by previous research, which has frequently employed such measures to assess L2 proficiency (see Ortega, 2003). However, our measures were limited, and their suitability for capturing AO effects and nativelikeness is debatable. While we did make an effort to capture the multidimensionality of both constructs, their definitions and operationalization are still very much work in progress (Bulté & Housen, 2012; Jarvis, 2013a; Pallotti, 2014). More fine-grained syntactic measures capturing, for example, sentence/utterance type, additional morphological measures, such as the number of exponents or forms taken by lexemes to express different grammatical and categorical functions (Pallotti, 2014), as well as collocational measures like those used by Bartning and colleagues or Erman and Warren (2000), might help to tap more carefully into the very advanced levels of L2

proficiency. As Erman et al. (2014) and Forsberg Lundell et al. (2013) have demonstrated, assessing advanced L2 proficiency by means of collocations appears to be a worthwhile endeavor and should be further explored in future studies on advanced L2 proficiency.

Finally, our findings are specific to the areas of grammatical and lexical complexity. There are of course other types of linguistic complexity (e.g., phonological, discourse, etc.) that one could consider investigating as well as other dimensions proposed by the CAF framework, such as accuracy and fluency. As Abrahamsson and Hyltenstam (2009) found, listener perception tapping into the dimension of phonological accuracy in combination with several measures of linguistic performance, representation, and processing promise to provide for a thorough investigation of the advanced L2 learner. Adding fluency and accuracy will yield an even more complete picture of advanced L2 proficiency.

Conclusion

In this study, we assessed the grammatical and lexical complexity of spontaneous oral productions by long-term L2 speakers, investigating how these complexity measures are affected by AO, LoR, continued L1 use, level of education, and other potential factors. In an attempt to capture the multidimensionality of grammatical and lexical complexity (Bulté & Housen, 2012; Jarvis, 2013a; Pallotti, 2014), we collected a number of measures for both aspects of linguistic complexity. By means of mixed-effect modeling, we assessed which variables would predict the various measurements. The analyses showed that syntactic and lexical complexity at the productive level was not affected by variables that would suggest a critical period or maturational constraints on L2 learning. Instead, level of education—an experiential factor—played a significant role, as would be expected also for native speakers. It thus seems that any possible maturational effects in our data are superseded by a general, non-L2-specific effect of level of education, and that very high levels of L2 proficiency in the morphosyntactic and lexical domain are attainable across all ages of onset. In addition, gender and (for lexical complexity) the use of German at work were also found to have an effect, which might be due to the nature of our sample. However, qualitative analyses of the three lowest and highest performing interviewees indicated that AO is more likely to play a role with regard to the acquisition of L2 grammar, as opposed to the L2 lexicon. This is because the highest grammar scores were obtained by interviewees with AO below 12 whereas the highest lexical scores were obtained by interviewees

with AO beyond 12. This result confirms previous findings, suggesting that advanced levels of L2 productive proficiency are attainable in the areas of grammar and lexis, independently of AO of L2 learning.

Final revised version accepted 15 May 2015

Notes

1 ND refers to one of the interviewees from a subcorpus of oral history testimonies obtained from the University of Southern California Shoah Foundation, The Institute for Visual History and Education.
2 Their amount of exposure may be moderated and limited by internal (e.g., motivation, attitude) as well as external (e.g., living situation, integration experiences, etc.) factors. But it most likely exceeds the amount of L2 exposure that classroom learners receive.
3 They were all native speakers of English who either grew up in a U.S. household where a language other than English was spoken or they were exposed to a foreign language while living abroad for many years (Hellman, 2011).
4 None of the interviewees spoke Yiddish. Several mentioned that they had started taking English at school in preparation for their emigration.
5 We acknowledge that the passive is not a purely morphological measure. It is rather a morphosyntactic feature because its use also requires syntactic restructuring as explained by König and Gast (2009).
6 This is unlike to what a hypernymic hierarchy would usually look like. Here a word located at the lower end is more concrete, while a word at higher end is more abstract.

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*, 187–214. doi:10.1017/S0272263112000022

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481–509. doi:10.1017/S027226310808073x

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249–306. doi:10.1111/j.1467-9922.2009.00507.x

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R* (1st ed.). Cambridge, UK: Cambridge University Press.

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

Bartning, I. (2012). High-level proficiency in second language use: Morphosyntax and discourse. In M. Watorek, S. Benazzo, & M. Hickmann (Eds.), *Comparative perspectives to language acquisition: A tribute to Clive Purdue* (pp. 170–187). Bristol, UK: Multilingual Matters.

Bartning, I., Forsberg, F., & Hancock, V. (2009). Resources and obstacles in very advanced L2 French: Formulaic language, information structure and morphosyntax. *EUROSLA Yearbook*, *9*, 185–211. doi:10.1075/eurosla.9.10bar

Bartning, I., Forsberg Lundell, F., & Hancock, V. (2012). On the role of linguistic contextual factors for morphosyntactic stabilization in high-level L2 French. *Studies in Second Language Acquisition*, *34*, 243–267. doi:10.1017/S0272263112000046

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, *13*, 116–137. doi:10.1191/026765897677670241

Bialystok, E., & Hakuta, K. (1995). *In other words: The science and psychology of second-language acquisition*. New York: Basic Books.

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

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics in Language Teaching*, *43*, 319–328. doi:10.1515/iral.2005.43.4.319

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 235–249. doi:10.1006/jmla.2000.2750

Blankevoort, C. G., Scherder, E. J. A., Wieling, M. B., Hortobágyi, T., Brouwer, W. H., Geuze, R. H., et al. (2013). Physical predictors of cognitive performance in healthy older adults: A cross-sectional analysis. *PLoS ONE*, *8*, e70799. doi:10.1371/journal.pone.0070799

Bley-Vroman, R. W., Felix, S. W., & Ioup, G. L. (1988). The accessibility of Universal Grammar in adult language learning. *Second Language Research*, *4*, 1–32. doi:10.1177/026765838800400101

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21–46). Amsterdam: John Benjamins.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, *63*, 544–573. doi:10.2307/415005

Crossley, S. A., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, *59*, 307–334. doi:10.1111/j.1467-9922.2009.00508.x

Crossley, S. A., Salsbury, T. L., McNamara, D., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*, 561–580. doi:10.1177/0265532210378031

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*, 219–253. doi:10.1075/lab.2.3.01dab

Dąbrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*, 604–615. doi:10.1016/j.langsci.2005.11.014

DeKeyser, R. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*, 52–67. doi:10.1111/j.1467-9922.2012.00737.x

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). Oxford, UK: Oxford University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–534.

Erman, B., Denke, A., Fant, L., & Forsberg Lundell, F. (2014). Nativelike expression in the speech of long-residency L2 users: A study of multiword structures in L2 English, French and Spanish. *International Journal of Applied Linguistics*, *25*, 160–182. doi:10.1111/ijal.12061

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text—Interdisciplinary Journal for the Study of Discourse*, *20*, 29–62. doi:10.1515/text.1.2000.20.1.29

Fellbaum, C. (2013). WordNet. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–8). Chichester, UK: Blackwell. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal1285/abstract

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104. doi:10.1006/jmla.1999.2638

Forsberg, F. (2010). Using conventional sequences in L2 French. *International Review of Applied Linguistics in Language Teaching*, *48*, 25–51. doi:10.1515/iral.2010.002

Forsberg Lundell, F., Bartning, I., Engel, H., Gudmundson, A., Hancok, V., & Lindqvist, C. (2013). Beyond advanced stages in high-level spoken L2 French. *Journal of French Language Studies*, *24*, 255–280. doi:10.1017/S0959269513000057

Forsberg Lundell, F., & Lindqvist, C. (2012). Vocabulary aspects of advanced L2 French: Do lexical formulaic sequences and lexical richness develop at the same rate? *Language, Interaction & Acquisition*, *3*, 73–92. doi:10.1075/lia.3.1.05for

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–375. doi:10.1093/applin/21.3.354

Golier, J. A., Yehuda, R., Lupien, S. J., & Harvey, P. D. (2003). Memory for trauma-related information in Holocaust survivors with PTSD. *Psychiatry Research*, *121*, 133–143. doi:10.1016/S0925-4927(03)00120-3

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202. doi:10.3758/BF03195564

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*, 311–343. doi:10.1177/0267658312461497

Hellman, A. B. (2011). Vocabulary size and depth of word knowledge in adult-onset second language acquisition. *International Journal of Applied Linguistics*, *21*, 162–182. doi:10.1111/j.1473-4192.2010.00265.x

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931. doi:10.1016/j.lingua.2009.06.004

Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, *34*, 361–394. doi:10.1017/S0142716411000737

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 538–588). Chichester, UK: Blackwell.

Hyltenstam, K., & Abrahamsson, N. (2012). Introduction. *Studies in Second Language Acquisition*, *34*, 177–186. doi:10.1017/S0272263112000010

Jarvis, S. (2013a). Capturing the diversity in lexical diversity. *Language Learning*, *63*, 87–106. doi:10.1111/j.1467-9922.2012.00739.x

Jarvis, S. (2013b). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 13–44). Amsterdam: John Benjamins.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99. doi:10.1016/0010-0285(89)90003-0

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*, 215–258. doi:10.1016/0010-0277(91)90054-8

Kemper, S. (1987). Life-span changes in syntactic complexity. *Journal of Gerontology*, *42*, 323–328. doi:10.1093/geronj/42.3.323

König, E., & Gast, V. (2009). *Understanding English-German contrasts*. Berlin, Germany: Schmidt Erich.

Labov, W. (2013). *The language of life and death: The transformation of experience in oral narrative*. Cambridge, UK: Cambridge University Press.

Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, *50*, 675–724. doi:10.1111/0023-8333.00143

Lenneberg, E. H. (1967). *Biological foundations of language*. Oxford, UK: Wiley.

Long, M. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics in Language Teaching*, *43*, 287–317. doi:10.1515/iral.2005.43.4.287

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*, 588–595. doi:10.1017/S1366728911000010

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Abingdon, UK: Routledge.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85–104. doi:10.1191/0265532202lt221oa

Marinova-Todd, S. H. (2003). *Native, near-native or non-native: Comprehensive analysis of the linguistic profiles of highly proficient adult second language learners*. Unpublished doctoral dissertation, Harvard Graduate School of Education, Boston.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, *34*, 9–34.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*, 241–259. doi:10.1515/iral.2007.011

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers. *Studies in Second Language Acquisition*, *25*, 351–398. doi:10.1017/S0272263103000159

Mulder, K., & Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Applied Linguistics*, *32*, 475–494. doi:10.1093/applin/amr016

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, *44*, 1–35. doi:10.1017/S0261444810000327

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555–578. doi:10.1093/applin/amp044

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*, 492–518. doi:10.1093/applin/24.4.492

Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). Berlin, Germany: Walter de Gruyter.

Pallotti, G. (2014). A simple view of linguistic complexity. *Second Language Research*, *31*, 117–134. doi:10.1177/0267658314536435

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*, 449–468. doi:10.1111/j.1467-1770.1980.tb00328.x

Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics*, *28*, 163–188. doi:10.1093/applin/amm008

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language (general grammar)* (2nd ed.). New York: Longman.

R Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. (Version 3.1.2). [Computer software].

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*, 5–42. doi:10.1111/tops.12078

Rothman, J. (2008). Why all counter-evidence to the Critical Period Hypothesis in second language acquisition is not equal or problematic. *Language and Linguistics Compass*, *2*, 1063–1088. doi:10.1111/j.1749-818X.2008.00098.x

Schiffrin, D. (2001). Language and history: Oral history as discourse. In D. Tannen & J. E. Alatis (Eds.), *Georgetown University Roundtable on Language and Linguistics* (pp. 84–113). Washington, DC: Georgetown University Press.

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

Schmid, M. S. (2007). The role of L1 use for L1 attrition. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 135–154). Amsterdam: John Benjamins.

Schmid, M. S. (2012). The impact of age and exposure on bilingual development in international adoptees and family migrants: A perspective from Holocaust survivors. *Linguistic Approaches to Bilingualism*, *2*, 177–208. doi:10.1075/lab.2.2.03sch

Schmid, M. S. (2014). The debate on maturational constraints in bilingual development: A perspective from first-language attrition. *Language Acquisition*, *21*, 386–410. doi:10.1080/10489223.2014.892947

Schmid, M. S., & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research*, *26*, 125–160. doi:10.1177/0267658309337641

Šišková, Z. (2012). Lexical richness in EFL students' narratives. *University of Reading: Language Studies Working Papers*, *4*, 26–36.

Stevens, G. (2006). The age-length-onset problem in research on second language acquisition among immigrants. *Language Learning*, *56*, 671–692. doi:10.1111/j.1467-9922.2006.00392.x

van Boxtel, S., Bongaerts, T., & Coppen, P.-A. (2003). Native-like attainment in L2 syntax. *EUROSLA Yearbook*, *3*, 157–181. doi:10.1075/eurosla.3.10box

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*, 239–263. doi:10.1016/j.jslw.2012.03.007

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, *12*, 233–265. doi:10.1177/026765839601200301

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*, e23613. doi:10.1371/journal.pone.0023613

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*, 741–747. doi:10.1093/applin/aml036

Yehuda, R., Golier, J. A., Halligan, S. L., & Harvey, P. D. (2004). Learning and memory in Holocaust survivors with posttraumatic stress disorder. *Biological Psychiatry*, *55*, 291–295. doi:10.1016/S0006-3223(03)00641-3

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Distribution of Residuals in Fixed-Effects Models.