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

Author Note

Cornelia Lahmann, English Department, University of Groningen

Rasmus Steinkrauss, English Department, University of Groningen

Monika Schmid, Department of Language and Linguistics, University of Essex/English Department, University of Groningen

This research was financed by the Dutch national research foundation (NWO; Nederlandse Organisatie voor Wetenschappelijk Onderzoek), Grant No. 360-70-420. We also thank the following archives for granting us the permission to use their oral history testimony data: USC 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, D.C.), Association of Jewish Refugees (London).

Correspondence concerning this article should be addressed to Cornelia Lahmann, University of Groningen, English Department, Faculty of Arts, P.O. Box 716, NL 9700AS Groningen

Email: cornelia.lahmann@gmx.de

Abstract:

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 which 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. Multifactorial analyses yielded consistently significant effects of gender and level of education for grammatical and lexical complexity. Additionally, L1 use at work was found to be a significant variable for lexical complexity, improving the model fit significantly. Conversely, AO did not figure significantly. We conclude that grammatical and lexical complexity at the stage of L2 ultimate attainment are the result of a complex interplay of variables which are more general to language learning and performance, rather than SLA-specific in nature.

Keywords: naturalistic L2 attainment, spontaneous speech, grammatical complexity, lexical complexity

58 years after escaping from Nazi Germany to England as a child, N.D.[1] recalled: "I think within six 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), e.g., "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 second language acquisition (SLA) (e.g., Granena & Long, 2013; Luk, De Sa, & Bialystok, 2011), but there is disagreement about the underlying mechanisms (see e.g., DeKeyser & Larson-Hall, 2009).

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, 2011). 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 found that, despite adult-onset learners' extensive L2 exposure within naturalistic settings [2] and their often very advanced L2 knowledge,

---

[1] N.D. refers to one of the interviewees from a sub-corpus of oral history testimonies obtained from the USC Shoah Foundation – The Institute for Visual History and Education at USC (University of Southern California).

[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 which classroom learners receive.

child-onset L2 learners are nevertheless more likely to score within native-speaker ranges than adult-onset L2 learners. Altogether, AO appears to be a robust factor which explains L2 attainment across multiple linguistic domains.

Since age coincides with distinct life stages, resulting in different experiences, researchers have argued against maturational constraints by pointing to a range of experiential factors (Birdsong, 2006). Amongst those are the amount and quality of (L2) input and bilingualism effects, i.e. 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 socio-psychological factors such as motivation and attitude towards 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 influenced their oral production, in particular their levels of grammatical and lexical complexity.

Theoretical background

The present study applied the Complexity, Fluency, Accuracy (CAF) approach, most frequently used to assess beginning to advanced stages of SLA in classroom-based settings (Housen, Kuiken, & Vedder, 2012), to investigate advanced L2 proficiency. It thereby contributes to the small body of existing CAF research of 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 which has addressed late stages of L2 development generally falls under the umbrella terms of ‘ultimate attainment’ and ‘fossilization’ (for reviews see Birdsong & Molis, 2001; Birdsong, 2005, 2006; DeKeyser & Larson-Hall, 2009; Hyltenstam & Abrahamsson, 2008; Long, 2005; Muñoz & Singleton, 2011; Rothman, 2008). Most studies 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, p. 176), it states that maturational constraints guide (L2) language acquisition, which after a certain age does no longer take place automatically (implicitly) on the basis of mere exposure. Positive evidence for the CPH would be to find not a single L2 user who started acquiring the L2 after the critical period to perform within native range (Birdsong & Molis, 2001). Furthermore, 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 different points in time. They found a discontinuity in the declining slope first for L2 phonology between the age of 4 to 5, next for L2 lexis and collocations between the age of 9 to 10, and finally for L2 morphosyntax around the age of 12. However, a number of studies did not find clear patterns of discontinuity in the generally declining slope (e.g., Bialystok, 1997; Bialystok & Hakuta, 1995; Bialystok & Miller, 1999) or a great number of adult learners with nativelike proficiency (see however White & Genesee, 1996). Therefore, socio-psychological 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 reviewed studies which employed primarily linguistic, behavioural tests as opposed to reaction time experiments or brain imaging for assessing the acquisition of L2 syntax, morphosyntax, and the lexicon.

Studies on grammar (syntax and morphosyntax). Studies in favour of the CPH or maturational constraints have frequently employed grammatical judgment tests (GJTs) (DeKeyser, 2000; Johnson & Newport, 1989, 1991). Fewer studies conducted a wide array of tests 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, the correct grammatical gender, or the correct morphological ending of nouns (e.g., Abrahamsson & Hyltenstam, 2009; Granena & Long, 2013). A great number of studies of the L2 attainment of syntax and morphosyntax looked at L2 English speakers, but with different 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 an 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 five years and their age at testing (AaT) ranged from the ages of 20 to 50. All of these studies point to significant age effects as most adult-onset

L2 learner did not fall within the (monolingual) native speaker range. Abrahamsson and Hyltenstam (2009) even found only a few cases of child-onset learners who actually performed to this yardstick. Overall, a vast amount of evidence confirms AO effects on long-term migrants' L2 proficiency. Given that older immigrants appear to be less successful in acquiring a second language, many scholars have argued in favour of maturational constraints.

In addition to the studies reviewed thus far, there is also some evidence from investigations of productive grammatical complexity, in particular of the morphosyntactic characteristics of spontaneous or semi-spontaneous L2 productions which highlight the role of age of onset. 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 ranging from 1 to 3 years (university students), 5 to 15 years and 15 to 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 based on speech production strengthen the claim that nativelike performance does no longer seem to be possible for adult-onset learners in that domain.

Evidence against strong claims for the CPH comes from a series of studies which applied similar methodologies. L1 backgrounds of participants varied and included English, Dutch and Russian (Hopp, 2010), Chinese (Bialystok, 1997), Korean (Flege, Yeni-Komshian, & Liu, 1999), as well as 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 in terms of AOs and LoRs of the participants is comparable to those in studies arguing for maturational constraints. However, findings point 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. Such studies found that the observed age of onset effects were confounded with factors such as L1-L2 similarities (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 of L2 ultimate attainment by Dąbrowska (2012), who reviewed studies on adult monolingual speakers' grammatical knowledge.

Altogether, practically all studies report rather robust age effects on L2 ultimate attainment. But given the fact that some studies found nativelike performance and a continuous rather than a discontinuous slope, alternative (possibly age-confounding) variables such as L1 development and use, typological distance between the L1 and the L2, motivation and attitude by the learner should be considered. As for the data under investigation, we note that grammatical complexity in spontaneous speech has rarely been assessed in studies of ultimate L2 attainment. To get a more accurate picture of factors which affect grammatical complexity at the level of advanced L2 proficiency, we therefore investigated migrant L2 learners in a naturalistic SLA setting where age of emigration corresponds with AO.

Studies on semantics and the lexicon. Turning from (morpho) syntax to the lexical domain, it appears that evidence for and against maturational constraints is limited. Studies which have clearly argued for a CPH in the lexical domain are e.g., Coppieters, (1987), Abrahamsson

and Hyltenstam (2009), and Granena and Long (2013). The testing battery administered by Granena and Long (2013) to L2 speakers of Spanish included lexical tasks on multi-word correction and completion, a picture-guided narrative and a two-word preference task, all tapping into lexical proficiency. Abrahamsson and Hyltenstam (2009) also employed several gap-filling tasks to test lexical knowledge and formulaic sequences. In comparison, Coppieters' (1987) approach 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. Across these studies significant differences between the L2 participants and native speaker controls were found. This lack of nativelike performance in particular among the adult-onset learners was interpreted in favour of maturational constraints.

On the other hand, several studies provide evidence against definitive age effects on the acquisition of the L2 lexicon (Hellman, 2011; Marinova-Todd, 2003; Montrul & Slabakova, 2003). Montrul and Slabakova (2003) administered a truth-value judgment task and a sentence-conjunction task to 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 on both tasks. More than half of these participants had been grouped as near-native. In another study of L2 English, Hellman (2011) administered several receptive vocabulary knowledge tests including the Peabody Picture Vocabulary Test (PPVT) and a Self-Rated Vocabulary 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 level range. Nevertheless, 76 per cent of the adult-onset L2 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 (2003), who found some adult-onset learners to be indistinguishable from native controls on tasks tapping into the lexicosemantic domain. This finding led Hellman (2011: 177) to argue that adult-onset L2 learners can achieve a native level in their L2 vocabulary. In addition, Hellman (2011) found five exceptional late-onset learners who outperformed the native speakers on all three administered tasks. All of them turned out to be highly educated and to possess outstanding intellectual giftedness which raises the question whether they are outliers and should be excluded from the sample. All studies mentioned 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). Such 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 also stems from studies on spontaneous speech, suggesting that even adult-onset learners can attain nativelike

[3] They were all native speakers of English who either grew up in a US 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, p. 167).

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 to 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. Altogether, 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, for the lexical domain studies on advanced L2 proficiency are generally scarce, whether they test receptive or productive lexis. Some evidence suggests that AO might have an effect, in particular for receptive lexical skills, but clearly other factors must be considered such as LoR and level of education.

To conclude, 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 the earlier attested methodological limitations. As for example Long (2005) showed, there are a number of issues concerning these studies, above all their often narrow methodological scope with respect to the dominant use of GJTs for testing. Furthermore, as demonstrated above, very little is known about L2 grammatical and lexical complexity at the stages of advanced L2 proficiency when assessed productively. Along the lines of suggestions made by Long (2005), DeKeyser (2013) presents a list of criteria toward better sampling for future CPH studies, amongst which are:

- LoR > 10 years
- Sufficient variability in LoR
- Spread in socio-economic status (SES)
- AaT below 50
- Participants who have almost exclusively used the L2
- Etc.

The present study aimed at taking a step towards these suggestions by extending the scope to very advanced L2 speakers who spent the majority of their lives in an L2 environment and used the L2 almost exclusively. Instead of GJTs we analysed their spontaneous productions on the basis of oral history interviews to find out whether grammatical and lexical complexity in the L2 are related to different internal and external-experiential background variables, including age of onset, length of residence, age at testing, and level of education. We chose for a range in ages of onset from 7 to 17 years, since a discontinuous decline is expected to be found for L2 lexis and morphosyntax according to recent finding by Granena and Long (2013).

Research questions

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

1. How does age of onset affect the complexity of L2 syntax in spontaneous oral production? Are there any other variables which might explain the acquisition of L2 grammatical complexity?

2. How does age of onset affect the complexity of L2 lexical use in spontaneous oral production? Are there any other variables which might explain the acquisition of L2 lexical complexity?

Method

Interviewees. The study analysed 102 oral history testimonies which were given by German-Jewish immigrants in the USA, the UK 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 to 17 with a mean age of onset (AO, which corresponded to their age at emigration) of 12.5 years. On average, our interviewees' length of residence (LoR) in any of the three English-speaking host countries was 61.3 years (range 41-73). Table 1 shows that the greater part of their lives was spent in the L2 environment. At the time of testing (AaI which corresponds to AaT) interviewees were on average 73.6 years old (57-87). The majority of interviewees were female (n = 60). The level of continued exposure to German (L1 Exp) after emigration was assessed by three independent raters 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 English-speaking community (through studies, work, and/or extracurricular engagements). Interrater agreement for all pairs was $r > .7$. We considered the

[4] None of the interviewees spoke Yiddish. Some of them mentioned that they started taking English at school in preparation for their emigration.

median value as the final L1 Exp score to avoid the influence of outliers. Interviewees' average L1 Exp was 4.34. 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 concrete 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 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 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 kind of vocational training in case they did not finish high school (mid); 46 of the participants graduated from college or university (high). For 11 of the participants we were unable to identify their Edu due to lack of background information. These were excluded from the analyses. Table 1 gives an overview of the background variables.

[Table 1: Overview independent variables]

We refrained from using a control group. Since all the data used in the analyses are from interviewees who emigrated under severe circumstances, it was not feasible to find a comparable monolingual control group (see also Schmid, 2002, p. 78).

Oral history testimonies. The data consisted of 102 oral history testimonies (autobiographical interviews) in which our participants narrated their lives before and after

emigration. We acquired these testimonies from several sources[5], including libraries and archives in Germany, the UK, and the USA.

With the rise of the oral history method over the past few decades (Pavlenko, 2007) generally accepted standards for such interviews 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 to 10, 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 outbreak 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.

[5] The data were obtained from the following institutions: Werkstatt der Erinnerung (Hamburg, Germany), Alte Synagoge Essen (Germany), Prof. Manfred Brusten (Wuppertal, Germany), USC Shoah Foundation Institute (Los Angeles, USA), Fortunoff Video Archive for Holocaust Testimonies (Yale University Library), Tauber Holocaust Library and Education Program (San Francisco, USA), United States Holocaust Memorial Museum (Washington, D.C., USA), and the Association of Jewish Refugees (London, UK). We thank these institutions for their kind permission to use the interviews in our study.

Given the focus of such testimonies on personal histories, they are primarily considered a historical source providing historians with a window into people's personal past. On the other hand, these narratives also represent a rich source of spontaneous speech which may give us insight into patterns of discourse (Labov, 2013; Schiffrin, 2001) as well as into processes of language development such as L1 loss (i.e. L1 attrition) and L2 acquisition (Schmid, 2002, 2012).

Data generation. The first step was to establish detailed transcripts for the first 30 minutes of each interview based on extensive guidelines. These guidelines (available from on this website: http://www.let.rug.nl/languageattrition/tools) adhere to the transcription standards of CHAT (MacWhinney, 2000) and specify what belongs on the main utterance line, how to deal with anonymization of names, places etc., the transcription of compound words, phrases and collocations, as well as epistemic phrases, acronyms, incomplete, contracted and dialectal forms. The transcripts also contain 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 sub-clausal units with the option of subordinated clauses associated with either (Foster et al., 2000, pp. 365–66). This approach allowed for the inclusion of fragment-like independent units common in naturalistic speech. These fragment-like units may be phrases which can be elaborated to a full clause based on content (such as the phrase "Five years." in response to the question "For how long were you there?"), or they may simply be irregular or nonsentences as identified by Quirk, Greenbaum, Leech, and Svartvik (1985, pp. 838–53). Intonation and pauses served as additional indicators of where to break an utterance. In most cases utterances stretched across several lines, the so-called main tiers.

For generating syntactic measurements, the first 1800 words of pruned speech, i.e. excluding repeated, retraced, reformulated or incomplete material, were annotated based on detailed guidelines (available from the authors). The syntactic annotation below each main tier included information on whether the tier was the start of an utterance and AS-unit, whether it is a main or subordinated 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 non-finite verbs and the length of noun phrases. Where applicable, additional tags were included to capture the type of relative clause (object vs. subject relative clause) and passive constructions. The following example illustrates 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 type-token-ratio (TTR)-based lexical measurements was done for the first 30 minutes of speech using the English MOR grammar of CLAN (MacWhinney, 2000).

For the generation of the hypernymy measure of lexical sophistication, the Coh-Metrix software (Graesser, McNamara, Louwerse, & Cai, 2004) was run on approximately the first 2000 words per transcript (about 1600 words of pruned speech).

Measures of grammatical and lexical complexity. In consideration of current discussions 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., the L2 grammar and the L2 lexicon in the present study) 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

multi-dimensional construct of complexity which should be operationalized using multiple measurements.

To capture the multi-dimensionality of grammatical and lexical complexity (see Bulté & Housen, 2012 for an illustration), we selected several measures for each based on three criteria. (1) The measures should have little overlap (conceptually) and capture distinct dimensions, e.g., diversity (i.e., the variation in grammatical and lexical choices) and sophistication (i.e. the use of advanced grammatical constructions and lexical items). (2) The measures should be applicable to spoken language, i.e. 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. (3) Finally, they should tap into distinct proficiency spheres of the L2 speakers that are beyond the advanced stage of L2 development, where especially morphosyntactic measures have been demonstrated to be indicative of non-nativeness (Bartning, 2012). Since English does not make use of extensive inflectional morphology, and given that we were looking at highly advanced L2 speakers, we chose to consider voice morphology by looking at the frequency of passive[6] use. All in all, we generated a variety of measures to assess grammatical and lexical complexity.

We selected 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 non-finite adverbial dependent clauses (DCs) per AS-unit to capture the sub-clausal level and mean number

[6] We acknowledge that the passive is not a purely morphological measure. It is rather a morphosyntactic feature since its use also requires syntactic restructuring as explained by König and Gast (2009, p. 123).

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 it has been found that length of noun phrases increases, rather than the number of subordinate clauses (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 multiple 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 on lexical measures (Crossley, Salsbury, McNamara, & Jarvis, 2011; Jarvis, 2013b) we selected four measures. For diversity, a simple type-token ratio (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 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 with the infrequent lemmas used across the overall corpus. Hypernymy refers to semantic relationships between words (for nouns and verbs only), i.e. associations between hypernyms (superordinate words, e.g., entity or furniture) and hyponyms (subordinate words, e.g., 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[7]. Thus, we measured the degree of abstractness of nouns and verbs used in the oral productions. As Crossley, Salsbury, and McNamara (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).

Analyses: mixed-effects models. Since each interviewee received multiple complexity scores, we conducted linear mixed-effects regression modeling, using the *lme4* package (version 1.1-7) for *R* (version 3.1.1.), with interviewee as a random-effect factor 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 accounting for the variability in the effect a certain predictor has. 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 model of the data (Blankevoort et al., 2013). The consideration of random intercepts and slopes prevents being anti-conservative (i.e. reporting too high *p*-values; Baayen, 2008; Baayen, Davidson, & Bates, 2008).

[7] 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.

To perform mixed-effects regression modeling on the different complexity scores, we first *z*-transformed all grammatical and lexical measures. Using the reshape package (version 0.8.5) in *R*, we merged the scores into a grammatical complexity and a lexical complexity score which were each again *z*-transformed.

Results

Grammatical complexity. The total number of grammatical complexity scores in our dataset was 455 based on 91 interviewees. Table 2 shows the means, standard deviations, minimum and maximum values for each individual grammatical complexity measure.

[Table 2: Overview of grammatical measures]

Table 3 shows the coefficients and associated statistics of the fixed-effect factors and covariates of the final mixed-effects regression model obtained by using our exploratory analysis (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%).

[Table 3: Linear mixed-effect model of grammatical complexity scores]

This model shows that male interviewees (β = .528, t = 3.775) and participants with a higher education (β = .228, t = 2.272) obtained overall higher complexity scores. The other potentially confounding variables (i.e. AO, AaI, LoR, and L1 Exp) did not reach significance by themselves or in interaction with any other variable and were therefore not included in the model. 11 interviewees had to be excluded due to missing information on their level of education.

Table 4 gives 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 interviewee. The AIC decreased by at least 2 with the step-wise addition of each fixed effect

factor, suggesting that the inclusion of gender and level of education in the final model is warranted.

[Table 4: Goodness of fit of the fixed-effects of the model for grammatical complexity]

Model criticism revealed that the distribution of residuals was more or less normal, as illustrated in Figure 1.

[Figure 1: Distribution of residuals of grammatical complexity model]

We did not find a significant effect for AO. A graphical look at the data (see Figure 2) shows the *z*-transformed scores for the individual grammatical complexity measures in relation to AO. It reveals that higher grammatical complexity scores were found across all AOs. The highest scores were obtained by 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.

[Figure 2: The relationship between age of onset and grammatical complexity]

A closer examination of several 'outstanding' cases, i.e. interviewees with the highest and the lowest score, 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 ages of onset ranged from 9 to 12, their interview ages from 68 to 81, and their lengths of residence from 59 to 71. All three of them had a medium to high level of education and were found to have a high degree of continued L1 exposure, ranging from 5 to 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 ages of onset ranging from 14 to 16 with ages at interview ranging from 68 to 78, and lengths of residence ranging from 52 to 64 years. Two of them had obtained a low level of education, whereas one had obtained a high level of education. Their

continued exposure to German ranged from medium (3) to high frequency (7). In terms of profession they varied from one interviewee pursuing mainly housewife activities to another working for the United Nations. This qualitative close-up partly confirmed the statistical model which showed gender and level of education effects (Table 3).

Lexical complexity. The total number of lexical complexity scores in our dataset was 300 observations based on 75 interviewees. Table 5 shows the means, standard deviations, minimum and maximum values for each individual lexical complexity measure.

[Table 5: Overview of lexical measures]

Table 6 shows the coefficients and associated statistics of the fixed-effect factors and covariates of the final mixed-effects regression model obtained by using exploratory analysis (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%).

[Table 6: Linear mixed-effect model of lexical complexity scores]

This model shows that interviewees who were male (β = .394, t = 2.714), highly educated (β = .277, t = 2.696), and who used German at work (β = .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 by themselves or in interaction with any other variable and were therefore not included in the model. 27 interviewees had to be excluded due to missing information on their Edu and whether they used German at work.

Table 7 gives an overview of how log Likelihood and AIC values changed with the step-wise inclusion of the significant fixed-effect predictors while keeping the random-effects structure constant (see Wieling et al., 2011). The inclusion of each additional predictor resulted in

an AIC decrease of at least 2, warranting the final model with the three fixed effect factors: gender, level of education, and the use of German at work.

[Table 7: Goodness of fit of the fixed-effects of the model for grammatical complexity]

Model criticism revealed that the distribution of residuals was more or less normal, as illustrated in Figure 3.

[Figure 3: The distribution of residuals for lexical complexity model]

A qualitative look at the interviewees with the highest and the lowest lexical complexity scores (see Figure 4) shows that the three interviewees with the highest scores (for the ratio of content words and total words and TTR) were two male and one female. Their ages of onset ranged from 12 to 16 and their interview ages from 71 to 81. Their lengths of residence ranged from 59 to 66 years. The three interviewees were found to have had frequent continued L1 exposure (5 to 6) and they had obtained a medium to high level of education. Whereas two of them pursued academic careers as university lecturers, one worked as a professional photographer. The interviewees with the lowest score (for the ratio of content words and total words and TTR) were two male and one female with ages of onset ranging from 8 to 11. For one of the interviewees we have no further background information. The other two interviewees were 65 and 66 years old at the time of the interview and had resided in the L2 country for 55 and for 57 years respectively. Both had continued L1 exposure ranging from some (3) to a high (7) degree and they had obtained a medium to high level of education. One of them worked as a nurse whereas the other was an engineer.

[Figure 4: The relationship between age of onset and lexical complexity scores]

Discussion

Age of onset effect. Our findings do not lend support to the critical period hypothesis and maturational constraints. In fact, the analyses did not reveal an effect of age of onset (AO), neither for grammatical complexity nor for lexical complexity at the word level (see Figures 2 & 4). A graphical look at the relationship between AO and grammatical and lexical complexity scores revealed a similar distribution of scores across all AOs. A slightly deviating picture emerged with respect to the distribution of very low and very high scores. For 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 between 8 and 12. A slightly 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 the highest scores. A qualitative inspection of the interviewees with the highest and the lowest complexity scores reveals differences in age at the time of the interview and the professions which the interviewees pursued. Interestingly we find that the interviewees who scored highest on the lexical measures were older than those who scored lowest. This suggests that lexical growth takes place well into old age (e.g., Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). The qualitative inspection further suggests that a profession which requires extensive use of and exposure to language such as teaching might benefit lexical complexity.

Altogether, the lack of evidence for age effects in contrast to what previous studies found (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 length of residence (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 changes 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 look at the lexical complexity scores showed that interviewees with higher ages at emigration obtained some of the highest scores. What may be at play here is that the L1 lexicon had already developed substantially and might have served as a foundation on which to build and enrich the L2 lexicon. A larger pool of L1 lexical items could enable 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 exerted by the Nazi regime. Especially those interviewees who were already adolescents at the time will have consciously 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 the outbreak of WWII. Third, finding no age of onset effects may be due to the chosen age range. While we cannot exclude the possibility the window of opportunity closes before the chosen age range, studies particularly on morphosyntax have 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).

Alternatively, our interviewees may be resembling Bartning and colleagues' (2012) most advanced group of Swedish L2 learners of French. With regard to lexical measures they did not find this group to differ significantly from native speakers. However, there were persistent

morphosyntactic deviances. This might be for 'cross-linguistic' reasons, a potential factor pointed to by Erman et al. (2014). As already mentioned, English is morphologically poor. French on the other hand has a complex morphology which is likely to be challenging for any L2 learner of French. This would also confirm Schmid's (2014) results. That we did not find an age effect for our interviewees who were all L2 speakers of English could suggest that they do mirror Bartning and colleagues' most advanced group of Swedish L2 learners of French who did not differ significantly from native speakers. However, this conclusion remains highly speculative.

Instead of AO effects which would be specifically related to the L2 learning context, our findings of gender and level of education effects suggest patterns to be expected in a group of native speakers of both sexes and with different levels of education. Hence, our interviewees' L2 proficiency at this level is not related to variables specific to the L2 learning context such as AO, but rather to general variables which might affect language learning in general and performance of L1 and L2 speakers alike, such as gender and level of education.

Gender effect. 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. The qualitative inquiry 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. Reasons for this gender effect might be related to our data set and the population under investigation. It is conceivable that the men in our sample were more likely to be the provider of the family, as would have been common at the time. Being responsible for their families, men were likely to engage in the labour market more actively than women. In turn, they would have had more opportunities to converse and to connect to English-speaking co-workers;

hence, to integrate more easily than their female counterparts who took over the role of being housewives. 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 if they chose to stay at home might have been more restricted to family, neighbours and fellow mothers. However, there is no correlation between gender and level of education as well as pursuing a professional career for the interviewees in our data set, suggesting that women were equally well-educated as men and did pursue professional careers. Evidently, there is no clear indication for why men obtained significantly higher scores than women on the basis of level of education and interviewees' careers. Our findings require further investigation. Immigrant communities where men traditionally hold the role of the provider may help to shed light on our interpretations.

Level of education effect. 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 factor according to Birdsong's (2006) classification, or a socio-psychological factor. 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 the SLA context but have been demonstrated for L1 proficiency, for L1 grammar (Dąbrowska, 2012; Dąbrowska & Street, 2006) as well as the L1 lexicon (Mulder & Hulstijn, 2011). Since we find this effect amongst L2 speakers while there is no significant effect of AO, we tentatively conclude that our interviewees show similar patterns as did the native speakers in the above mentioned studies. Level of education appears to have a positive effect on both L2 grammar and L2 lexical knowledge.

A qualitative view on our data revealed that those interviewees with the highest scores for grammatical and lexical complexity had not only obtained a higher level of education, but they were also professionally accomplished mainly as business men or academics. The type of work 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 at work. 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 asks for speculative explanations as there are no studies on complexity which have addressed this particular relationship according to our knowledge. However, as Wolter (2006) suggests, the L1 lexicon can be both beneficial and hindering to the acquisition of an L2 lexicon. In our case, for lexical complexity there appears to be a beneficial effect of bilingualism, i.e. the 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, the continuous learning and use of the first language at work might also encourage the L2 speaker to expand his L2 vocabulary. Furthermore, the extensive linguistic experience of bilinguals probably enables them to use language much more creatively, allowing for more diverse productions. However, these findings require further investigation.

Conclusion

In this study we assessed the grammatical and lexical complexity of spontaneous oral productions by long-term L2 speakers and how they are affected by age of onset, length of

residence, 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 modelling we assessed which variables would predict the various measurements. Considering previous research arguing against the CPH we also took into account several experiential factors including the level of education and the use of German at work. The analyses showed that syntactic and lexical complexity at the productive level were not affected by variables which would indicate a critical period or maturational constraints. Instead, the experiential factor level of education played a significant role, as would be expected also for native speakers. It seems thus that any possible maturational effects in our data are superseded by a general, non SLA-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.

There are several limitations to the study. First, there may be variation in terms of the trauma our interviewees underwent as a result of the persecution of Jews in Europe. Unfortunately, precise clinical details were not available to us and could not be controlled for. In consideration of this particular characteristic and their experience of emigration in general, we refrained from a monolingual, native speaker control group comparisons. Future studies could attempt to find suitable control groups with similar traumatic refugee experiences. For example, a group of native speakers such as war veterans who have likely suffered from traumatic experiences could be considered. However, it is questionable to what extent their trauma would be comparable and whether it would affect their language production in a similar manner.

Furthermore, such war veterans would have probably been slightly older when they went to war, as compared to our interviewees who were mostly children or young adolescents when they were confronted with anti-Semitism. The difficulty of finding a control group underlines the special characteristics of our interviewees due to which we can only draw tentative conclusions regarding their nativelikeness.

A second limitation concerns the rather high age of our interviewees at the time of the interview. As Long (2005) has suggested, 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. It has been shown also for native speakers that syntactic complexity in their diary entries decreased with increasing age (e.g., Kemper, 1987). On the other hand, it has been argued that lexical growth continues (e.g., Ramscar et al, 2014).

A third 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 choice of measures was limited and their suitability for capturing age of onset effects and nativelikeness is debatable. While we did make an effort to capture the multi-dimensionality 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 e.g., sentence/utterance type, additional morphological measures such as the number of exponents (the number of 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 endeavour and should be further explored in future studies on advanced L2 proficiency. Finally, our findings are constrained to the areas of grammatical and lexical complexity. There are of course other types of linguistic complexity (e.g., phonological, discourse, etc.) which one could consider to look at as well as other dimensions proposed by the CAF framework, i.e. accuracy and fluency. As Abrahamsson and Hyltenstam (2009) found, listener perception tapping into the dimension of phonological accuracy in combination with a series of measures of linguistic performance, representation, and processing promise a more thorough investigation on the advanced L2 learner. Adding fluency and accuracy will yield a more complete picture of advanced L2 proficiency. Nevertheless, we hope to have demonstrated that advanced levels of L2 productive proficiency are attainable in the areas of grammar and lexis, independently of someone's age of onset.

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*(02), 187–214. http://doi.org/10.1017/S0272263112000022

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306. http://doi.org/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*(6), 716–723. http://doi.org/10.1109/TAC.1974.1100705

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R* (1st ed.). Cambridge, UK ; New York: 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*(4), 390–412. http://doi.org/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*(1), 185–211. http://doi.org/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*(02), 243–267. http://doi.org/10.1017/S0272263112000046

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, *13*(2), 116 –137. http://doi.org/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*(02), 127–145.

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 319–328.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*(s1), 9–49.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249. http://doi.org/10.1006/jmla.2000.2750

Blankevoort, C. G., Scherder, E. J. A., Wieling, M. B., Hortobágyi, T., Brouwer, W. H., Geuze, R. H., & van Heuvelen, M. J. G. (2013). Physical predictors of cognitive performance in healthy older adults: A cross-sectional analysis. *PLoS ONE*, *8*(7), e70799. http://doi.org/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), 1–32. http://doi.org/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 Publishing.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, *63*(3), 544–573. http://doi.org/10.2307/415005

Crossley, S. A., Salsbury, T. L., McNamara, D., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*(4), 561–580. http://doi.org/10.1177/0265532210378031

Crossley, S. A., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, *59*(2), 307–334. http://doi.org/10.1111/j.1467-9922.2009.00508.x

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253. http://doi.org/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*(6), 604–615. http://doi.org/10.1016/j.langsci.2005.11.014

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(04), 499–533. http://doi.org/http://dx.doi.org.proxy-ub.rug.nl/

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

DeKeyser, R., & Larson-Hall, J. (2009). 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: Oxford University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 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*. http://doi.org/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*(1), 29–62.

Fellbaum, C. (2013). WordNet. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–8). Chichester: Blackwell Publishing Ltd. 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*(1), 78–104. http://doi.org/10.1006/jmla.1999.2638

Forsberg, F. (2010). Using conventional sequences in L2 French. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(1), 25-51. http://doi.org/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*(2), 255–280.

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*(1), 73–92.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics*, *21*(3), 354 –375. http://doi.org/10.1093/applin/21.3.354

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*(2), 193–202. http://doi.org/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*(3), 311–343. http://doi.org/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*(2), 162–182. http://doi.org/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*(4), 901–931. http://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H., & Schmid, M. S. (2011). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, 1–34. http://doi.org/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 Publishing.

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: Blackwell Publishing Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470756492.ch17/summary

Hyltenstam, K., & Abrahamsson, N. (2012). Introduction. *Studies in Second Language Acquisition*, *34*(02), 177–186. http://doi.org/10.1017/S0272263112000010

Jarvis, S. (2013a). Capturing the diversity in lexical diversity. *Language Learning*, *63*(s1), 87–106. http://doi.org/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 Publishing.

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*(1), 60–99. http://doi.org/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* 1. *Cognition*, *39*(3), 215–258.

Kemper, S. (1987). Life-span changes in syntactic complexity. *Journal of Gerontology*, *42*(3), 323–328. http://doi.org/10.1093/geronj/42.3.323

Kemper, S., Rash, S., Kynette, D., & Norman, S. (1990). Telling stories: The structure of adults' narratives. *European Journal of Cognitive Psychology*, *2*(3), 205–228. http://doi.org/10.1080/09541449008406205

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

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

Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, *50*(4), 675–724. http://doi.org/10.1111/0023-8333.00143

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

Long, M. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 287–317.

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*(04), 588–595. http://doi.org/10.1017/S1366728911000010

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

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85–104. http://doi.org/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, MA.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, *34*(1), 9–34.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL - International Review of Applied Linguistics in Language Teaching*, *45*(3), 241–259.

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers. *Studies in Second Language Acquisition*, *25*(03), 351–398. http://doi.org/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*(5), 475 –494. http://doi.org/10.1093/applin/amr016

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, *44*(01), 1–35.

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

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: Walter de Gruyter.

Pallotti, G. (2014). A simple view of linguistic complexity. *Second Language Research*, 0267658314536435. http://doi.org/10.1177/0267658314536435

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language 1. *Language Learning*, *30*(2), 449–468. http://doi.org/10.1111/j.1467-1770.1980.tb00328.x

Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics, 28*(2), 163-188. http:// doi.org/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.

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*(1), 5–42. http://doi.org/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*(6), 1063–1088. http://doi.org/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, D.C.: Georgetown University Press.

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

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 Publishing.

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*(2), 177–208. http://doi.org/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*(4), 386–410. http://doi.org/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*(1), 125 –160. http://doi.org/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*(4), 671–692. http://doi.org/10.1111/j.1467-9922.2006.00392.x

Van Boxtel, S. van, Bongaerts, T., & Coppen, P.-A. (2003). Native-like attainment in L2 syntax. *EUROSLA Yearbook*, *3*(1), 157–181.

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*(3), 239–263. http://doi.org/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*(3), 233 –265. http://doi.org/10.1177/026765839601200301

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613. http://doi.org/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*(4), 741–747. http://doi.org/10.1093/applin/aml036

Table 1

Overview independent variables

	n	Mean (*SD*)	Range/Categories		
AO	101	12.15 (2.6)	7-17		
LoR	100	61.33 (6.12)	41-73		
AaI	101	73.59 (6.97)	57-87		
L1 Exp	98	4.34 (1.44)	1-7		
L1 at work	80		Yes: 14	No: 66	
Gender	102		M: 42	F: 60	
Edu	91		Low: 11	Mid: 34	High: 46

Note: AO = age of onset, LoR = length of residence, AaI = age at interview, L1 Exp = continued L1 exposure, L1 at work = use of German at work, M = male, F = female, Edu = level of education.

Table 2

Overview of grammatical measures

Dimension	Type	Level	Statistical construct	Mean (*SD*)	Min	Max
Diversity	Syntactic	Sentence & clause	words per AS-unit	9.47 (1.34)	6.18	12.53
		Sub-clause	DCs per AS-unit	0.48 (0.13)	0.20	0.83
			non-finite adverbial DCs per AS	0.09 (0.05)	0.01	0.26
		Phrase	words per NP	1.79 (0.15)	1.51	2.12
Sophistication	Morphological		passives per clause	0.05 (0.02)	0.01	0.14

Note: AS-unit = analysis of speech unit, DC = dependent clause, NP = noun phrase. Here we show the original values before any necessary transformations were performed.

Table 3

Linear mixed-effect model of grammatical complexity scores

Fixed effect	Estimate	Std. Error	*t*-value[1]	*p*-value
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*

Table 4

Goodness of fit of the fixed-effects of the model for grammatical complexity

Additional fixed effects	Log – likelihood increase	AIC decrease	Evidence ratio	Likelihood ratio test	Additional degrees of freedom
Random intercept only					
+ Gender male (vs. female)	5.63	9.26	102.56	$p < .001$	1
+ Level of education	72.85	143.70	>1000	$p < .0001$	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.

Table 5

Overview of lexical measures

Dimension	Lexical level	Statistical construct	Mean (*SD*)	Min	Max
Diversity	All words (number & range)	Type-token ratio (TTR)[a]	0.27 (0.02)	0.22	0.33
Density		Content words/total words	0.32 (0.02)	0.25	0.38
Sophistication	Infrequent content words	Frequency bands	0.20 (0.05)	0.1	0.32
	Selected content words (interrelatedness)	Hypernymy mean scores	1.42 (0.11)	1.16	1.71

Note: Frequency bands were calculated for nouns, verbs, and adjectives only. Mean hypernymy was calculated for nouns and verbs only. Here we show the original values before any necessary transformations were performed.

[a] Given that we controlled for text length, we decided to generate a simple TTR for lexical diversity. More advanced TTRs have been suggested over the years in response to the influence of text length on this measure (see e.g., Malvern & Richards, 2002).

Table 6

Linear mixed-effect model of lexical complexity scores

Fixed effect	Estimate	Std. Error	*t*-value	*p*-value
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**

Table 7

Goodness of fit of the fixed-effects of the model for grammatical complexity

Additional fixed effects	Log – likelihood increase	AIC decrease	Evidence ratio	Likelihood ratio test	Additional degrees of freedom
Random intercept only					
+ Gender male (vs. female)	3.958	5.91	19.25	$p < .01$	1
+ Level of education	59.83	117.66	>1000	$p < .0001$	1
+ German at work	94.71	187.42	>1000	$p < .0001$	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.