Measuring complexity in long-term L2 speakers of English and L1 attriters of German

Lahmann, Cornelia; Steinkrauss, Rasmus; Schmid, Monika S.

Published in:
International Journal of Applied Linguistics

DOI:
10.1111/ijal.12259

Publication date:
2019

Received: 8 May 2017 | Revised: 2 March 2019 | Accepted: 24 March 2019

DOI: 10.1111/ijal.12259

SPECIAL ISSUE ARTICLE

WILEY

Measuring linguistic complexity in long-term L2 speakers of English and L1 attriters of German

Cornelia Lahmann[1] | Rasmus Steinkrauss[2] | Monika S. Schmid[3]

[1] University of Mannheim, Germany

[2] University of Groningen, The Netherlands

[3] University of Essex, United Kingdom

Correspondence
Rasmus Steinkrauss, University of Groningen, The Netherlands.
Email: r.g.a.steinkrauss@rug.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 360-70-420

Linguistic complexity is neither easily defined nor measured. The challenge in finding reliable ways to measure linguistic complexity is even more pronounced when the variation of contexts in which complexity is measured is taken into account. This paper therefore aims at finding measures for assessing syntactic and lexical complexity that are sensitive and non-overlapping even in a less frequently studied context: spontaneous speech of L1 attriters and highly proficient L2 English speakers. To evaluate operationalizations of syntactic and lexical complexity, several previously proposed measures are applied to a corpus of spontaneous speech and a principal component analysis is conducted. The results show which of the measures group together as underlying variables of syntactic and lexical complexity and how this can inform future operationalizations of complexity.

KEYWORDS

L1 attrition, L2 acquisition, lexical complexity, syntactic complexity

Linguistische Komplexität ist sowohl schwer zu definieren als zu messen. Die Schwierigkeit, zuverlässige Maße für Komplexität zu finden, zeigt sich besonders dann, wenn versucht wird, auch die Kontexte, in denen Komplexität gemessen wird, in den Messungen zu berücksichtigen. Der vorliegende Artikel untersucht daher, welche Maße syntaktische und lexikalische

Komplexität in einem selten erforschten Kontext—Spontansprache sehr fortgeschrittener L2-SprecherInnen und MuttersprachlerInnen, die unter Erstsprachverlust leiden—zielgenau und verlässlich abbilden können. Um unterschiedliche Operationalisierungen syntaktischer und lexikalischer Komplexität zu vergleichen, werden verschiedene in der Literatur bisher verwendete Maße auf einem Korpus von Spontansprache angewandt und mit Hilfe einer Hauptkomponentenanalyse ausgewertet. Die Resultate zeigen, welche Maße gemeinsam die zugrundeliegenden Faktoren syntaktischer und lexikalischer Komplexität in aussagekräftiger Weise erfassen können. Die Resultate können zukünftige Operationalisierungen linguistischer Komplexität erleichtern.

SCHLÜSSELWÖRTER

L1 Sprachverlust, L2 Erwerb, Lexikalische Komplexität, Syntaktische Komplexität

1 | INTRODUCTION

The concept of complexity is a difficult and elusive one, as is evident from the often vague definitions that are offered for it. For example, in the Oxford Dictionary of English the noun "complexity" is defined as a state or quality of being "intricate" or "complicated." Other definitions associate complexity with terms like complicatedness, difficulty, or cost. The interrelatedness of these terms makes it hard to define complexity clearly. Additional difficulty arises due to the multidimensionality of the construct, i.e., an object's complexity is composed of multiple dimensions. Language for instance is complex at various levels (e.g., at the level of the word, the phrase, or the sentence) and with respect to different linguistic aspects (grammatically, lexically, phonologically, etc.).

This lack of clarity of the complexity construct with respect to its definition and operationalization as a multidimensional construct and the purposes it serves is also reflected in studies on, for example, the complexity of L2 productions. First, the numerous ways of how complexity has been measured across studies makes it difficult to compare them. Second, despite these many ways of operationalization, L2 studies may capture complexity incompletely by not tapping sufficiently into its multiple sub-dimensions. Third, valid operationalizations of complexity should be applicable to a variety of contexts where measuring complexity is of interest. Hence, there is a need for a theoretical renewal of the construct (Ortega, 2012).

In this paper, we aim at testing recent proposals for refining the complexity construct for the purposes of studying L2 productions. More specifically our goal is to investigate whether the analytic nature of complexity as outlined in Bulté and Housen's (2012) proposal can be empirically attested when complexity is measured in a context that is not frequently investigated: a corpus of spontaneous speech of highly advanced bilinguals. The application of the complexity construct to such a dataset is especially taxing for the following two reasons: first, most studies investigating complexity in L2 productions focus on written, not spoken samples (Ortega, 2012). Second, most studies applying measures of complexity target intermediate to advanced L2 learners, not highly proficient speakers. However, the contexts where linguistic complexity might be measured are much more varied. Therefore, evaluating the recent

proposals on the definition and operationalization of linguistic complexity on spoken productions of highly proficient bilinguals may show the limits of these proposals, and highlight the issues that need to be considered when assessing complexity in a range of varied contexts.

We will start our investigation by summarizing Bulté and Housen's (2012) approach and discussing it in relation to other approaches to complexity in the SLA literature. Based on this discussion, we will select specific measures of complexity and apply them to our corpus to test whether they empirically support the analytic view of complexity according to Bulté and Housen's (2012, 2014) taxonomies of grammatical and lexical complexity.

2 | THEORETICAL BACKGROUND

In an attempt to define and renew the complexity construct, Bulté and Housen (2012, 2014) (see also Bulté, 2013) take an analytic view and propose hierarchical taxonomies for grammatical and for lexical complexity, which reflect how the two constructs have been conceived in the L2 literature. In these taxonomies a distinction is made between a theoretical, an observational and an operational level.

At the theoretical level, the authors (Bulté & Housen, 2012, 2014) distinguish system complexity from structure complexity. System complexity pertains to the composition and organization of a specific L2 system, for example, the range of different words or grammatical structures a learner knows, while structure complexity pertains to the (learner-independent) complexity of the linguistic items, structures or rules that make up the learner's L2 system, for example, the degree of sophistication of the learner's lexical and grammatical knowledge.

The outlines for the observational and operational level differ between grammatical and lexical complexity, whereby grammatical complexity includes both syntactic and morphological complexity. Regarding grammatical complexity, a distinction is made between diversity and sophistication, each with respect to both syntactic and morphological complexity. The multidimensionality becomes evident as syntactic complexity is subdivided into sentential, clausal and phrasal complexity, whereas morphological complexity is subdivided into inflectional and derivational complexity. For lexical complexity a distinction between density, i.e., the relative amount of lexical words to function words, diversity, compositionality and sophistication, i.e., the relative amount of infrequent words is drawn.

These observational constructs translate into several measures illustrated at the operational level, including measures of length and a subclause ratio as well as the number of syntactic arguments per clause for capturing syntactic complexity. For morphological complexity measures include the number of dependents per phrasal head, the number of verb inflections per verb, the number of verb forms such as the frequency of tensed forms and modals, and the number of derivational affixes per word. For lexical complexity, measures include the lexical word function words ratio and various type-token ratios among others. Altogether these measures are meant to capture grammatical and lexical complexity as a multi-dimensional construct.

With this, Bulté and Housen (2012) offer a fine-grained picture of the complexity construct in structural terms, from its definition all the way to its operationalization. However, the measures may tap into the same sub-constructs. These issues will be taken up again after reviewing other, related proposals to define and operationalize complexity.

In these proposals, three different strands may be distinguished. First, there is a recent approach that Pallotti (2014) terms the “simple view” and which is narrower than Bulté and Housen's (2012) proposal discussed above, second, there are analytic approaches which are compatible with Bulté and Housen's (2012) proposal but focus only on one specific aspect of complexity, and third, there are more holistic views of complexity which stress the interrelatedness of complexity with the other two dimensions of the CAF triad (accuracy and fluency) and the importance of contextual factors (e.g., modality and audience).

In contrast to Bulté and Housen's (2012) analytic view, Pallotti's (2014) “simple view” is interested in the complexity of a text only, not in that of the learner system which is affected by processing costs, proficiency and development. Accordingly, Palotti suggests to limit investigations of complexity to a restricted set of measurements

relating to the structure of the linguistic system only, such as the number of forms taken by lexemes to express grammatical information.

Complementary to Pallotti's simple view are proposals that elaborate a sub-aspect of an analytic view of complexity. Two proposals which figure prominently in the literature focus on lexical complexity: Jarvis (2013a, 2013b) explores the various properties of lexical diversity, while Crossley, Salsbury, McNamara, and Jarvis (2011a, 2011b) concentrate on lexical sophistication. Both suggest several measures such as, for diversity, the total number of lexemes in a text and the use of infrequent words, and for sophistication, the hypernymy (relationship between subordinate and superordinate words) and familiarity of words,[1] among others. The two proposals thus support an analytic view of complexity. At the same time, the proposed fine-grained measures run the risk of overlapping with each other: frequency-based measures, for example, have traditionally been viewed as a measure for lexical sophistication, not diversity.

Finally, holistic approaches to complexity either underline the importance of taking contextual factors into account (Ortega, 2012), or point at the interrelatedness of complexity with accuracy and fluency as suggested in previous CAF studies (Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012; Osborne, 2011, 2013; Skehan, 2009). Contextual factors may include topic, register, audience, modality and genre (Ortega, 2012; Rimmer, 2006, 2008), and Bartning, Forsberg Lundell, and Hancock (2012) stress the importance of relating complexity to accuracy and fluency for a better understanding of advanced/highly proficient L2 users.

In summary, we have identified three perspectives on complexity within the CAF framework: the simple, the analytic, and the holistic view. The purpose of this study is to identify and test measures of complexity that are valid even in a less frequently investigated setting. The holistic views are informative for this study because they argue for taking context and proficiency level into consideration, i.e., addressing the variation in contexts where complexity is measured. Pallotti's proposal on the other hand seems too narrow for the current purpose as disregarding measures of sophistication might not enable us to capture differences in advanced levels of proficiency. Therefore, a broader, analytical approach to complexity as advocated by Bulté and Housen (2012) will be taken in this study.

Each of the three approaches addresses various shortcomings with respect to definition and operationalization of complexity in SLA studies. Four issues are relevant in the current context: (i) the overlap of measures; (ii) and the variation in contexts investigated, more specifically the language mode studied; (iii) the assessment of advanced proficiency levels; and (iv) the prevalence of English data. Given the goal of the current study to select several measures which capture the multidimensionality of complexity and then test this multidimensionality by assessing oral complexity in two groups of advanced bilinguals, one issue to overcome (that this study shares with any study investigating linguistic complexity) is the problem of overlapping measures tapping into the same dimensions. For grammatical complexity, not only length and subordination should not be measured (see Bulté & Housen, 2012; Šišková, 2012). Instead, morphological measures as well as measures capturing specific types of syntactic constructions—depending on the language under investigation—may be added. For the domain of lexical complexity, a careful selection of the measures suggested by Jarvis (2013a, 2013b) and Kyle and Crossley (2015) might prevent overlap and offer the additional potential to account for different modalities and different proficiency levels and, hence, varied contexts.

Another issue is the prevalence of studies analyzing written output (Ortega, 2012). Few CAF studies have assessed speech, let alone spontaneous speech in a natural setting. This might result in different kinds of complexification strategies than we would expect for speech or writing under more controlled conditions. Therefore, more finely tuned measures of complexity should be chosen to capture complexity in varied contexts. Measures targeting different ways of complexifying sentences or utterances could be relevant here, as subordination in speech is usually achieved by means of finite dependent clauses (Beaman, 1984; Biber, Gray, & Poonpon, 2011), whereas in writing noun phrase elaboration can be observed (Michel, Kuiken, & Vedder, 2007; Ortega, 2003). Finding the latter characteristic in spontaneous speech would imply an increase in complexity. Also, due to the shared context of speech and the pressure of real time processing, there is an avoidance of the specification of reference, i.e., the subject is reduced to a pronoun or even dropped (Leech, 2000: 695). Measuring the length of subject noun phrases might therefore be a way to discriminate between different levels of complexity in speech. Finally, for lexical complexity,

studies show measures of sophistication such as frequency, concreteness, imageability, meaningfulness and polysemy to be closely related to speech (Louwerse, McCarthy, McNamara, & Graesser, 2004).

The third issue pertains to the assessment of complexity in productions by highly advanced bilinguals as opposed to beginning, intermediate, and advanced L2 learners which are usually covered by CAF studies. To our knowledge only a few CAF investigations look at speech productions by highly proficient L2 speakers who have been immersed in L2 naturalistic settings for an extensive period (e.g., Bartning, Forsberg, & Hancock, 2009; Forsberg, 2010; Bartning et al., 2012; Forsberg Lundell & Lindqvist, 2012). Therefore, a selection of measures that previous studies have not employed widely might be necessary and morphosyntactic measures as well as measures targeting multi-word units may help to address the issue of varied contexts and to distinguish amongst speakers with advanced language proficiency levels.

The fourth and final issue concerns the language(s) under investigation. The majority of CAF studies focus on English as an L2. However, languages differ typologically and, hence, different measures may have to be used for capturing advanced language proficiency levels in different languages (Brezina & Pallotti, 2016). In studies of L1 attrition of German, syntactic measures of word order and of specific syntactic constructions in German[2] were applied (M.S. Schmid, 2002, 2012) since German requires the verb in second position for main clauses and in final position for subordinate clauses. Due to its grammatical case and gender system, German is also morphologically more complex than English. Therefore, it is of interest to take a comparative perspective on complexity measures, applying them to both German as an L1 in the context of attrition and English in the context of L2 acquisition.

3 | RESEARCH QUESTION

The present study aims at characterizing learners in terms of linguistic complexity and identifying the internal structure of the complexity construct, in particular syntactic, morphological and lexical complexity, on the basis of evaluating a range of measures. In doing so, we hope to contribute insights into the multidimensionality of complexity. The following research question will be addressed:

RQ: To what extent does the analytic view of complexity hold when looking at a corpus of spontaneous speech by highly advanced bilinguals?

Based on the literature reviewed above, we expected that a combination of different measures of complexity is able to tap into different aspects of complexity, confirming a multidimensional, analytic view. The rationale for focusing on the analytic view is the nature of the data. As we set out to assess oral proficiency by means of applying various complexity measures, the question arises how to measure complexity in sufficient detail so as to find any variation among the generally proficient speakers of L1 German and L2 English. The analytic view provides the most fine-grained picture of complexity.

4 | METHOD

4.1 | Participants

The data of this study consist of transcripts of oral interviews from two distinct groups of highly advanced German-English bilinguals. All participants were L1 speakers of German who had emigrated to an English-speaking country where they predominantly spoke English as their L2, and they were interviewed either in German or in English. Those interviews given in German were used to investigate possible L1 attrition, while those interviews given in English were used to investigate L2 knowledge. For the purposes of the current study, the participants will be referred to as L1 German attriters or L2 English speakers respectively, depending on the language of the interview. In total, the corpus consists of 73 L1 German attriters (L1A) and 102 L2 English speakers (L2S) with German as their L1.[3] Both

groups share a German-Jewish background. All participants had arrived in the UK, the USA, and Australia as children and adolescents between the ages of 7 to 17 (their age at emigration corresponds to their age of onset, henceforth: AO). They had all been forced to escape the Nazi regime at some point between the pogrom on the night of the 9th November 1938 and the outbreak of WWII in September 1939. Their emigration took place either unaccompanied or together with their close family members including parents and siblings.

While our participants' early biographies in Germany are rather homogeneous and mostly set against a bourgeois, middle-class background, we find more variation after emigration. Upon arrival in their host countries, some could complete their education whereas others needed to enter the labor market immediately. This led to some variation especially with respect to the participants' educational and professional development. About half of them completed university (high Edu) and eventually became managers, doctors, and university professors. The other half completed a professional/vocational training (mid Edu) and mostly took on manual jobs. Several of our interviewees did not finish their formal education (low Edu).

Depending on their circumstances, even after emigration our participants were still exposed to their L1 German, but to varying degrees. To establish this variable (L1 exposure: L1 Exp), three independent raters were asked to score each participant's quantity and quality of L1 exposure on a scale from 1 (low) to 7 (high). Raters based their final score on the occurrence of statements regarding (a) avoidance of speaking German, (b) manner of emigration (adopted or not), (c) contact with family members, (d) origin of marital partner (German-speaking, English-speaking, or other), (e) continued use of German, and (f) integration into English-speaking community (through studies, work, or extracurricular engagements). Ratings were given on a scale from 1 (little exposure) to 7 (much exposure). Inter-rater agreement for all pairs was $r > 0.7$. The median value was taken as the final L1 exposure score to avoid the influence of outliers.

At the time of the interview L2Ss and L1As were somewhere between 57 to 87 years old (age at interview: AaI). Thus, they had spent most of their lives in the L2 community with lengths of residence (LoR) ranging from 41 to 73 years. For a detailed summary of these variables for each group see Table 1.

No control group was taken into consideration due to the practical limitations of finding a comparable control group that would have experienced similarly severe conditions (see Lahmann, Steinkrauss, & M.S. Schmid, 2016).

TABLE 1 Overview of the corpus

	M (*SD*)		Range/Categories					
	L2S (n = 102)	*L1A (n = 73)*	*L2S (n = 102)*			*L1A (n = 73)*		
AO[a]	12.15 (2.67)	13.8 (2.08)	7–17			7–17		
LoR[b]	61.33 (6.12)	60.3 (5.5)	41–73			46–70		
AaI[c]	73.59 (6.97)	74.08 (6.14)	57–87			58–85		
L1 Exp[d]	4.34 (1.44)	4.62 (1.49)	1–7			1–7		
Gender			M: 42	F: 60		M: 38	F: 35	
Edu[e]			Low: 11	Mid: 34	High: 46	Low: 9	Mid: 22	High: 31

[a]AO = age of acquisition;

[b]LoR = Length of residence;

[c]AaI = Age at interview;

[d]L1 Exp = First language exposure;

[e]Edu = Level of education (numbers do not add up due to missing data)

4.2 | Data

4.2.1 | The corpus

The dataset consists of a corpus of 175 oral history interviews, of which 102 are in English given by the group of L2Ss and 73 are in German given by the group of L1As. More specifically, the data we are looking at are Holocaust oral histories which were audio- or video-recorded. While these data are usually a source for historians, for linguists they offer a wealth of spontaneous language (Schiffrin, 2001; M. S. Schmid, 2002). For more information regarding the current corpus, see Lahmann et al. (2016).

4.2.2 | Data processing procedure

Data generation took place in three steps: transcription, annotation, and generation of raw data. Each interview was transcribed until a total length of about 1,800 words was reached. All transcriptions were in CHAT format (see MacWhinney, 2000) according to detailed guidelines (see https://languageattrition.org/resources-for-researchers/coding-your-data), yielding a corpus of 185,400 words in English and 134,000 words in German. Except for three German transcripts which were 1,300, 1,355 and 1,667 words long due to the overall shorter length of the interview, each transcript (per speaker) was on average 1,800 words long excluding repeated, retraced or reformulated passages/materials. All transcripts were prepared, checked and annotated by native and near-native transcribers of the respective languages.

For transcription, utterances were identified based on intonation and pauses. An utterance could consist of several clauses with associated subordinate clauses. Within utterances, AS-units (an independent or sub-clausal unit plus any subordinated clauses associated with either; Foster, Tonkyn, & Wigglesworth, 2000) were marked. All transcripts adhere to the minimum standard for CHAT files (MacWhinney, 2000).

During the second step all interviews were manually coded for aspects of grammatical complexity. Lemmatization and part-of-speech tagging was done automatically using the free TreeTagger software developed by the University of Stuttgart (H. Schmid, 1995) for the German data or the mor-program in CLAN (MacWhinney, 2000) for the English data.

In a final step we counted the grammatical and morphological/lexical information encoded in the tagging. Calculations were done by using dedicated macros written in the scripting language of Microsoft Word, VBA, and by the CLAN software. From these data the measurements were calculated.

4.3 | Measures

The choice of measures was informed by the analytic view of complexity and the four considerations mentioned above: (1) overlapping measures; (2) the assessment of spontaneous speech produced by (3) highly advanced bilinguals; and (4) the languages under consideration. A variety of measures were retained which have been hypothesized to capture complexity at different levels of grammar and the lexicon. This results in several measures capturing utterance complexity, namely at the utterance/sentence, clausal and phrasal level.

Table 2 gives an overview of the syntactic complexity measures used. Three measures cover the level of sentence and clausal complexity. All three are overall measures of length. As Norris and Ortega (2003) pointed out, these measures are rather crude indicators of complexity and they do not inform us of the type of complexification. Therefore, several measures were added to explore the clausal and the phrasal level further.

For morphological complexity, several measures of inflectional and derivational morphology were used to capture both grammatical diversity and sophistication. These are partly language specific: for the L1 German data, the number of subjunctives and complex verb phrases (VPs), i.e., verb phrases consisting of one finite and two or more non-finite verb forms, were calculated. Table 3 presents an overview of these measures.

TABLE 2 Overview of syntactic complexity measures

Syntactic level	Statistical construct
Sentence & clause	*words per utterance (WD_UTT)* *words per AS-unit (WD_AS)* *words per clause (WD_C)*
Sub-clause	*dependent clauses (DCs) per AS-unit (DC_AS)* *non-finite adverbial DCs per AS (for English) (NONFDC_AS)* *object subject relative clause ratio (OBJREL_SUBREL)*
Phrase	*words per NP (WD_NP)* *words per SUBJ (WD_SUB)*

For lexical complexity, the taxonomy by Bulté and Housen (2012) shows that the various constructs—diversity, density, compositionality and sophistication—are less clear-cut and not as easy to separate; analyses by Šišková (2012) confirm this overlap. Table 4 gives an overview of the lexical complexity measures used. The measures selected all focus on content words (nouns, verbs, adjectives).

As measure of diversity a simple type-token ratio (TTR) was used for the English data as all transcript were of the same length, and the Guiraud index was taken for the German data due to the shorter text length of three of the interviews. As measures of sophistication a corpus-internal frequency measure (see Verspoor, Schmid, & Xu, 2012) was applied that calculates the relation of each speaker's use of the most frequent (F1) and most infrequent (F5) 20% of content words in relation to their use by the other participants. In addition, two corpus-external frequency measures based on the spoken part of the British National Corpus (BNC) for English were obtained, namely for individual lexical items and for trigrams, i.e., multiword units composed of three lexical items occurring together such as "the red cat." The trigram measure was added since recent investigations by Erman, Denke, Fant, and Forsberg Lundell (2014; see also Forsberg Lundell & Lindqvist, 2012) found that the use of multiword units distinguished the most advanced from the less advanced L2 speakers. Finally, several measures gauged the depth of the lexical knowledge in the English data: hypernymy and polysemy measures using Coh-Metrix (for details see Crossley, Salsbury, & McNamara, 2009; Crossley et al., 2011a, 2011b) as well as concreteness (ease of description[4]) and imageability (ease of creating a mental image) measures of word meanings based on TAALES (Kyle & Crossley, 2015).

TABLE 3 Overview of morphological complexity measures

Morphological level	Statistical construct
	passives per clause (for English) (PASS_C) *subjunctives per clause (for German) (KONJ_C)* *complex VPs per clause (for German) (KV_C)*

TABLE 4 Overview of lexical complexity and respective measures

Lexical level	Statistical construct
Diversity	*Type token ratio (for English) (TTR)* *Guiraud (for German) (GUI)*
Sophistication	*Frequency Bands for frequent content words (F1) and infrequent content words (F5) (F1RATIO & F5RATIO)* *BNC spoken for content words (for English) (BNC.Spoken.Freq.CW.log)* *BNC spoken trigram frequency (for English) (BNC.Spoken.Trigram.Freq.log)* *Hypernymy (for English)* *Polysemy (for English)* *Concreteness (for English) (MRC.concreteness. CW)* *Imageability (for English) (MRC.imageability. CW)*

4.4 | Analytical procedure

The final dataset was split into the L2S dataset consisting of 102 cases for nine grammatical complexity variables and nine lexical complexity variables and the L1A dataset consisting of 73 cases for 10 grammatical complexity variables and three lexical complexity variables. The large difference between the number of lexical complexity variables in the two datasets is due to technical limitations. There are simply more sophisticated tools available for assessing English data than German data. Each dataset was first explored using hierarchical clustering and then validated by means of principal component analysis.

Hierarchical cluster analysis (HCA) serves to find similar sets of objects within a dataset, in our case clusters of grammatical and lexical complexity measures. The analysis was run with the R-software (R Core Team, 2015) and the *hclust* function using Ward's minimum variance method. In a HCA, initially each object, i.e., each grammatical and lexical measure, is assigned to its own cluster. In a step-wise fashion an algorithm proceeds iteratively joining the two most similar (or least dissimilar) clusters (measures). This procedure continues until there is a single cluster, one for grammatical and one for lexical complexity measures. The function outputs a dendrogram illustrating the hierarchical cluster structure of the dataset. For more details on the analysis see Murtagh and Legendre (2011).

To verify the hierarchical cluster analysis we ran a principal component analysis (PCA) on the same sets of data using the *R* packages *GPArotation* (Bernaards & Jennrich, 2005) and *psych* (Revelle, 2016). PCA aims at displaying the pattern of similarity in the original variables of a dataset, i.e., in our grammatical and lexical complexity measures, as a set of new orthogonal variables called principle components. The PCA will be assessed on the basis of the following criteria: First, Bartlett's test and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy check whether the variables correlate well enough but not too highly, which would indicate multicollinearity and cause problems in determining the unique contribution to a factor of the variables that are highly correlated (see Field, Miles, & Field, 2012: 771). Based on the eigenvalues as depicted in a scree plot, the number of components to be extracted is determined. Through a rotation, the loading of each variable (measure) on one of the extracted components can be maximized, while minimizing the loadings on all other factors. The final output is a table of factor loadings, showing to what extent each original variable loads onto each component. The RQ will be evaluated based on the PCA outcomes.

5 | RESULTS

5.1 | Grammatical complexity

5.1.1 | L2 English speakers

The dendrogram shows that a hierarchical structure of one cluster in the main and one additional cluster emerged from the cluster analysis of the grammatical complexity measures in the English data (see Figure 1).

The cluster in the main combines the following measures: adverbials per AS-unit, words per utterance, words per AS unit and DCs per AS-unit, passives per clause, words per subject, words per clause and words per NP (including subject NPs). All may be considered measures of length at the sentence, clausal and phrasal level. A second cluster consists of a more fine-grained measure of subordination, the object subject relative clause ratio.

To validate the HCA a principal component analysis (PCA) was carried out using the same dataset. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis KMO = 0.65 (‘mediocre” according to Kaiser, 1974), and all KMO values for individual items except for one were > 0.55, which is above the acceptable limit of 0.5. We kept the DC_AS “item” which was only slightly below the limit at.48. Bartlett's test of sphericity, χ^2

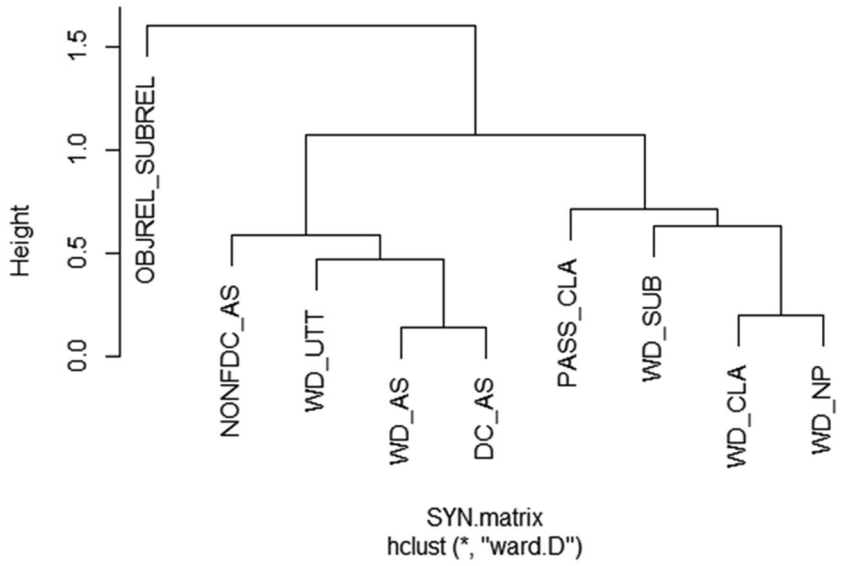

FIGURE 1 Dendrogram for L2 English grammatical complexity measures

(36) = 912.37, $p < 0.001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Two components had eigenvalues over Kaiser's criterion of 1 and in combination explained 63.11% of the variance. The scree plot was slightly ambiguous and would justify retaining two or three components. Given the hypothesized multidimensionality of complexity, three components were retained in the final analysis.

Table 5 shows the factor loadings after oblique rotation. The items that cluster on the same components suggest that component 1 represents complexity in terms of overall length, component 2 represents complexity in terms of

TABLE 5 Pattern matrix for L2 English grammatical complexity measures

	Oblique rotated factor loadings		
Item/measure	1 (Length)	2 (Subordination)	3 (Specific subordination)
WD_SUB	**0.83**	−0.27	0.01
WD_CLA	**0.81**	0.17	−0.05
WD_NP	**0.77**	0.11	−0.11
PASS_CLA	**0.66**	−0.16	−0.21
WD_AS	**0.59**	**0.57**	0.04
WD_UTT	**0.56**	**0.33**	0.24
NONFDC_AS	−0.13	**0.94**	−0.12
DC_AS	0.22	**0.77**	0.10
OBJREL_SUBREL	−0.06	−0.05	**0.94**
eigenvalues	3.32	2.24	1.04
% of variance	36.89	24.89	11.56
Alpha	.58	.55	

The abbreviations are explained in Table 2 and 3. The figures in bold indicate which factor each measure loads most strongly on.

subordination, and component 3 includes only a single measure, object relative clause per subject relative clause ratio. The PCA thus confirms the clusters found in the HCA. In Table 5 Alpha indicates the reliability of the scale which should be at least within the range of 0.7 to 0.8 or above. This is not the case for factor 1 and 2 which may be related to their overlap as illustrated in the matrix.

5.1.2 | L1 German attriters

For the L1 German dataset the HCA resulted in a hierarchical structure of two clusters between which the variables were significantly different in the main (see Figure 2).

The cluster in the main consists primarily of length measures at the utterance, AS, clausal and phrasal (NP and subject phrase) level. They combine with a general measure of subordination (DC per AS) to form the left strand of the dendrogram. On the right strand the next most plausible cluster in the hierarchy consists of morphological measures (passives, subjunctives and complex verb phrases per clause) which cluster together with a relative clause ratio.

To validate the HCA, a PCA was carried out using the same dataset. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis KMO = 0.64 ('mediocre'; Kaiser, 1974), and all KMO values for individual items except for one were > 0.53. We kept the DC_AS "item" which was only slightly below the limit at 0.49. Bartlett's test of sphericity, χ^2 (45) = 584.4, $p < 0.001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis showed that three components had eigenvalues over Kaiser's criterion of 1 and in combination explained 69.2% of the variance. The scree plot was slightly ambiguous and would justify retaining one or three components. Given the hypothesized multidimensionality of complexity and our approach to the English grammatical data, three components were retained in the final analysis.

Table 6 shows the factor loadings after oblique rotation. The items that cluster on the same components suggest that component 1 represents complexity in terms of length (and subordination), component 2 represents morphological complexity, and component 3 consists of the object subject relative clause ratio. Again, the results of the HCA are confirmed by the PCA.

Alpha, indicating the reliability of the scale, should be at least within the range of 0.7 to 0.8 or above. This is the case only for the second component which again may be related to the tight relationships between the factors.

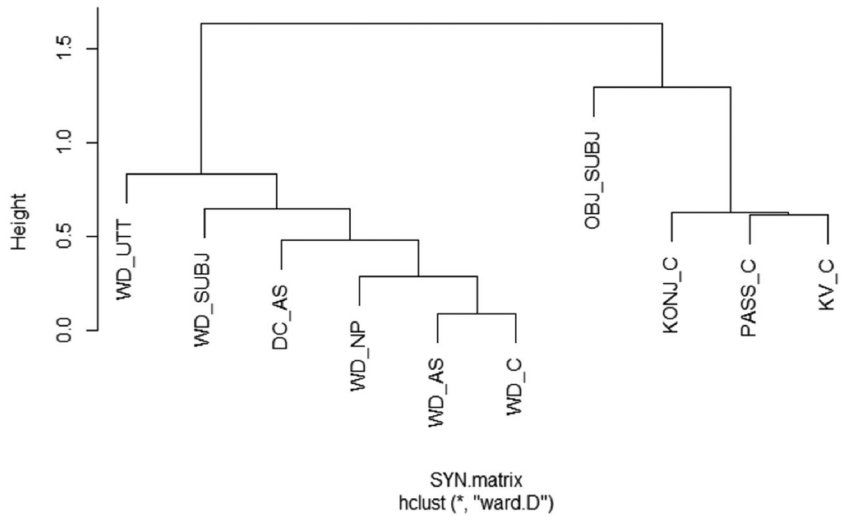

FIGURE 2 Dendrogram for L1 German grammatical complexity measures

TABLE 6 Pattern matrix for L1 German grammatical complexity measures

	Oblique rotation factor loadings		
WD_NP	**0.91**	**−0.09**	**0.19**
WD_CLA	**0.90**	0.10	0.08
WD_AS	**0.86**	0.25	−0.07
WD_SUBJ	**0.76**	−0.37	−0.23
DC_AS	**0.58**	0.35	−0.24
WD_UTT	**0.43**	−0.01	−0.32
KV_CLA	−0.07	**0.79**	0.04
KONJ_CLA	0.21	**0.73**	0.12
PASS_CLA	0.10	**0.63**	−0.35
OBJ_SUBJ	0.05	0.02	**0.91**
eigenvalues	3.67	1.97	1.28
% of variance	36.7	19.7	12.8
Alpha	.26	.61	

The abbreviations are explained in Table 2 and 3. The figures in bold indicate which factor each measure loads most strongly on.

5.2 | Lexical complexity

5.2.1 | L2 English speakers

For the L2 English data the HCA on the lexical complexity measures showed that two clusters in the main emerged (see Figure 3).

One cluster contains the F5 Ratio and the TTR, as well as hypernymy, imageability and concreteness. The second cluster contains the two external frequency measures, the F1 ratio and Polysemy. The low height indicates the high degree of similarity between the measures in each cluster.

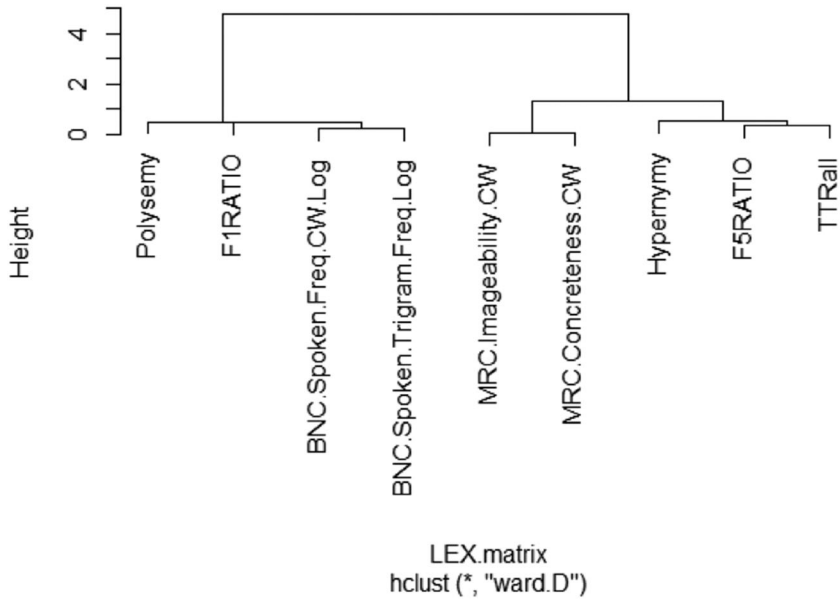

FIGURE 3 Dendrogram for L2 English lexical complexity measures

To validate the HCA a PCA was carried out using the same dataset. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis KMO = 0.72 ("good" according to Kaiser, 1974), and all KMO values for individual items were > 0.68. Bartlett's test of sphericity, χ^2 (15) = 352.60, p < 0.001, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Two components had eigenvalues over Kaiser's criterion of 1 and in combination explained 71.88% of the variance. The scree plot justified retaining three components.

Table 7 shows the factor loadings after oblique rotation. The items that cluster on the same components suggest that component 1 represents lexical diversity, component 2 represents measures of lexical abstractness, and component 3 represents frequency-based measures (capturing frequent word use). Components 1 and 2 correspond to the right cluster found in the HCA, while component 3 corresponds to the left cluster. Alpha indicates the reliability of the scale, which should be at least within the range of 0.7 to 0.8 or above. This is not the case for factors 2 and 3.

5.2.2 | L1 German attriters

Because there were only three lexical complexity measures for the German data, only a cluster analysis, but no PCA was carried out. The dendrogram of the HCA (see Figure 4) illustrates however that the frequency band 5 ratio (representing the use of least frequent tokens within the corpus) and the Guiraud cluster together, whereas the frequency band 1 ratio (representing the use of the most frequent tokens within the corpus) clusters separately.

5.3 | Summary of the results

For both datasets, both the HCA and PCA identified three clusters of grammatical complexity measures. One cluster included mainly length measures while another cluster consisted of just one measure, a specific subordination measure (object subject relative clause ratio). The remaining cluster consisted of two more general subordination measures in the English data, and of language-specific morphological complexity measures in the German data. However, given the low alpha values, the reliability of the clusters is questionable. There seems to be a considerable overlap between measures, but more so in the English than the German dataset.

TABLE 7 Pattern matrix for L2 English lexical complexity measures

	Oblique rotated factor loadings		
Item/measure	**1**	**2**	**3**
TTR	**0.92**	0.05	0.13
F5 Ratio	**0.91**	−0.14	−0.06
Concreteness	−0.07	**1.01**	0.06
Imageability	−0.02	**0.97**	−0.03
Hypernymy	0.49	**0.53**	−0.12
Polysemy	0.10	0.01	**0.98**
F1 Ratio	−0.48	0.07	**0.49**
BNC Spoken Trigram Frequency	−0.44	−0.12	**0.48**
BNC Spoken Frequency	−0.43	−0.35	**0.46**
eigenvalues	2.86	2.52	1.99
% of variance	31.78	28.00	22.11
Alpha	.91	.59	.54

The abbreviations are explained in Table 4. The figures in bold indicate which factor each measure loads most strongly on.

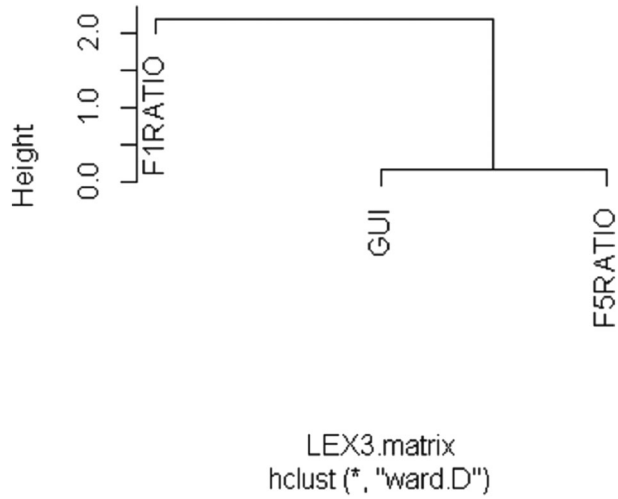

FIGURE 4 Dendrogram for L1 German lexical complexity measures

For the lexical complexity measures, a general division was found in both datasets between measures capturing lexical diversity, and measures capturing the use of frequent words. In the English data, the additional BNC-based frequency measures clustered together with the other measures for frequent word use, while the measures for the use of abstract words constituted a component of their own.

6 | DISCUSSION

The results of both datasets confirm the multidimensionality of grammatical and lexical complexity. However, the quantitative/statistical multidimensionality does not correspond one to one with the conceptual multidimensionality of both constructs.

For *grammatical complexity*, the principal component analysis revealed a cluster of length measures and a cluster of subordination ratios for the English L2 data. These measures correspond to the grammatical complexity measures most frequently used in the CAF literature (Bulté & Housen, 2012; Ortega, 2012). This distinction between measuring length and amount of subordination is also supported by the German data. While the only subordination ratio applied to the German data did cluster with the length measures, the reliability of this cluster was quite limited (alpha = 0.26) and excluding the subordination ratio and the average length of utterances from that principal component revealed a considerable increase in reliability (alpha = 0.6). This points to a certain independence of the subordination ratio from the length measures, as indicated by the clusters found in the English data. The fact that the subordination ratio clustered with the length measures in the German data might be an artefact of the number of measures used for this language: had there been more than one subordination ratio, a separate subordination cluster might have emerged just as it did in the English data. That the subordination ratio clustered together with length measures is probably due to the fact that an increase in subordination always results in an increase in AS-unit length as well. Altogether, the results thus support the long-standing practice of measuring both length and ratio of subordinate clauses as separate aspects of grammatical complexity.

At the same time, a closer look at the length measures seems warranted. In the current study, length measures at all levels of the utterance (phrase, clause, AS-unit and utterance) correlated highly and clustered together, indicating that they tap into the same dimension. This is at odds with Norris and Ortega's (2003) suggestion that

measures of clausal length (targeting phrasal elaboration) and supra-clausal length should be distinguished, and Michel et al.'s (2007) findings that phrasal complexity measures are distinct from other measures. The cause for this discrepancy might be the mode of the language sample studied here. Reliably identifying an utterance in speech is notoriously difficult (as indicated by the increase in reliability of the first component in the German data when utterance length is excluded), and measuring phrasal/clausal elaboration might be less informative in speech (as opposed to written, academic registers; Biber, 2006), where the length of coherent units is relatively limited (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Leech, 2000). This would mean that the well-defined and well-identifiable AS-unit might be best suited as unit of reference when assessing complexity in speech, with the added benefit that it offers a sound basis for calculating a subordination ratio (see above). Further length-based measures seem unnecessary.

Apart from the divide between length and subordination measures, we found that the object subject relative clause ratio clustered separately from all other measures. While we included this ratio to capture different types of subordination, we acknowledge that it may be a measure of difficulty rather than structural complexity. As psycholinguistic research has demonstrated, object relative clauses are cognitively more demanding than subject relative clauses (King & Just, 1991; King & Kutas, 1995; Traxler, Morris, & Seeley, 2002). Still, the results point to the added value of targeting language-dependent, specific clause types when assessing complexity, such as cleft sentences which are known to be infrequent in English (Roland, Dick, & Elman, 2007) or sentences in German where a syntactic constituent other than the subject occupies the pre-verbal position in the clause (e.g., M. S. Schmid, 2012).

Finally, the analysis of the German data revealed a separate cluster of morphological measures. This is in line with Bartning's (1997) research that stresses the necessity for capturing morphological traits in advanced learners of L2 French which distinguish them from native speakers. Hitherto, measures of morphology may generally have been neglected due to the large number of studies done on L2 English and English being a morphologically poor language where those measures are less informative. The current result confirms the need to take into account cross-linguistic variation (De Clercq, 2014; Bernardini & Granfeldt, 2019; Kuiken & Vedder, 2019) and to establish reliable measures of morphological complexity (Pallotti, 2014).

Altogether these results confirm a multidimensional, analytic view of grammatical complexity, calling for the use of several measures: length and general subordination, but also specific sentence types and morphology.

For *lexical complexity*, the analysis of both datasets revealed a cluster targeting lexical diversity. That cluster included a measure that was originally aimed at capturing lexical sophistication, namely the corpus-internal frequency measure for the least frequently used lexical items. That this measure still clustered together with the diversity measure (TTR/Guiraud) is presumably because an increased use of more infrequent lexical items is likely to result in a greater diversity of words used.

A second cluster that emerged in both datasets consisted mainly of measures for frequent lexical items. For the English data, this cluster included the corpus-external BNC-measures and a measure for polysemy. In particular the polysemy and the BNC-trigram frequency measure share the fact that they target more frequently used lexical items which can explain the emergence of this cluster, but the finding runs contrary to Šišková's (2012) results where measures which cover the most frequently used words in the BNC corpus correlated with lexical diversity measures.

In the English L2 data, we found an additional third cluster of measures capturing the use of abstract words. The reason for finding more clusters for the English data is the greater number of lexical measures applied to that dataset, in particular measures which have been proposed to assess lexical sophistication (Kyle & Crossley, 2015). The inclusion of these measures was motivated by the aim to assess advanced language proficiency, and lexical sophistication has been shown to increase with growing L2 proficiency (Crossley et al., 2009, 2011a).

The analyses confirm that lexical complexity is a multidimensional construct. What emerges from the data are two distinct aspects of lexical complexity: lexical diversity and lexical sophistication. In the English dataset (to which a wider range of measures was applied), lexical sophistication could be divided into two related aspects: one capturing the semantic dimension of abstractness/concreteness and hypernymy, and one related to the use of frequent or polysemous words, i.e., general words. That this latter aspect may be interpreted as related to lexical sophistication

is supported by the fact that a measure assessing the use of multiword units (trigrams) also clusters with this aspect: the frequency of multiword units occurring in advanced L2 speech has been found to be significantly lower than for L1 speakers (e.g., Erman et al., 2014; Forsberg Lundell et al., 2013). Taken together, the use of frequent or polysemous words, i.e., general words, and of multiword units which are typical for native speakers is a dimension of lexical sophistication along which language users of different proficiency levels may be distinguished. All in all, this suggests that the number of lexical complexity measures might be reduced to three, capturing the use of frequent, infrequent, and abstract lexical items.

7 | CONCLUSION

Our goal was to shed light on two constructs—grammatical and lexical complexity—frequently used in L2 experimental studies, especially those falling within the CAF framework. We did so by applying the taxonomies proposed by Bulté and Housen (2012) to a context that is not frequently investigated in the context of CAF research: a corpus of oral history testimonies given by highly advanced bilingual speakers, namely L2 speakers and L1 attriters. Our findings confirm the multidimensionality of the two constructs, even though some clusters deviated to some extent from the ones frequently proposed in analytic approaches to complexity (e.g., the lexical cluster encompassing frequency measures is not in line with Šišková's (2012) results). There was also considerable overlap between measures, implying the possibility to reduce the number of measures traditionally used, but still capturing the multidimensionality of complexity. The results also indicate the need for more fine-grained subdimensionalities of complexity. For example, grammatical complexity can be measured more generally (length and subordination) but also at a deeper level with respect to different types of constructions (clause types, morphology). This finding may be related to the specificity of our dataset, i.e. the very advanced levels displayed by our bilingual speakers, and the mode, i.e. speech. Nevertheless, it would be interesting to include additional, recently proposed measures of lexical diversity as well as other measures of lexical sophistication to investigate whether they can capture additional aspects.

There is yet little longitudinal research capturing the long-term development of grammatical and lexical complexity (Lambert & Kormos, 2014), and also this study looked at one point in time only, However, the methodology of the current study might be applied to a longitudinal design. In doing so, one could investigate the composition of clusters over time for changes with different levels of proficiency. If such changes were to be found, a rigid taxonomy of complexity might not turn out to be ideal and a broader construct might be needed. Indeed, Ortega and Byrnes (2008: 282) write with respect to the overall CAF framework that "the narrow focus on accuracy, fluency, and complexity traits in isolatable domains such as lexis and grammar does not capture defining aspects of advanced levels of ability, particularly the textual oriented, socially embedded, and situationally motivated nature of language use that addresses a vast area of concerns in human life." Finally, comparable native control groups should be considered for future studies.

Overall, this study is in support of more fine-grained analytic approaches to complexity, but also acknowledges holistic views such as those taken by Ortega, Byrnes, Bartning and others. We hope our findings will encourage researchers to search for a more commonly acceptable construct definition and operationalization that has the potential to hold across a varied range of contexts. Such a construct should also be broad enough to be applied to very advanced levels of spoken language productions, for example, by long-term immigrants who are well-established in their host countries and who are highly advanced in their L2, but who might suffer from varying degrees of attrition in their L1.

ACKNOWLEDGEMENTS

This research was financed by the Dutch national research foundation (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), Grant number 360-70-420.

ENDNOTES

[1]See Coh-Metrix (Graesser, McNamara, & Kulikowich, 2011; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Louwerse, McCarthy, & Graesser, 2010)

[2]In main clauses it was determined how frequently sentences contained a topicalized element other than the syntactic subject, i.e. the number of sentences of the type XVS. She also counted the number of main clause constructions in which discontinuous word order was used, i.e. sentences with a finite and a non-finite verb where the two are split to frame other constituents such as postverbal subjects, objects, prepositional phrases, etc.

[3]The two groups consist of different participants. None of the participants gave an interview in both languages.

[4]The TAALES tool assumes that, for example, word meanings that can be described by simply pointing at an object are more concrete than word meanings that can only be explained by using other words.

ORCID

Cornelia Lahmann https://orcid.org/0000-0002-7543-6963
Rasmus Steinkrauss https://orcid.org/0000-0002-3643-8704

REFERENCES

Bartning, I. (1997). L'apprenant dit avancé et son acquisition d'une langue étrangère. *Acquisition et Interaction en Langue étrangère*, *9*, 9–50.

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. https://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*(2), 243–267. https://doi.org/10.1017/S0272263112000046

Beaman, K. (1984). Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In D. Tannen (Ed.), *Coherence in spoken and written discourse* (pp. 45–80). Norwood, NJ: Ablex.

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Bernardini, P., & Granfeldt, J. (2019). Linguistic complexity and cross-linguistic variation in L2 English, L3 French and L4 Italian. This issue

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/scl.23

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Ltd.

Brezina, V., & Pallotti, G. (2016). Morphological complexity in written L2 texts. *Second Language Research*, *95*(1), 99–119. https://doi.org/10.1177/0267658316643125

Bulté, B. (2013). *The development of complexity in second language acquisition: A dynamic systems approach* (Unpublished doctoral dissertation). Vrije Universiteit Brussel, *Brussels*.

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. https://doi.org/10.1075/lllt.32.02bul

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

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

Crossley, S. A., Salsbury, T., McNamara, D., & Jarvis, S. (2011b). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, *45*(1), 182–193. https://doi.org/10.5054/tq.2010.244019

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

De Clercq, B. (2014). The development of morphological complexity: A cross-linguistic study on L2 French and English. Paper presented at the Colloquium on Linguistic Complexity, Brussels.

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*(2), 160–182. https://doi.org/10.1111/ijal.12061

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, CA: Sage publications.

Forsberg, F. (2010). Using conventional sequences in L2 French. *IRAL—International Review of Applied Linguistics in Language Teaching*, *48*(1), 25–51. https://doi.org/10.1515/iral.2010.002

Forsberg Lundell, F., Bartning, I., Engel, H., Gudmundson, A., Hancock, 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. https://doi.org/10.1093/applin/21.3.354

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

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. https://doi.org/10.3758/BF03195564

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–473. https://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/lllt.32

Jarvis, S. (2013a). Capturing the diversity in lexical diversity. *Language Learning*, *63*(s1), 87–106. https://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. https://doi.org/10.1075/sibil.47.03ch1

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/bf02291575

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

King, J., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376–395. https://doi.org/10.1162/jocn.1995.7.3.376

Kuiken, F., & Vedder, S. C. (2019). Syntactic complexity across proficiency levels and across languages: An investigation of L2 and L1 writing in Dutch, Italian and Spanish. This issue.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757–786. https://doi.org/10.1002/tesq.194

Lahmann, C., Steinkrauss, R., & Schmid, M. S. (2016). Factors affecting grammatical and lexical complexity of long-term L2 speakers' oral proficiency. *Language Learning*, *66*(2), 354–385. https://doi.org/10.1111/lang.12151

Lambert, C., & Kormos, J. (2014). Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition. *Applied Linguistics*, *35*(5), 607–614. https://doi.org/10.1093/applin/amu047

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

Louwerse, M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society* (pp. 843–848). Mahwah, NJ: Erlbaum.

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

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292–330. https://doi.org/10.1080/01638530902959943

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.

Murtagh, F., & Legendre, P. (2011). Ward's hierarchical clustering method: Clustering criterion and agglomerative algorithm (arXiv e-print No. 1111.6285). Retrieved May 3, 2019 from http://arxiv.org/abs/1111.6285

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 716–761). Oxford: Blackwell Publishing Ltd. Retrieved May 3, 2019 from http://onlinelibrary.wiley.com/doi/10.1002/9780470756492.ch21/summary

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. https://doi.org/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: Walter de Gruyter. https://doi.org/10.1515/9783110229226.127

Ortega, L., & Byrnes, H. (2008). Theorizing advancedness, setting up the longitudinal research agenda. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 281–300). New York/London: Routledge.

Osborne, J. (2011). Fluency, complexity and informativeness in native and non-native speech. *International Journal of Corpus Linguistics*, *16*(2), 276–298. https://doi.org/10.1075/ijcl.16.2.06osb

Osborne, J. (2013). Fluency, complexity and informativeness in native and non-native speech. In G. Gilquin, & S. De Cock (Eds.), *Errors and disfluencies in spoken corpora* (pp. 139–161). Amsterdam: John Benjamins Publishing.

Pallotti, G. (2014). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117–134. https://doi.org/10.1177/0267658314536435

R Core Team. (2015). R: A language and environment for statistical computing. Retrieved May 3, 2019 from http://www.r-project.org

Revelle, W. (2016). Psych: Procedures for personality and psychological research—Google scholar. Northwestern University, Evanston, Illinois, USA. Retrieved May 3, 2019 from https://CRAN.R-project.org/package=psych

Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, *23*(4), 497–519. https://doi.org/10.1191/0265532206lt339oa

Rimmer, W. (2008). Putting grammatical complexity in context. *Literacy*, *42*(1), 29–35. https://doi.org/10.1111/j.1467-9345.2008.00478.x

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

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, H. (1995). Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-workshop* (pp. 47–50). Dublin. Retrieved 3 May, 2019 from http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger

Schmid, M. S. (2002). *First language attrition, use and maintenance: The case of German Jews in anglophone countries*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/sibil.24

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. https://doi.org/10.1075/lab.2.2.03sch

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

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Traxler, M. J., Morris, R. K., & Seeley, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90. https://doi.org/10.1006/jmla.2001.2836

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. https://doi.org/10.1016/j.jslw.2012.03.007

How to cite this article: Lahmann C, Steinkrauss R, Schmid MS. Measuring linguistic complexity in long-term L2 speakers of English and L1 attriters of German. *Int J Appl Linguist*. 2019;29:173–191. https://doi.org/10.1111/ijal.12259