BRILL

Language Dynamics and Change 3 (2013) 218–244

brill.com/ldc

Learning Complex Features: A Morphological Account of L2 Learnability

Job Schepens[a, b], Frans van der Slik[a, c] and Roland van Hout[a]
a) Centre for Language Studies, Radboud University Nijmegen, The Netherlands
b) International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
c) Department of Linguistics, University of the Free State, Bloemfontein, South Africa

j.schepens@let.ru.nl

Abstract
Certain first languages (L1) seem to impede the acquisition of a specific L2 more than other L1s do. This study investigates to what extent different L1s have an impact on the proficiency levels attained in L2 Dutch (Dutch L2 learnability). Our hypothesis is that the varying effects across the L1s are explainable by morphological similarity patterns between the L1s and L2 Dutch. Correlational analyses on typologically defined morphological differences between 49 L1s and L2 Dutch show that L2 learnability co-varies systematically with similarities in morphological features. We investigate a set of 28 morphological features, looking both at individual features and the total set of features. We then divide the differences in features into a class of increasing and a class of decreasing morphological *complexity*. It turns out that observed Dutch L2 proficiency correlates more strongly with features based on increasing morphological complexity ($r = -.67$, $p < .0001$) than with features based on decreasing morphological complexity ($r = -.45$, $p < .005$). Degree of similarity matters ($r = -.77$, $p < .0001$), but increasing complexity seems to be the decisive property in establishing L2 learnability. Our findings may offer a better understanding of L2 learnability and of the different proficiency levels of L2 speakers. L2 learnability and L2 proficiency co-vary in terms of the morphological make-up of the mother tongue and the second language to be learned.

Keywords
morphological complexity; WALS; adult language learning; L2 learnability; speaking proficiency

1. Interaction between Second Language Learning and Linguistic Structure

Children seem to learn languages easily, in a natural way, unlike adults, who often struggle when learning to understand a second language and express themselves in it. Their struggle can often be noticed in their use of L2 morphology, as inflected forms are often missing or incorrect (for L2 Dutch, see Ol-

DOI: 10.1163/22105832-13030203

denkamp, 2013). Previous research on L2 learning impediments has taken different perspectives on L1-L2 linguistic differences, for example by means of (1) contrastive analysis (Lado, 1957; Odlin, 1989; Towell and Hawkins, 1994; Weinreich, 1963), (2) linguistic distance (Chiswick and Miller, 2005; Van der Slik, 2010), and (3) morphological complexity (Dahl, 2004; Lupyan and Dale, 2010; McWhorter, 2007; Nettle, 2012).

The notion of morphological complexity is relevant for explaining patterns of variation in the morphological make-up of languages. Language contact has a direct impact on morphological complexity, in particular in combination with mechanisms of adult language learning. Correlational evidence obtained from typological data (Lupyan and Dale, 2010; Nettle, 2012) indicates a decrease in morphological complexity of languages when the number of L2 learners increases. These studies confirm on a larger scale what is observed in smaller scale acquisition studies (Ionin and Wexler, 2002; Lardiere, 1998): adult learners have persistent problems in L2 acquisition, especially in acquiring L2 morphosyntax.

If complexity is so essential, it is tempting to conclude that some languages are easier to learn for adults than others. Trudgill (1983, 2011) points to Dauenhauer and Dauenhauer (1998), who investigated reversing language shift in Tlingit, Haida, and Tsimshian. They conclude that "the languages of Southeast Alaska are intrinsically more difficult to learn than Maori or Hawaiian because of their more complex grammars and phonologies." Variation between languages in their morphological make-up and complexity has a strong influence on how these languages are transmitted in language contact scenarios (Andersen, 1988; Braunmüller, 1990; Dahl, 2004; Kusters, 2003). The consequences are, as Trudgill argues, that the 'easier' languages are highly analytical (less complex morphology, more lexical means), often because they have experienced more contact. Language complexity is linked to adult L2 learning difficulty, although the precise mechanisms involved are far from clear. As Trudgill (2011: 41) notes, "Dahl (2004: 39) prefers to suppose that complexity and L2 difficulty are not actually identical but simply 'related.' "

The main aim of the present study is to investigate whether data regarding adult L2 learning, in particular L2 Dutch, reveal effects of morphological distance (differences) and complexity. We have shown earlier that the lexical distance between L2 Dutch and the L1s of L2 learners is systematically correlated with L2 Dutch proficiency (Schepens et al., 2013). Secondly, we want to investigate the additional value of morphological L1-L2 distance measures in comparison to the impact of lexical L1-L2 distance we found earlier. We hypothesize that differences in morphological make-up in general and differences in morphological complexity in particular account for the L2 learnability of Dutch. More specifically, we expect that L2 learnability is lower when the L1 is morphologically less complex as compared to the L2.

The notion of L2 learnability may help to shed more light on the likelihood of L1-dependent biases in learning L2 linguistic features. Typologically relevant linguistic features for many languages can be found in the online World Atlas of Language Structures (henceforth WALS) database (Dryer and Haspelmath, 2011). Lupyan and Dale (2010) used the WALS data to define a set of 29 morphological features on which they based their correlational study on language structure and population sizes. They ordered the variants of those features (i.e., the feature values) on a complexity scale. We employ the morphological set of features they extracted from the WALS database, also making use of the complexity scales they defined. We systematically compare the Dutch variants of the morphological features with the variants in the L1s. For every feature in the L1 involved, we check if its variant is morphologically identical, more complex, or less complex as compared to Dutch.

Thus far, there are no large scale correlational studies of L2 learnability bias in adult L2 learning that encompass the structure of L1s. The correlational study of Lupyan and Dale (2010) implicitly assumes that all L1s are equally responsible for effects of population size on morphological complexity. A strong point of the present study is that we relate L2 proficiency scores to the structural features of the L1s of learners of L2 Dutch. The concept of L1-dependent L2 learnability can thus shed more light on the likelihood of L1-dependent biases in the L2 learnability of linguistic features.

To determine L2 proficiency levels in Dutch for learners who speak a typologically wide variety of L1s, we use speaking proficiency scores of Dutch as an L2 for speakers of 73 different L1s. The database allows for evaluation of morphological distance and complexity by means of a statistical analysis of more than 50,000 L2 proficiency scores.

In the following section, we define morphological complexity and provide an overview of current evidence for the relationship between morphological complexity and adult language learning. In the methods and results sections, we describe the development and testing of the impact of morphological distance and morphological complexity. We test the added value of morphological distance in relation to lexical distance between Dutch and the L1s involved. In the final section we discuss our findings and present directions for further study.

2. Morphology and Adult L2 Learning

When a set of morphological features is available for a set of languages, distances in terms of differences can be counted in a straightforward way by establishing whether the languages in question have the same feature value or not. Making comparisons in terms of morphological complexity is more difficult, however.

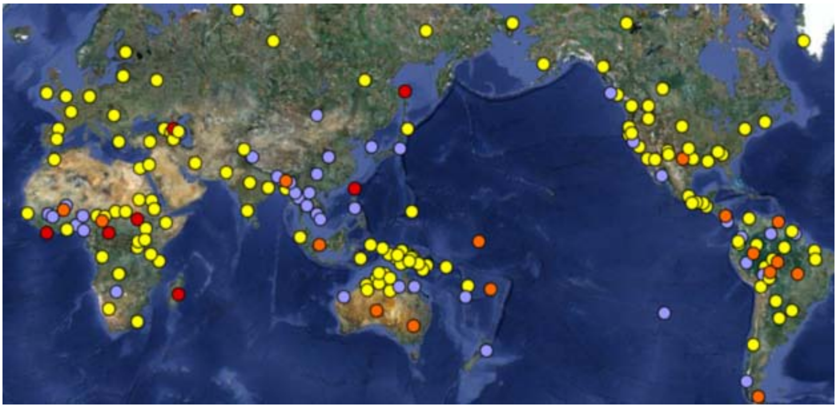

Figure 1. Verbal person marking (100): neutral (violet and red) versus non-neutral alignment (yellow and orange). Verbal subject marking for person and number (feature 29): none (violet and orange) versus other than none (yellow and red) (Dryer and Haspelmath, 2011)

Morphological complexity can be defined as the extent to which a language makes use of modifications of words (Nettle, 2012). This definition is in accordance with the notion of structural complexity of linguistic expressions (Dahl, 2004), and fits information theory in terms of compressibility (Juola, 1998; Lupyan and Dale, 2010). It is also in accordance with the notion of complexity in terms of L2 acquisition difficulty (Kortmann and Szmrecsanyi, 2012; Kusters, 2003: 6). Complexity reflects the investment needed for an adult L2 learner to acquire another language. It quantifies languages with more inflectional morphology as more complex relative to more isolating languages, based on the assumption that morphology is harder to acquire in an L2 than in an L1.

WALS provides data in terms of feature values across languages with varying degrees of inflectional morphology. Consider person/number marking on the verb, for example. Many languages mark person and number of the subject on the verb; however, in languages of Southeast Asia this is quite uncommon, as can be seen in Fig. 1. WALS contains at least 29 morphological features whose values range from less complex lexical variants to more complex inflectional devices (Lupyan and Dale, 2010). An overall degree of morphological complexity can be obtained by pairwise comparisons of the morphological complexity of feature values. Using the lexical-inflectional rank orders given in Table 1 of Lupyan and Dale (2010) as scales, languages can be compared and evaluated in terms of their morphological complexity.

This approach to morphological complexity challenges the traditional view that structural complexity is distributed uniformly across languages (Hockett, 1958: 180). It allows for cultural-evolutionary mechanisms that affect the development of complexity (Sampson et al., 2009). There is, in fact, recent evidence for the existence of cultural-evolutionary mechanisms in language structure (Evans and Levinson, 2009). For example, differences in language structure may be due to differences in genetic bias (Dediu and Ladd, 2007; Hunley et al., 2008) and population size (Wichmann and Holman, 2009; Wichmann et al., 2008). Thanks to, in all likelihood, the better availability of typological databases such as WALS, researchers are beginning to quantify structures cross-linguistically on a large scale.

Table 1 highlights the distinctions between morphologically less and more complex dimensions of language according to the linguistic niche hypothesis of Lupyan and Dale (2010). These authors hypothesize that the differences in social structure between esoteric and exoteric niches affect language structure. Languages with a relatively high number of L2 learners, as found in the exoteric niche, are more likely to use lexical means of expression. In contrast, languages spoken in the esoteric niche are supposedly more complex morphologically, as they adapt to an L1-facilitative structure.

Table 1. Dimensions in which morphologically more and less complex languages are assumed to differ

Dimension	Morphologically less complex	Morphologically more complex
Restrictedness	Ambiguous	Overspecified
Linguistic Strategy	Lexical / word order	Inflectional / conjugational
Learning Mechanism	Selection (facilitates L2)	Redundancy (facilitates L1)
Linguistic Type	Isolating	Synthetic
Cultural Type	Exoteric	Esoteric
Population	High, many adult learners	Low, many child learners

This observed negative relationship between population size and the degree of morphological complexity is in accordance with research from multiple disciplines. Studies in historical linguistics show that within many language families, morphological inflection has been lost because of changes in community structure (Kortmann and Szmrecsanyi, 2012; Kusters, 2003; McWhorter, 2002, 2007, 2011; Miestamo et al., 2008; Trudgill, 2001, 2002, 2011). Breaking down population size into specific L1/L2 community size estimates confirms the importance of the number of L2 learners compared to the whole population size (Bentz and Winter, 2013). Psycholinguistic studies and studies in language acquisition have come up with abundant evidence of learning differences between children and adults (Blom et al., 2006; Flege et al., 1999; Johnson and Newport, 1989; McDonald, 2000; Prévost and White, 2000). In addition, artificial language learning studies have uncovered a weaker bias for regularization in adult language learners as compared to child language learners (Smith and Wonnacott, 2010).

L2 learnability differences can be further illustrated with respect to the expression of verbal inflection by Mandarin Chinese learners of Dutch. Since no verbal inflection exists in Mandarin, one would expect these learners to prefer short verb forms corresponding to the stem of a verb. Oldenkamp (2013: 53) showed that Mandarin Chinese L2 learners of Dutch use verbal inflections less than Moroccan Arabic L2 learners of Dutch (whose native language does have verbal inflection). Hence, the realization of inflection in the L2 may depend on the degree of inflection in the L1.

In a previous study (Schepens et al., 2013), we showed that state exam data can be used successfully to compare how well lexical measures of linguistic distance explain differences in proficiency in L2 Dutch. Two different lexical measures of linguistic distance between the L1 and L2 were tested for their explanatory value of L1 variance in L2 proficiency scores (Gray and Atkinson, 2003; Wichmann et al., 2010). It was concluded that the effect of the L1 on learning L2 Dutch is a distance effect, as the linguistic distance between the L1 and L2 explained differences in L2 proficiency to a large extent (75.1%). This success raises the question as to whether morphology can explain the L1 variance in L2 proficiency levels even better. Does it have additional value?

Our first hypothesis is based on the observation that differences in morphological distance and complexity across L1s exist, and the premise that the more inflectional morphology an adult language learner needs to acquire, the lower L2 learnability is. Morphological distance is a result of either more or less morphology between an L1 and an L2. As a baseline, we expect that a higher distance between the L1 and the L2 is related to a lower L2 learnability, but that such a distance effect can be explained better in terms of complexity.

More specifically, we expect that the impact of morphology on L2 learnability is consistently present across families despite family-specific biases in the morphological make-up of languages. Recent studies show how some features are more stable than others (Dediu and Levinson, 2012) and how feature distributions depend on lineage-specific trends (Dunn et al., 2011); see for an overview Wichmann (in press). We therefore expect the impact of morphological differences to vary depending on the lineage in which the features evolved. Although we assume that an L2 learnability bias itself is not lineage-specific, a family bias is likely to affect its impact and could potentially conceal effects of morphological differences on L2 learnability.

Furthermore, measures of morphological distance or complexity may explain why a strong effect of lexical distance on L2 learnability can be observed across Indo-European languages. We hypothesize that morphological differences explain differences in L2 proficiency scores better than current measures of lexical distance.

3. Methods and Data

3.1. *Proficiency Scores of L2 Dutch*

A unique database is available in the Netherlands, consisting of L2 proficiency scores for the state exam *Dutch as a Second Language* for more than 50,000 participants. The exams are administered by the official Board of Examinations in the Netherlands, and developed by a large test battery constructor (Central Institute for Test Development; Cito) and the independent Bureau of Intercultural Evaluation. The exam is tailored to higher education; passing it is a requirement for individuals wanting to obtain admission to certain Dutch educational programs. The full exam consists of speaking, writing, listening, and reading tasks, for which proficiency scores are available for most participants. The speaking part of the exam comprises 14 tasks that are similar to one another, in which participants are required to provide information, give instructions, etc., and has to be completed in 30 minutes. The spoken language is evaluated by two independent examiners on both content and correctness according to a formal protocol. The pass level is upper-intermediate, comparable to the B2 level of the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Council of Europe, 2001).

Using the results of speaking exams for L1s for which at least 20 L2 proficiency scores were available, it is possible to compare 73 languages (L1s) with Dutch (L2). Following WALS (Dryer and Haspelmath, 2011), the 73 L1s come from 35 different genera which belong to 14 language families. Of these 73 languages, 39 are Indo-European and 34 are non-Indo-European. In the latter group, we have eight Niger-Congo languages, six Afro-Asiatic, four Austronesian, three Altaic, three Uralic, two Dravidian, and two Creole languages (Haitian and Papiamentu), as well as one Kartvelian (Georgian), one Austro-Asiatic (Vietnamese), one Sino-Tibetan (Chinese), and one Tai-Kadai language (Thai), and, finally, Japanese and Korean.

The L2 proficiency scores were annotated with control variables taken from questionnaire information on gender, educational level, length of residence, age at arrival in the Netherlands, and additional language background(s). Enrollment levels in higher education in the country of origin (World Bank, 2011) were included as well. We calculated *adjusted proficiency* scores for each L1 language. Adjusted proficiency is the by-L1 adjustment (BLUP) as taken from a multilevel model with the control variables as fixed effects and random effects for the L1 (mother tongue), L2 (additional language acquired before learning

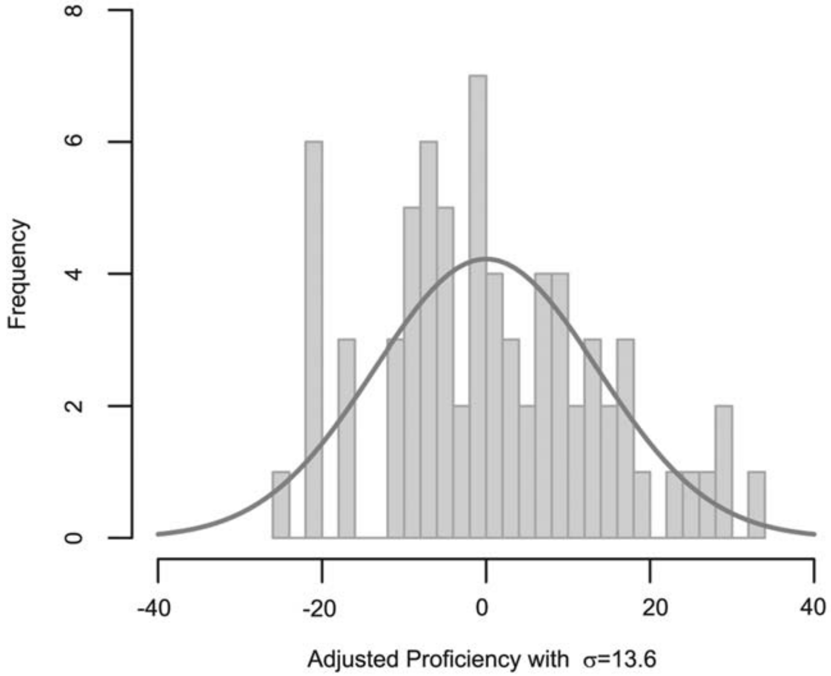

Figure 2. The distribution of adjusted proficiencies exhibits positive skew

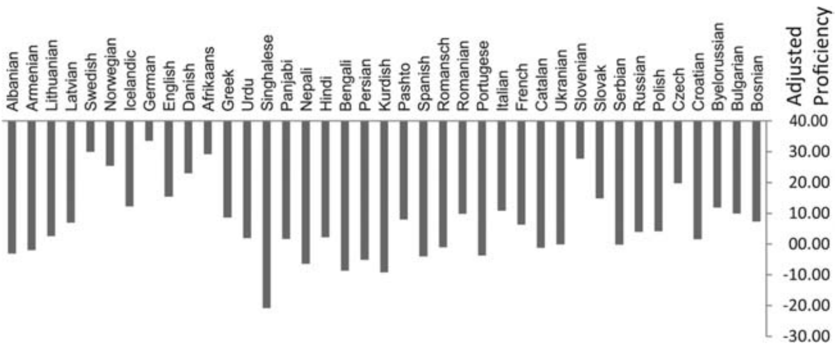

Figure 3. The distribution of adjusted proficiency among 33 non-Indo-European languages

L2 Dutch), L1-L2 combinations, and countries (Schepens et al., submitted). The adjusted proficiency measures were extracted with the function *ranef* from the R (R Core Team, 2013) lme4 package (Bates et al., 2011). The distribution of the

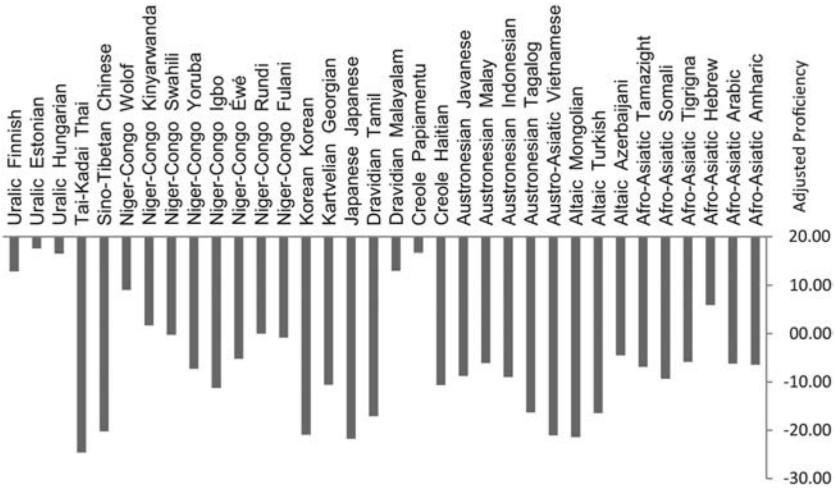

Figure 4. The distribution of adjusted proficiency among 39 Indo-European languages

adjusted proficiency scores for the 73 L1s is visualized in Fig. 2, where zero indicates the average adjusted proficiency score across L1s. Figures 3 and 4 provide an overview of the L1-specific L2 Dutch adjusted speaking proficiency scores for non-Indo-European and Indo-European L1s, respectively. The proficiency scores are generally higher for Indo-European languages (Fig. 3); some exceptions are the Uralic languages, which score higher than many Indo-European languages, and Singhalese, an Indo-European language, whose score is among the lowest overall. In the present study, the proficiency scores are used as the dependent variable.

3.2. *Morphological Feature Values*

Typological features are structural properties of language that represent dimensions of cross-linguistic diversity (Dryer and Haspelmath, 2011). A subset of 29 morphology-specific feature values was extracted from WALS by Lupyan and Dale (2010; note that feature number 26 in their ordering involves two WALS features). These 29 features cover a broad range of morphological dimensions (e.g., agreement, verb inflection, articles) and feature markings (e.g., no plurality vs. obligatory plurality). For our study, we first retrieved all the available feature values from WALS for all the 74 languages in our set (73 L1s plus Dutch). This resulted in a set of 1123 values, excluding all the missing feature values. We filled in six missing feature values for Dutch on the basis of the information provided by the ANS (Algemene Nederlandse Spraakkunst = “Dutch General Syntax”; Haeserijn et al., 1997).

All feature values of all the languages included were transformed into three measures in comparing Dutch and the 73 L1s: similarity, increasing complexity, and decreasing complexity. Similarity is 1 for an identical value of a feature and 0 for any other value. Increasing complexity is based on the observed patterns reported in Table 1 of Lupyan and Dale (2010). The measure distinguishes between languages that are less complex than Dutch for a specific feature versus languages that are equally or more complex than Dutch. The score of 1 indicates that a value in a specific L1 is either equal to Dutch or higher in the complexity ordering, the value of 0 indicates that a value of a specific L1 is lower in the complexity ordering than Dutch. The third transformation defines decreasing complexity from the perspective of the L1s, distinguishing between an equal or lower L1 complexity (coded as 1) versus a higher L1 level of complexity (coded as 0). In all three measures, the 1 is used to indicate equal and the 0 to indicate a difference. It is possible to compare the correlations between adjusted proficiency and each of the three measures in order to test which measure best explains variance in proficiency scores.

Lupyan and Dale (2010) report one feature pattern that seems reversely related to a complexity ordering, namely WALS feature no. 34: Coding/Occurrence of Plurality. Lupyan and Dale’s analysis indicates that obligatory plurality marking is more likely for languages in the exoteric niche. This runs counter to the fact that, according to the linguistic niche hypothesis, exoteric languages are generally more likely to use lexical strategies. In contrast, optional plurality marking (using either a word, affix, or clitic) or no plurality marking is more likely for languages in the esoteric niche. This seems to be a contradiction, considering the linguistic properties of rank ordering of the other features. If languages with obligatory plurality, like Dutch, are considered more complex than languages with no or optional plurality marking—contrary to the findings of Lupyan and Dale (2010)—plurality marking correlates strongly with proficiency scores (.651, $p \langle$ .0001, N = 34). It is thus debatable whether obligatory marking is of high or low complexity, and we remove this feature from the set of features analyzed here, resulting in a set of 28 features.

In 12 out of the 28 features considered in the present study, no other language is less complex than Dutch. Examples of languages that, for (almost) every feature value present in Dutch, have a feature value that is either equally or more complex are Hungarian (13 equally or more complex out of 15 observed values), German (12 out of 16), and Georgian (10 out of 13). On the other hand, languages in which feature values of lower complexity, as compared to Dutch, are predominant include Tagalog (9 less complex values out of 14), Vietnamese (12 out of 15), and Chinese (10 out of 12). We first report one-by-one comparisons of feature patterns to proficiency and then evaluate overall measures of linguistic distance based on a combination of these patterns.

3.3. *Data Analysis*

Beyond straightforward morphological similarity between L1 and L2, L2 learnability involves learning an increasing (higher) or decreasing (lower) level of morphological complexity, depending on the morphological features of the L1 and L2. We test whether increasing and decreasing complexity produce significant differences in L2 learnability and whether such a distinction is superior to a similarity-based morphological analysis of L2 learnability.

For evaluating the patterns of individual features, we compute feature-specific point-biserial correlation coefficients between adjusted proficiency scores and the three types of binary feature values (similarity, increasing complexity, decreasing complexity). We compute distances for all three types using a *sum of weighted features*. The weights are the correlations of each morphologically different feature that is more complex in Dutch. We divide these sums by the number of features for which information was available in WALS.

The distance scores were added to the original dataset. We fit linear mixed effects regression models (LMEM) to the adjusted proficiency scores using the *lme4* package in R. We model adjusted proficiency as a function of each of the three distance measures separately in three specific models with one fixed effect each, including a random effect for language family, as well as random slopes quantifying the by-family variance in proficiency. LMEMs can be used for modeling nested dependencies in random variance at the family level (Atkinson, 2011). Separate regressions for each family suffer from data sparseness and are likely to reveal family-specific idiosyncrasies (Jaeger et al., 2011; Levy and Daumé, 2011). By-family variance is the result of family-specific bias causing languages within a family to be more similar to each other. LMEMs control for such bias by fitting random intercepts and slopes. The random intercepts reflect, by assumption, the normally distributed family-specific intercepts, which capture systematic deviations in proficiency from the average family. The random slopes reflect family-specific relationships between distance measures and L2 proficiency. We choose to include by-family random slopes as it may a priori be expected that family bias moderates the relation between morphological distance and proficiency. This theoretical motivation of the random effect structure avoids overfitting to a particular sample (Barr et al., 2013), in this case a selection of languages from multiple families.

In all, we make use of a LMEM with morphological distance as a language-level fixed predictor and random intercepts and slopes across families (Gelman and Hill, 2006). The model has six parameters, comprising three variance components (variance across families, languages, and random slopes), one covariance coefficient (between slopes and families), one fixed effect, and an overall intercept. Adding another distance measure to this model (morphological or lexical) involves estimation of an additional random slope and a more complex covariance structure (random intercepts × distance measure 1, random intercepts × distance measure 2, distance measure 1 × distance measure 2).

4. Results

4.1. *Feature Patterns*

A distinction between increasing and decreasing complexity is unnecessary if it does not lead to a better explanation of L2 learnability differences than plain similarity does. Increasing and decreasing complexity together should explain at least as much variance in proficiency scores as similarity alone. In addition, if learning additional inflectional morphology is relatively hard for L2 learners, increasing complexity should match the similarity effect better than decreasing complexity does.

Table 2 shows the correlations between adjusted proficiency scores and measures of similarity, increasing complexity, and decreasing complexity for each of the 28 features. For morphological similarity, all eight significant correlations are positive, ranging between .31 and .68, meaning that a structurally different value in the L1 is often associated with a lower proficiency score. We assume that the negative non-significant values reflect sample fluctuations. Table 2 includes the number of languages for which information was available as well, a number that varies between 23 and 53. The varying numbers have an impact on the correlations observed, but it is hard to tell what the precise effects are. The global pattern obviously is that differences lead to impediments, although this is not a consistent finding for all features. Features spread in their effect on proficiency.

This overall view is confirmed for the features with increasing complexity. Some significant correlations are even a bit higher than their similarity counterparts. Seven correlations are significant, a subset of the eight significant similarity correlations (the exception is feature number 57, coding of possessives). Table 2 shows that increasing complexity captures most of the effects found for similarity. The feature patterns for increasing complexity are highlighted in more detail in Table 3. The correlations are ordered from high to low to illustrate which features affect proficiency the most. The patterns on each row are ordered from lower to higher complexity, e.g. no past tense ⟨ past tense. As Dutch has a past tense, languages with no past tense have a lower morphological complexity than Dutch. The positive correlations (14 out of 16) indicate that L1s with a relatively low proficiency score are likely to be less complex than Dutch. Seven out of sixteen correlations are significant and positive, meaning that they are in agreement with the observed feature value orderings in Lupyan and Dale (2010). Two feature patterns have a negative correlation, but in both cases the correlations are non-significant. Negative correlations can arise because of statistical fluctuation due to the current selection of L1s in the present study.

Table 2. 28 WALS feature numbers, the number of languages with the respective feature available, and correlations between adjusted proficiency scores and three different structural measures. A blank cell indicates that no language is either more or less complex than Dutch for that feature

WALS No.	Languages	Similarity	Increase	Decrease
100	30	.68***	.68***	
102	30	.59***	.68***	-.05
29	25	.51**	.71***	-.23
66	39	.43**	.43**	
112	53	.36**	.33*	.09
57	37	.35*		.35*
92	46	.34*	.34*	
26	51	.31*	.31*	
73	41	.29		.29
74	45	.26		.26
20	20	.26	.26	
75	44	.23		.23
22	23	.22	.33	.02
67	39	.19		.19
76	44	.18		.18
41	30	.16	-.32	.36*
65	39	.14	.14	
77	32	.12	.12	
36	37	.09		.09
28	25	.06	.06	
70	51	.04	.04	
59	24	.03		.03
38	40	.00	-.05	.14
49	44	-.28		-.28
98	23	-.22		-.22
48	30	-.13		-.13
37	43	-.09	.06	-.17
101	46	-.06		-.06

*Signif. codes: ***: p ⟨ .001, **: p ⟨ .01, *: p ⟨ .05*

Table 3. The feature hierarchy orders features with the highest impact from high to low. Impact is based on Pearson correlations between predicted and observed differences in morphological complexity between the L1s and Dutch (L2). The patterns in the first column point out which feature values are considered less complex (⟨) than the value of Dutch (always in last position)

Short Description (WALS no.), tested pattern	r
1. Syncretism in Verbal Person/Number Marking (29), none ⟨ syncretic	.71
2. Alignment of Verbal Person Marking (100), neutral (absent) ⟨ accusative	.68
3. Person Marking on Verbs (102), no person marking ⟨ agent only	.68
4. Past Tense (66), no past tense ⟨ past tense	.43
5. Polar Question Coding (92), question particle ⟨ no question particle	.34
6. Coding of Negation (112), word/affix/double ⟨ negative particle	.33
7. Inflectional Synthesis of the Verb (22), 0–1 ⟨ 2–3 categories per word	.33
8. Inflectional Morphology (26), little affixation ⟨ strongly suffixing	.31
9. Fusion of Inflectional Formatives (20), isolating ⟨ concatenating	.26
10. Perfective/Imperfective (65), no grammatical marking ⟨ grammatical marking	.14
11. Coding of Evidentiality (77), no evidential ⟨ indirect only	.12
12. Case Syncretism (28), no case marking/core and non-core cases ⟨ core only	.06
13. Definite Articles (37), no articles/demonstrative word ⟨ word distinct from demonstrative	.06
14. Morphological Imperative (70), no imperatives/singular/plural ⟨ 2nd person number-neutral	.04
15. Indefinite Articles (38), no articles/indefinite word same as 'one' ⟨ indefinite word distinct from 'one'	-.05
16. Distance Distinctions in Demonstratives (41), no distance contrast ⟨ two-way contrast	-.32

The correlations for increasing complexity suggest that the L2 learnability of Dutch is most hampered when the L1 has no syncretism in verbal person/number marking (no. 29), no person marking on the verb (no. 100), and/or no agreement marking at all (no. 102). In addition, L2 learnability of Dutch is affected when the L1 has no past tense (no. 66), has a specialized lexical unit for marking a question (no. 92), and encodes negation with words, verbal affixes, or a double-negative construction rather than a negative particle 'not' (no. 112). To a lesser extent, L2 learnability is likely to be hampered when the L1 has predominantly isolating morphology (no. 26). Some correlations did not reach significance, for example: making use of lexical items more than morphology (no. 20), morphologically marking the perfective/imperfective distinction (no. 65), or the number of possible verb categories (no. 22).

What may we expect for decreasing morphological complexity? Learning to make use of less complex morphology in an L2 seems less difficult than learning to make use of more complex morphology. If this weren't the case, linguistic structures of esoteric languages should be characterized by less complex morphology, which is not true. Acquiring a language with less complex morphology than is present in the learner's native language should be relatively easier than acquiring a language with more complex morphology. The implication is that decreasing complexity should correlate less strongly, if at all, with proficiency scores than increasing complexity.

The third column in Table 2 gives the correlations between proficiency scores and decreasing complexity. For decreasing morphological complexity, correlations for 19 features are available. Two features have a significant positive correlation. The first one, with the highest correlation, comprises distinctions in demonstratives related to distance (no. 41, $r = .36$, $p \langle .05$). The second feature is affixal possessive marking (no. 57, $r = .35$, $p \langle .05$), which Dutch does not have. Unlearning to code possessives thus has a significant positive effect—that is, it makes it harder to acquire L2 Dutch. Furthermore, although the other patterns are non-significant by themselves, together they may still have an effect: the seven negatively correlating patterns could suggest that a decrease in complexity is beneficial to learning, instead of adding difficulty. In all, only few significantly correlating feature patterns based on decreasing morphological complexity are found. This strengthens the evidence for the importance of increasing complexity in L2 learning.

4.2. *Combining Feature Patterns*

By combining feature patterns, a distance measure can be developed to assess the general effect of the complete set of relevant features. Several methods exist to optimize the weighting of the features involved. We decided to avoid any suggestion of maximizing our results by taking a mechanical approach based on the sample correlations we found in our data. To this end, we computed overall scores by weighting all relevant features by their correlations. We only included languages for which more than five feature values were available, reducing the subset of languages to 49. The three resulting distance measures for similarity, increasing complexity, and decreasing complexity give the distances from either Indo-European or non-Indo-European L1s to Dutch. For example, the maximum observed increasing complexity score before dividing is 4.289 for Vietnamese, which is the sum of all 16 correlations in Table 2 with four exceptions: one missing value and three equally complex feature values. Vietnamese has articles, makes perfective/imperfective distinctions, and has distance distinctions in demonstratives. In all other feature correlations, Vietnamese is less complex than Dutch. Dividing the weighted sum by the number of available features (15), Vietnamese gets a score of .286 for increasing morphological complexity. The larger the distance score, the more complex Dutch is as compared to an L1.

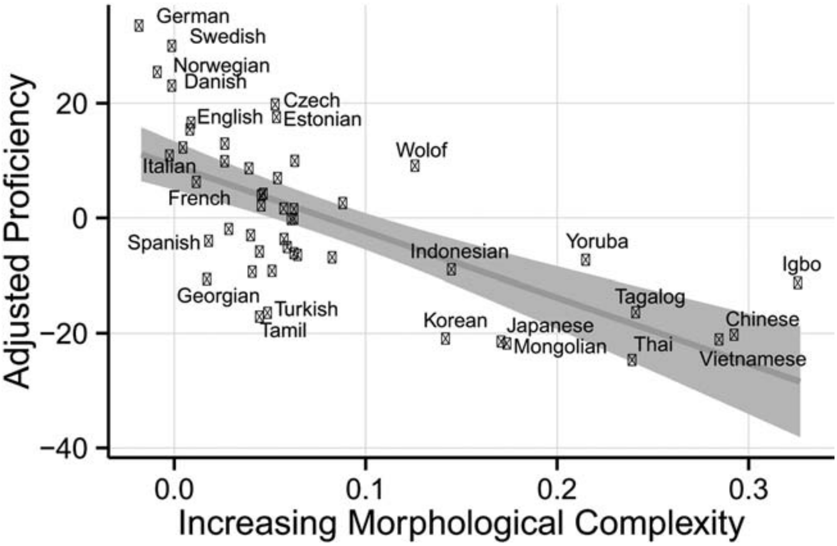

Figure 5. The relationship between speaking proficiency scores (y-axis) and a weighted sum of features that are less complex in the L1 than in Dutch (x-axis)

Having computed the three distance measures, we find that similarity and proficiency are strongly correlated (r = -.77, p ⟨ .0001), and similarity is more strongly correlated with increasing complexity (r = .78, p ⟨ .0001) than with decreasing complexity (r = .60, p ⟨ .0001). Increasing complexity correlates more strongly with proficiency (r = -.67, p ⟨ .0001) than decreasing complexity (r = -.45, p ⟨ .005). Decreasing and increasing complexity are not significantly correlated (r = .23, p = .11). Figure 5 shows a scatterplot of adjusted L2 proficiency scores and increasing complexity from the L1 perspective. “0” on the x-axis means that an L1 has more or exactly the same degree of morphological complexity as compared to Dutch. This is why most Germanic languages are situated here. Further to the right on the x-axis are L1s with a lower degree of morphological complexity. These languages are mostly not related to Dutch, like Vietnamese. Many Indo-European languages are clustered together. The grey line is a linear fit to the adjusted proficiency scores with 95 % confidence intervals added. The linear fit explains a substantial amount (45 %) of the variance in proficiency.

We compare maximum LMEMs to exactly the same model without the respective predictor added (a null model) by means of likelihood ratio tests. A maximal LMEM is a mixed effects model with random intercepts and slopes (Barr et al., 2013). There is a significant effect of similarity on proficiency ($\chi^2(1)$ = 6.86, p ⟨

.01). When replacing similarity by increasing complexity, the effect is still significant ($\chi^2(1) = 5.02$, $p \langle .05$). However, when replacing similarity with decreasing complexity, the effect is not significant anymore ($\chi^2(1) = 0.28$, $p = .60$).

A maximal LMEM with similarity as a fixed effect is not significantly more likely than a maximal LMEM with increasing complexity as fixed effect (evidence ratio of 6.9). An evidence ratio (Spiess, 2013) of 10 or more indicates strong evidence. This means that increasing complexity explains the same amount of variance in proficiency scores as similarity. On the other hand, decreasing complexity provides the least evidence: the similarity model is 34,454.5 times more likely than a decreasing complexity model. Combining increasing and decreasing complexity does not lead to a better model than similarity alone (evidence ratio of 1.9). We conclude that cross-linguistic morphological similarity effects seem to be largely built up from the degree of increasing morphological complexity in the L2.

Contrasting increasing with decreasing complexity, we find that a LMEM model with increasing complexity is significantly more likely than a decreasing complexity model (evidence ratio of 4,977.3). Adding increasing complexity to a model containing decreasing complexity (and random slopes for increasing complexity) improves the model significantly ($\chi^2(1)= 5.36$, $p \langle .05$), whereas adding decreasing complexity to a model containing increasing complexity does not ($\chi^2(1)= .82$, $p = .366$). The distance measure based on increasing morphological complexity seems to overshadow and even nullify decreasing complexity.

4.3. *Family Bias*

The relation between complexity and proficiency may be moderated by a family bias. The LMEMs reported on thus far include estimated family-specific random intercepts and slopes for each language family with two or more languages available. The family-specific slopes are shown as dotted lines in Fig. 6. The slope of the solid line, which is constant across families, is taken from a model without random slopes. The random slopes are consistently positive across families. The effect is relatively strong for Indo-European and Uralic L1s, as the dotted line for the Indo-European and Uralic families are steeper than the solid slope. This indicates a bias for a relatively high degree of similar morphology, which is unsurprising as Dutch is Indo-European itself. The steep slope for Uralic may have been affected by the Indo-European estimate, as only 3 Uralic languages are available. For the other three families there is a lower by-family adjustment to the slope, indicating that, within these families, a larger complexity difference is needed for the same difference in adjusted proficiency. Indo-European and Uralic L1s also have a higher random intercept than other L1s, meaning that they are expected to perform better irrespective of their morphological complexity.

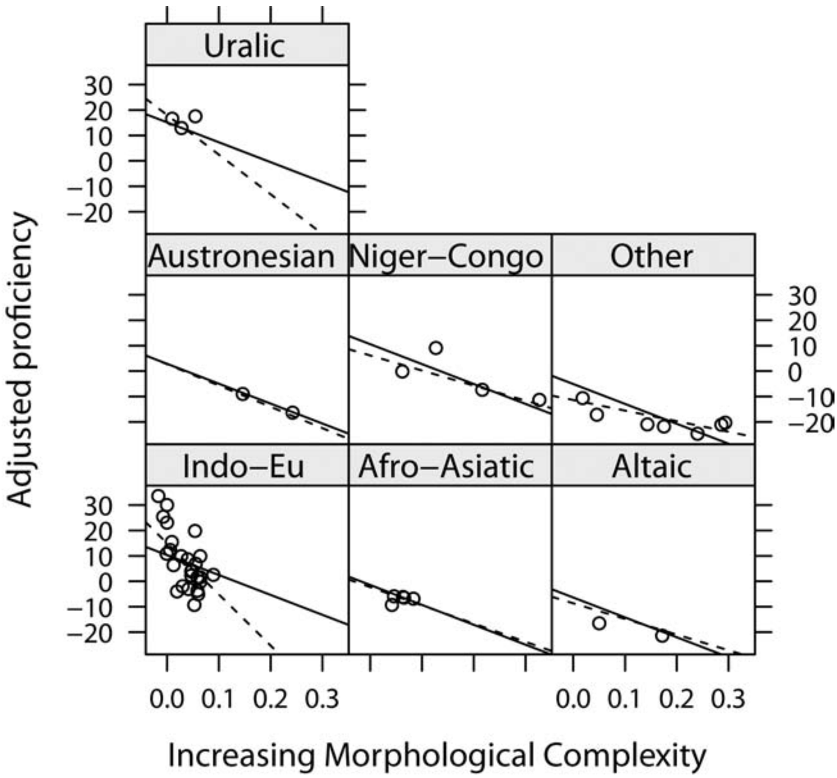

Figure 6. Between-family variation as estimated by a random intercepts and slopes model (dotted lines) and a random-intercept model (solid lines). The label "Other" contains L1s from families with 1 or 2 languages available in our sample

The relation between complexity and proficiency varies across language families as the variance across neither the random slopes nor intercepts is zero. Figure 6 shows that the intercepts differ along the y-axis and the slopes differ in steepness. However, the estimated random slopes are highly (but not perfectly) collinear with the estimated random intercepts (r = -.805). The random slope for Uralic languages is less steep than predicted by the random intercepts alone, and the slope for Niger-Congo languages is steeper than predicted by the random intercepts alone. The 95 % confidence intervals for the random intercepts of Uralic, Indo-European, and the "Other" category (families with only one or two languages available, see Fig. 6) do not contain 0, indicating that these random intercepts are different from the fixed intercept. Similarly, the 95 % confidence intervals for the random slopes of Indo-European and the "Other" category do not contain 0, indicating that these slopes are different from the fixed slope. The family bias is not strong enough (in the present set of languages) to reverse directions of any of the random slopes. In a different or larger sample of languages, family bias may play a more critical role.

4.4. *Lexical Distance*

The previous paragraphs show that morphological similarity and increasing complexity are closely related. However, it is still unclear whether the observed effect of increasing complexity cannot be reduced to a lexical distance effect. In other words, we need to investigate whether increasing complexity explains some variance that is not explained by the lexical distance model alone.

Lexical distances are a successful model of similarity effects in L2 learnability (see Section 2). Here, we use lexical distances as computed by summing over inferred branch lengths from an Indo-European language family tree (Gray and Atkinson, 2003). Missing values are replaced with inferred distances from the ASJP tree (Wichmann et al., 2010), as calculated by applying the Levenshtein-based LDND distance measure to version 13 of the ASJP database (Schepens et al., 2013). The set of 49 L1s contains 26 Indo-European L1s. Correlating lexical distances with proficiency scores reveals that the lexical distances are more strongly correlated with proficiency ($r = -.80$, $p \langle .0001$) than similarity ($r = -.65$, $p \langle .001$), increasing complexity ($r = -.68$, $p \langle .0001$), and decreasing complexity ($r = -.15$, $p = .4588$).

Within Indo-European languages, lexical distance ($F(1) = 43.71$, $p \langle .0001$), similarity ($F(1) = 17.32$, $p \langle .0001$) and increasing complexity ($F(1) = 20.37$, $p \langle .0001$) are significant predictors of L2 proficiency scores, while decreasing complexity is not ($F(1) = 0.57$, $p = .4588$). Lexical distance is a better model than all of the three morphological measures (evidence ratios for similarity: 613.4, for increasing complexity: 244.1, and for decreasing complexity: 530,273.7). Within Indo-European languages, similarity and increasing complexity models are both equally likely (evidence ratio of 2.5), while both being better than decreasing complexity (evidence ratio for similarity: 864.4, increasing complexity: 2172.5). Lexical distance is a significant addition to all three morphological measures used within this language family (similarity: $F(1) = 17.9$, $p \langle .0001$, increasing complexity: $F(1) = 15.1$, $p \langle .0001$, decreasing complexity: $F(1) = 40.6$, $p \langle .0001$), whereas none of the three morphological measures adds significantly to lexical distance (similarity: $F(1) = 1.9$, $p \langle .1779$, increasing complexity: $F(1) = 2.0$, $p \langle .1726$, decreasing complexity: $F(1) = 0.1$, $p \langle .7662$). Within Indo-European languages, lexical distance is thus a better model than either similarity or increasing complexity. Decreasing complexity is worst.

Lexical distances between non-Indo-European languages and Dutch are not available in Gray and Atkinson (2003). In order to incorporate non-Indo-European languages, we assume that their distance to Dutch is maximal. The maximal lexical distance in our subset of Indo-European languages is the distance of Albanian to Dutch. The correlation of lexical distance with proficiency scores (r = -.73, $p \langle .0001$) is similar to the correlations of similarity and increasing complexity with proficiency scores (similarity: r = -.77, $p \langle .0001$, increasing complexity: r = -.67, $p \langle .0001$, decreasing complexity: r = -.45, $p \langle .0001$). Adding similarity ($\chi^2(1) = 6.05, p \langle .05$) and increasing complexity ($\chi^2(1) = 9.05, p \langle .01$) to a lexical distance model improves model fit significantly, but adding decreasing complexity makes no difference ($\chi^2(1) = 0.68, p = .41$). Vice versa, adding lexical distance to either a similarity model ($\chi^2(1) = 1.07, p \langle .302$) or an increasing complexity model ($\chi^2(1) = 2.09, p \langle .148$) does not improve the model significantly, but adding lexical distance to a decreasing model almost reached .05 ($\chi^2(1) = 3.84$, $p = .0501$), indicating that similarity and increasing complexity already account for lexical distance. The best model in a sample including non-Indo-European languages is the lexical distance model, as it is more likely than all three morphological models (evidence ratio for similarity: 218.7, increasing complexity: 1,513.8, decreasing complexity: 7,534,621). We already saw above that there is no strong evidence for favoring the similarity model above increasing complexity (evidence ratio of 6.9).

We conclude that adding non-Indo-European languages enhances the role of increasing morphological complexity in explaining L2 learnability differences. Lexical differences can no longer account for distances between Indo-European (Dutch) and non-Indo-European languages. This outcome strengthens the pivotal role of morphology. The effect of increasing morphological complexity is also seen within the Indo-European language family, as indicated by the high correlation between lexical and morphological distance for the Indo-European languages.

5. Discussion and Conclusion

This study investigated the relation between proficiency measures of adult language learning and cross-linguistic differences in morphological similarity and complexity between 49 different L1s and L2 Dutch. Most of the morphological complexity patterns observed by Lupyan and Dale (2010) were also present across L2 learners of Dutch. To our knowledge, no other study investigates systematically across a large number of languages to what extent morphological similarity and complexity determine L2 proficiency. We used the notion of L2 learnability to capture L1 properties that co-determine adult L2 learning. Our measures were three morphological measures based on similarity, increasing complexity and decreasing complexity. The study employed L2 speaking proficiency scores as a new type of data in the study of cultural-evolutionary mechanisms in language structure (Nettle, 2012). In the present section, the results are discussed with respect to the relation between morphological complexity and L2 learnability, with respect to variation across lineages, and with respect to other measures of linguistic differences.

First of all, morphological similarity correlated significantly with proficiency in eight out of 28 features. Seven out of these eight correlations are a result of increasing complexity and one correlation is a result of decreasing complexity. For these seven features, L2 learnability is lower, the less morphologically complex the L1 is compared to the L2. See Tables 2 and 3 above for the feature-specific correlations.

An overall measure for morphological similarity as computed by combining feature-specific correlations yields a correlation of -.77 ($p < .001$). An overall measure of increasing morphological complexity correlates more strongly with L2 Dutch proficiency scores ($r = -.67$, $p < .001$) than decreasing complexity does ($r = -.45$, $p < .005$). We did not try to optimize these results, for instance by excluding simpler or more complex features of Dutch that correlate with high L2 Dutch proficiency. We wanted to include the whole set to test if morphology has an overall effect on L2 learnability. The individual outcomes for the separate features reflect the overall pattern with similarity and increasing complexity as stronger effects, decreasing complexity being the weaker component. The outcomes provide confirmatory evidence for the validity of the cross-linguistic patterns of morphological complexity that Lupyan and Dale (2010) observed.

The overall outcomes turn out to be robust when controlling for language family biases with LMEMs: an increasing complexity model is 4,977.3 times more likely than a decreasing complexity model. Replacing similarity with increasing complexity (neglecting all differences due to decreasing complexity) does not result in worse model fit. Thus, learning complex morphology seems to be more difficult for adult language learners with less morphologically complex L1s. The finding is consistent with conclusions from psycholinguistic studies (Ionin and Wexler, 2002; Lardiere, 1998; Mitchell and Myles, 2004), e.g., in terms of use of articles, case systems, past/future tense, etc. The effect of morphological complexity in particular remained consistently present and constant when controlling for family-specific biases.

What is the relationship between the morphological distance measures and lexical distance? A comparison between increasing complexity and lexical distance suggests that increasing complexity adds independent explanatory value to a lexical account of L2 learnability. Adding increasing complexity to a lexical distance model improved model fit significantly ($\chi^2(1) = 9.05$, $p < .01$). The complexity measure is more versatile than the lexical distance measure because the morphological measure can be employed for all language families. Again, substituting complexity for similarity does not result in a better model fit. It was concluded that morphological and lexical distance measures are closely related.

Does the way the L1 impacts learning an L2 explain cultural-evolutionary mechanisms of language variation and change? In a social setting, dialect structures are being influenced by incomplete transmission due to adult learning (Labov, 1972). Quantitative diachronic investigations of the role of L2 learning, however, suffer from a lack of data to determine the mechanisms in more detail: "The results need more sophisticated multivariate and comparative analysis, and perhaps most pressingly, the cultural-evolutionary mechanisms involved need to be isolated and identified" (Nettle, 2012). Our findings can be helpful. They indicate that transmission of complex morphology in adults is hampered, which is in line with experimental and longitudinal studies of adult L2 learning of morphosyntax (Birdsong and Molis, 2001; Flege et al., 1999). Moreover, adult L2 learning depends on the complexity of a learner's L1 and the complexity of the target language.

Historical linguists have argued that some languages are morphologically more complex than others because of cultural-evolutionary mechanisms that accommodate adult language learning (McWhorter, 2011; Trudgill, 2011). It may be the case that languages gradually adapt to the common cultural practice of speaking foreign languages, similarly to linguistic adaptation to growth of literacy (Levinson and Gray, 2012). Interestingly, the impact of increasing morphological complexity is not equal across all morphological features. Our outcomes provide a rank order or hierarchy of feature impact. Morphological differences low in the hierarchy of impact may be more accessible for transfer, and features higher up in the hierarchy may be more likely to cause substrate effects. The L2 literature offers evidence that supports this hierarchy. For example, Oldenkamp (2013) concludes that verbal inflection of Arabic, Chinese and Turkish learners of Dutch influences production of L2 Dutch constructions with verbal inflection.

The concept of L2 learnability overlaps with cross-linguistic influence (CLI), but both are useful for different theoretical discussions. L2 learnability is a scale measuring the extent to which L2 proficiency depends on the L1 across learners. At the level of the learner, the L2 literature focuses on differences between L2 and L1 proficiency, e.g. differences due to critical period, CLI, and transfer effects. At the level of language structure, we are interested in whether one language is easier to learn as an L2 than another. Some L1s are more complex than others (e.g. with respect to morphology), which may affect how difficult it really is to learn an L2 with a complex morphology. We find that this is, on average, the case across the different L1s included in our study. So, in a general sense, morphological complexity plays a decisive role, although the extent to which it determines L2 learnability depends on the relative differences between L1 and

L2 morphological complexity. In addition, the concept of L2 learnability entails that L1 acquisition has been completed before the onset of L2 learning. In other words, L2 learnability applies to successive language learning.

The data that are used here provide a unique possibility to assess, on a large scale, quantitative effects of structural differences between L1s. The Dutch-specific proficiency scores are affected by demographic and geographical factors. In a pre-analysis, we controlled for several individual differences. Furthermore, we expect that our findings for L2 Dutch will also play a role in studying the acquisition of other second languages. L2 learnability is not a symmetric notion for all pairs of languages: it depends on which language is the L1 and which language is the L2. It is an empirical question to what extent the internal feature weights may need to be reconfigured for testing on different L1s or L2s. Including additional L2s will give us data to investigate in more detail which morphological features stand out as complex and which do not, and give us a better insight into the complexity of linguistic structures.

A large sample of many linguistic features from a wide variety of languages is provided by WALS. Lupyan and Dale (2010) used a broad selection of features to assess the correlation between population size and the use of morphological vs. lexical encoding strategies. The present study assesses whether Dutch L2 learnability is associated with these same observed feature patterns, as ranging from lexical to morphological. As it turns out, the feature patterns that Lupyan and Dale observed correlate strongly with L2 learnability with only few exceptions. Although most of the observed patterns reported in Lupyan and Dale indicate high complexity in small languages, plurality was one of the features in their study that did not follow this trend. However, in our study, we found that morphological plurality marking actually is harder to learn in an L2. These conflicting findings may be due to differences in the subsets of languages between the two studies. With respect to distance distinctions (no. 41) and coding of possessives (no. 57), the two features for which our study showed decreasing, rather than increasing complexity to be harder to learn, the subset of languages might have played a role as well.

LMEMs are a conservative way of modeling lineage and language-specific factors that affect whether or not a language is more or less complex than Dutch morphologically. On average, even after adjusting for random between-family variation, the features with increasing morphological complexity still correlate strongly with L2 learnability. It remains an open question to what extent the variation in specific lineages supports this claim. The data available for the learnability of Dutch as an L2 and the structural configurations of the L1s of the learners together provide a large-scale quantitative source of evidence for the hypothesis that morphological complexity across languages may be constrained by adult language learning over longer periods of time.

Acknowledgments

We are grateful to the secretary of the State Examination of Dutch as a Second Language for providing the data used in our study. We thank Kate Bellamy, Michael Dunn, Michele Gubian, Gary Lupyan, Theo Bongaerts, and three anonymous reviewers for comments on an earlier version of this manuscript and editorial help.

References

Andersen, Henning. 1988. Center and periphery: Adoption, diffusion and spread. In Jacek Fisiak (ed.), *Historical Dialectology: Regional and Social*, 39–83. Berlin: Walter de Gruyter.

Atkinson, Quentin D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332: 346–349.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278.

Bates, Douglas, Martin Maechler, and Ben Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. Downloadable at http://cran.r-project.org/web/packages/lme4/index.html (accessed December 1, 2013).

Bentz, Christian and Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3(1): 1–27.

Birdsong, David and Michelle Molis. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44(2): 235–249.

Blom, Elma, Daniela Polisšenská, and Fred Weerman. 2006. Effects of age on the acquisition of agreement inflection. *Morphology* 16(2): 313–336.

Braunmüller, Kurt. 1990. Komplexe Flexionssysteme—(k)ein Problem für die Natürlichkeitstheorie? *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43(5): 625–635.

Chiswick, Barry R. and Paul W. Miller. 2005. Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development* 26(1): 1–11.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Dahl, Östen. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.

Dauenhauer, Nora M. and Richard Dauenhauer. 1998. Technical, emotional, and ideological issues in reversing language shift: Examples from Southeast Alaska. In Lenore A. Grenoble and Lindsay J. Whaley (eds.), *Endangered Languages: Current Issues and Future Prospects*, 57–98. Cambridge: Cambridge University Press.

Dediu, Dan and Robert D. Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *PNAS* 104(26): 10944–10949.

Dediu, Dan and Stephen C. Levinson. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLoS ONE* 7(9): e45198.

Dryer, Matthew S. and Martin Haspelmath (eds.). 2011. *The World Atlas of Language Structures Online*. Max Planck Digital Library. Accessible at http://wals.info/ (accessed December 1, 2013).

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473 (7345): 79–82.

Evans, Nicholas and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(05): 429.

Flege, James E., Grace H. Yeni-Komshian, and Serena Liu. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language* 41(1): 78–104.

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st ed. Cambridge: Cambridge University Press.

Gray, Russell D. and Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426 (6965): 435–439.

Haeserijn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, and Maarten C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. 2nd ed. Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn. Accessible at http://ans.ruhosting.nl/e-ans/index.html (accessed December 1, 2013).

Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.

Hunley, Keith, Michael Dunn, Eva Lindström, Ger Reesink, Angela Terrill, Meghan E. Healy, George Koki, Françoise R. Friedlaender, and Jonathan S. Friedlaender. 2008. Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS Genetics* 4(10): e1000239.

Ionin, Tania and Kenneth Wexler. 2002. Why is "is" easier than "-s"?: Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research* 18(2): 95–136.

Jaeger, T. Florian, Peter Graff, William Croft, and Daniel Pontillo. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15(2): 281–319.

Johnson, Jacqueline S. and Elissa L. Newport. 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.

Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5(3): 206–213.

Kortmann, Bernd and Benedikt Szmrecsanyi. 2012. *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: Mouton de Gruyter.

Kusters, Wouter. 2003. *Linguistic Complexity. The Influence of Social Change on Verbal Inflection*. PhD dissertation, Universiteit Leiden. Downloadable at http://www.lotpublications.nl/publish/articles/000501/bookpart.pdf (accessed December 1, 2013).

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lado, Robert. 1957. *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Michigan: University of Michigan Press.

Lardiere, Donna. 1998. Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research* 14(4): 359–375.

Levinson, Stephen C. and Russell D. Gray. 2012. Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences* 16(3): 167–173.

Levy, Roger and Hal Daumé. 2011. Computational methods are invaluable for typology, but the models must match the questions. *Linguistic Typology* 15(2): 393–399.

Lupyan, Gary and Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1): e8559.

McDonald, Janet L. 2000. Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics* 21(03): 395–423.

McWhorter, John H. 2002. What happened to English? *Diachronica* 19(2): 217–272.

—— 2007. *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars*. New York: Oxford University Press.

—— 2011. *Linguistic Simplicity and Complexity: Why Do Languages Undress?* Boston: Mouton de Gruyter.

Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson. 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins Publishing Company.

Mitchell, Rosamond and Florence Myles. 2004. *Second Language Learning Theories*. London: Arnold Publishers.

Nettle, Daniel. 2012. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1597): 1829–1836.

Odlin, Terence. 1989. *Language Transfer: Cross-linguistic Influence in Language Learning*. New York: Cambridge University Press.

Oldenkamp, Loes. 2013. *The Trouble with Inflection of Adult Learners of Dutch. A Study on the L1-L2 Interplay of Morphosyntactic and Phonetic-Phonological Factors*. PhD dissertation, Radboud University Nijmegen.

Prévost, Philippe and Lydia White. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16(2): 103–133.

R Core Team. 2013. R: A language and environment for statistical computing. Downloadable at http://www.r-project.org/ (accessed December 1, 2013).

Sampson, Geoffrey, David Gil, and Peter Trudgill. 2009. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.

Schepens, Job, Frans Van der Slik, and Roeland Van Hout. 2013. The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In Lars Borin and Anju Saxena (eds.), *Comparing Approaches to Measuring Linguistic Differences*, 199–230. Berlin: Mouton de Gruyter.

Schepens, Job, Frans Van der Slik, and Roeland Van Hout. Submitted. The L2 impact on acquiring Dutch as a L3: The L2 distance effect. In Dirk Speelman, Kris Heylen, and Dirk Geeraerts (eds.), *Mixed Effects Regression Models in Linguistics*. Berlin: Springer.

Smith, Kenny and Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116: 444–449.

Spiess, Andrej Nikolai. 2013. qpcR: Modelling and analysis of real-time PCR data. Downloadable at http://cran.r-project.org/web/packages/qpcR/index.html (accessed December 1, 2013).

Towell, Richard and Roger D. Hawkins. 1994. *Approaches to Second Language Acquisition*. Clevedon, UK: Multilingual Matters.

Trudgill, Peter. 1983. *On Dialect: Social and Geographical Perspectives*. Oxford: Basil Blackwell.

—— 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5(2–3): 371–374.
—— 2002. *Sociolinguistic Variation and Change*. Edinburgh: Edinburgh University Press.
—— 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.
Van der Slik, Frans W.P. 2010. Acquisition of Dutch as a second language. *Studies in Second Language Acquisition* 32(03): 401–432.
Weinreich, Uriel. 1963. *Languages in Contact*. The Hague: Mouton.
Wichmann, Søren. In press. Diachronic stability and typology. In Claire Bowern and Bethwyn Evans (eds.), *Handbook of Historical Linguistics*. London: Routledge.
Wichmann, Søren and Eric W. Holman. 2009. Population size and rates of language change. *Human Biology* 81(2–3): 259–274.
Wichmann, Søren, Eric W. Holman, Dik Bakker, and Cecil H. Brown. 2010. Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications* 389(17): 3632–3639.
Wichmann, Søren, Dietrich Stauffer, Christian Schulze, and Eric W. Holman. 2008. Do language change rates depend on population size? *Advances in Complex Systems* 11(03): 357–369.
World Bank. 2011. School enrollment, secondary (% gross). UNESCO, Institute for Statistics report. Accessible at http://data.worldbank.org/indicator/SE.SEC.ENRR (accessed January 3, 2013).