

EMPIRICAL STUDY

L1 and L2 Distance Effects in Learning L3 Dutch

Job J. Schepens,^{a,b,c} Frans van der Slik,^a and
Roeland van Hout^a

^aRadboud University, ^bMax Planck Institute for Psycholinguistics, and ^cUniversity of Rochester

Many people speak more than two languages. How do languages acquired earlier affect the learnability of additional languages? We show that linguistic distances between speakers' first (L1) and second (L2) languages and their third (L3) language play a role. Larger distances from the L1 to the L3 and from the L2 to the L3 correlate with lower degrees of L3 learnability. The evidence comes from L3 Dutch speaking proficiency test scores obtained by candidates who speak a diverse set of L1s and L2s. Lexical and morphological distances between the L1s of the learners and Dutch explained 47.7% of the variation in proficiency scores. Lexical and morphological distances between the L2s of the learners and Dutch explained 32.4% of the variation in proficiency scores in multilingual learners. Cross-linguistic differences require language learners to bridge varying linguistic gaps between their L1 and L2 competences and the target language.

Keywords cross-linguistic difference; L1 distance effect; L2 distance effect; lexicon; morphology; L3 learning

Introduction

Besides factors such as learners' age and amount of exposure, learning an additional language appears to be affected by linguistic distances between

This article greatly benefited from comments by Jared Linck and two other reviewers, as well as Pavel Trofimovich, Theo Bongeaerts, Michael Dunn, and the HLP/Jaeger lab. We have no competing interests to report. Example elicitation instruments can be downloaded from https://www.hetcvte.nl/item/voorbeeldexamens_en. Aggregated data from this study are accessible in the Supporting Information online.

Correspondence concerning this article should be addressed to Job J. Schepens, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: schepens@mpib-berlin.mpg.de

previously learned languages and the target language. Linguistic distances can be defined through measures that quantify how distinct linguistic structures are, for example, at the lexical or morphological level. The larger the linguistic distance between a first language (L1) and a second language (L2), the lower the L2 learnability (Schepens, Van der Slik, & Van Hout, 2013a, 2013b; Van der Slik, 2010). We define L2 learnability as the degree to which the L1 facilitates or impedes the learning of a L2, such that L2 learnability characterizes learning difficulties that depend on the L1.

The concept of L2 learnability may be applied to learning or acquiring any additional language, be it a L2, third language (L3), or any subsequent language. However, this is clearly an oversimplification of how previously learned languages affect subsequent language learning. A growing number of studies targeting L3 learning have provided evidence for the role of both learners' L1 and their L2 in learning a L3. Accordingly, in L3 acquisition, the relative impact of the L1, as compared to the influence of the L2, may depend on: (a) which language is more similar to the L3 (Ahukanna, Lund, & Gentile, 1981; Rothman, 2011; Singleton, 1987); (b) the degree of learners' proficiency in the L1 and L2 (Lindqvist, 2010; Ringbom, 2007; Williams & Hammarberg, 1998); and (c) the status of the L2, as being relatively important by itself (Bardel & Falk, 2007; Falk & Bardel, 2011; Flynn, Foley, & Vinnitskaya, 2004; Hammarberg, 2001). The current study investigates the relationship between linguistic distance and L3 learning in an analysis of L3 Dutch speaking proficiency for 39,300 multilinguals, some of whom reported speaking only one language (L1) while most reported knowledge of other languages (L2s) prior to learning Dutch.

Theoretical Background

L3 Learnability

Does a L2 have an influence on L3 speaking proficiency and, if so, is its effect different from L1 influence? The L3 literature suggests that both L1 and L2 typology in relation to the L3 and the L1 and L2 proficiency levels play a role (Cenoz, 2003; Jaensch, 2013; Murphy, 2005). However, it is unclear whether L2 typological similarity is more or less important than L1 typological similarity and whether L2 proficiency is more or less important than L1 proficiency. In addition, L2 influence may differ between productive and receptive modalities, between written and spoken language use, and across learning stages. Learner-based variables, such as motivation, intelligence, years of full-time education, educational quality, age, and gender, may play

a certain role as well. The present study focuses on speaking proficiency and on language-related variables while controlling for learner-based variables as well as systematic variation in learners' countries of birth.

The current understanding of typology effects on L3 learnability suggests that a typological similarity or overlap between languages leads to positive cross-linguistic influences, both for L1 to L2 influence (Ard & Homburg, 1983; Kellerman & Sharwood Smith, 1986; Odlin, 1989), and for L2 to L3 influence (Cenoz, 2001). L1 negative transfer is more likely at lower proficiency levels (Odlin, 1989). There is not always a one-to-one correspondence between objective typological similarity and the typological similarity perceived by learners. A negative perception of typological similarity may lead to negative transfer (Jarvis & Pavlenko, 2008), overall or in a specific linguistic domain. The present study focuses on objective typological similarity.

There are at least three specific explanations for the way typological similarity influences L2 to L3 transfer. These are primarily based on the acquisition of syntactic properties, such as relative clauses (Flynn et al., 2004), negation placement (Bardel & Falk, 2007), verb second (Bohnacker, 2006), and word order (Rothman, 2010). First, the cumulative enhancement model (Flynn et al., 2004) predicts that the effect of the L2 (a) is not absorbed by the L1, (b) is either neutral or positive, and (c) is more beneficial for learning a L3 than having no L2 at all. Second, the L2 status factor model (Bardel & Falk, 2007) and the findings of Bohnacker (2006) predict that a L2 can obscure or impede transfer effects between the L1 and the L3, depending on L2 status and L2 proficiency. This suggests a prominent role for the L2, outranking L1 influence. Third, the typological primacy model (Rothman, 2010) predicts that either the L1 or the L2 will transfer into the L3, depending on the highest typological similarity (see Table 3 for an overview of these predictions). L2 transfer has also been observed in the lexical domain for nonnative function words (De Angelis, 2005). We test the predictions of these models by investigating additive and/or interactive influences of learners' L1s and L2s on their learning of L3 Dutch.

Does the mind structure L2 knowledge in a similar way as it structures L1 knowledge? This is a relevant question because L1–L3 typological similarity can be of lower importance than L2–L3 similarity for learning a L3 (Bardel & Falk, 2007). This L1–L2 difference in transfer is hypothesized to result from the representational nature of L2 knowledge: Adults acquire L2 knowledge initially on an explicit/declarative basis before they can proceduralize it

(Ellis, 2005; Paradis, 2009; Ringbom, 2007), and this explicit/declarative knowledge, as compared to procedural L1 knowledge, may be more beneficial for L3 learning. Thus, if L1 or L2 declarative knowledge is available, it may be more likely to affect the learning of the declarative knowledge in an additional language, compared to L1 procedural knowledge (Falk, Lindqvist, & Bardel, 2013). In addition, learners develop enhanced awareness or metalinguistic skills based on their L1 and L2 knowledge (Cenoz, 2003; Jessner, 2012, 2014). However, it is unclear to what extent there is a global language-independent multilingualism factor or whether this factor is composed of additive effects of specific characteristics of the languages learned.

Linguistic Distance

We hypothesize that L1 and L2 linguistic distances affect L3 learnability. The degree to which the L2 (or its absence) facilitates or impedes learning of a specific additional language can be defined as L3 learnability and can be estimated through proficiency scores of L3 learners. Linguistic distance measures the degree of similarity between languages, which is often used for language classification (Ruhlen, 1991; Trask, 2000). Distance measures include those that focus on qualitative differences at the level of family and genus and those that target quantitative distances based on the degree of linguistic differences (Greenberg, 1956; Nichols, 1992). Qualitative notions of linguistic distance based on language-family relations (Lewis, Simons, & Fennig, 2013) are useful for language learning studies in which a small number of languages are compared. For example, a study of the influence of Basque (as L1 or L2) versus Spanish (as L1 or L2) on English as L3 shows that a Basque background has a less positive effect on learning English than Spanish, irrespective of its status as L1 or L2 (Cenoz, 2001). The exact quantitative linguistic distance between Basque and English is not straightforward to measure, but it seems obvious that Basque, an isolate language, is more distant from English than Spanish, as English and Spanish are Indo-European languages.

For a comparison across a large number of L1s and L2s, a quantitative measure of linguistic distance is required to determine the effects of linguistic differences. Semi-quantitative measures can be used, which are based on the number of levels in the family tree that languages share (Adsera & Pytlikova, 2012; Desmet, Weber, & Ortuño-Ortín, 2009; Ispording & Otten, 2014), but such crude measures cannot distinguish between the similarity of Spanish and French to English, or Basque and Chinese to English. We therefore focused on linguistic data to provide measures of linguistic distance with more detail than is possible by counting nodes in language family trees.

Recently, typological resources have become available that are useful for measuring linguistic distances. For example, basic vocabulary word lists (Dyen, Kruskal, & Black, 1992) have been used to statistically estimate the most likely time depth of Indo-European languages based on models of evolutionary language change over time (Bouckaert et al., 2012; Gray & Atkinson, 2003). Similarly, basic vocabulary data for multiple language families have been used to reconstruct language family trees based on the number of shared lexical forms (Holman et al., 2011). Both approaches compute distance measures between languages based on lexical differences. In addition, structural data (Dryer & Haspelmath, 2011) have been used for reconstructing family relationships (Dunn, Terrill, Reesink, Foley, & Levinson, 2005) and linking development of morphological differences across language families to changes in social structures (Lupyan & Dale, 2010). As adult learners often experience problems with adding additional morphology to L1 morphology (Ionin & Wexler, 2002), languages that are learned relatively often by adults may show a gradual reduction in morphological complexity over time (Trudgill, 2011).

Here we used both lexical and morphological distance measures to predict L3 learnability. For lexical distance measures, we used measures of evolutionary change within the Indo-European language family (Gray & Atkinson, 2003). We also computed morphological distances, as these overcome the limitation of having lexical distances within one language family only. Morphological distance measures are available for languages from non-Indo-European language families as well. Schepens et al. (2013a) used 29 morphological features extracted from the World Atlas of Language Structures (WALS; Dryer & Haspelmath, 2011) to construct three measures of morphological distance. They measured morphological similarity between Dutch and other languages, the degree of increasing morphological complexity from the perspective of a particular language toward Dutch, and the degree of decreasing morphological complexity from the perspective of a particular language toward Dutch. Both increasing morphological complexity and morphological similarity emerged as significantly better predictors for explaining variation in speaking proficiency scores than decreasing morphological complexity. Increasing morphological complexity could replace morphological similarity without losing explanatory value, which justifies our choice of this complexity measure. Increasing morphological complexity is based on linguistic differences that are more morphologically complex in Dutch than in a L1 or L2. In contrast to the lexical distance measure, the complexity measure is asymmetric. Therefore, the increase in morphological complexity from Chinese to Dutch is not necessarily the same as the increase from Dutch to Chinese.

A previous investigation already established both L1 morphological and L1 lexical distance effects in a study which included both monolingual and multilingual learners (Schepens et al., 2013a). Therefore, we expect that lexical and morphological measures together cover a full range of distance effects from very distant to very similar L1s and L2s in multilingual learners of L3 Dutch, including Indo-European as well as non-Indo-European languages. For example, lexically similar languages can differ in morphological complexity (e.g., English–German). Besides lexical and morphological differences, languages also differ along syntactic (Dunn, Greenhill, Levinson, & Gray, 2011) and phonological dimensions (Moran & Blasi, 2014). The study of syntactic and phonological cross-linguistic differences is, however, beyond the scope of the present investigation.

Large-Scale Studies of Speaking Proficiency

Large scale studies of speaking proficiency have to deal with four challenging issues: (a) the difficulty of accurately and validly measuring language proficiency, (b) the availability and richness of information about learners' backgrounds, (c) the number of observations, and (d) the cross-classified nature of languages and countries. This study uses the data from the official state exam of Dutch as a Second Language (STEX) to study L3 speaking proficiency. In the past, large-scale studies have relied on self-reported proficiency measures (Hakuta, Bialystok, & Wiley, 2003). One of the reasons why these can be biased is that learners compare themselves to each other (Finnie & Meng, 2005; McArthur & Siegel, 1983; Siegel, Martin, & Bruno, 2001) and can thus systematically over- or underestimate their L2 skills. For instance, correlations between self-reported measures of proficiency and quick objective measures of proficiency are not strong and vary across groups, reaching .50 for Dutch learners of English and .30 for Korean learners of English (Lemhöfer & Broersma, 2012).

In contrast, STEX is a database of language proficiency scores; it provides testing scores that have been collected through a formalized assessment protocol and features large-scale assessments that are characteristic of census data. STEX aims to objectively measure the communicative competences of L2 learners of Dutch at the B2 level of speaking proficiency of the Common European Framework of Reference for Languages (Council of Europe, 2001; Hulstijn, Schoonen, de Jong, Steinel, & Florijn, 2012). The STEX database is large enough to cope with the second and third issue discussed above, because speaking scores are available for more than 50,000 learners with information on language background and key individual characteristics. This is a large sample, as compared to similar studies (e.g., the German Socio-Economic Panel

analyzed by Isphording & Otten, 2011). However, the final issue, which pertains to the cross-classified nature of languages and countries, is quite challenging because we have to deal in our case with the cross-classification not just between the learners' L1s and their countries of birth but also between their L1s and L2s. In the sample, the L1s and L2s occur in many but definitely not all possible language combinations. This partially cross-classified structure makes it difficult to disentangle the independent contributions of each country of birth, L1, and L2. For example, approximations of the linguistic distance effect from Norwegian to Dutch will become more precise when scores for L1 Norwegian speakers from countries besides Norway would be available as well. This is a recurrent issue in cross-country observational samples (Fearon, 2003).

The Current Study

In the current study, we investigate adult language acquisition of L3 Dutch, taking into account both the L1 and, if present, the best spoken L2. Our main hypothesis is that the L2–L3 distance is a robust factor in explaining proficiency differences in L3 learning, in addition to and independent from the impact of the L1–L3 distance. Confirming this hypothesis implies that there is a robust effect of being multilingual as well, in the sense that multilinguals and monolinguals differ in L3 performance. However, does being multilingual help or hinder L3 learning? In order to answer this question, we compare multilinguals to a monolingual baseline group. Our second hypothesis is that the L2–L3 distance effect is more facilitative when the L2–L3 distance is small. However, confirming the hypothesis that there is a L2 distance effect does not indicate how strong this effect is, compared with the L1 distance effect. Our third hypothesis, therefore, states that the L2 distance effect is weaker than the L1 distance effect, as the L1 is generally learned earlier and more intensively. Our overall goal is to understand to what extent multilingual learners can benefit from general effects of being multilingual in L3 learning and to what extent the effect of L2 distance is facilitative, as compared to the effect of the L1.

To address these issues, we used the same set of speaking proficiency test scores used by Schepens et al. (2013a, 2013b), based on the STEX exam taken by immigrants from around the world; we tested the explanatory power of both lexical and morphological linguistic distance measures in predicting L3 Dutch speaking scores, controlling for confounding variables, both at the level of the individual learner (exposure, age, gender, education) and his/her country of origin (educational quality). Using the STEX data set, this study examines how learners' L1 and L2 interact in L3 learnability, thus investigating cross-linguistic influences in L3 acquisition Cenoz, Hufeisen, & Jessner (2001), and tests L1

and L2 effects as predicted in L3 learning theories. Our comparative approach is different from longitudinal studies targeting L3 learnability in adults; rather than comparing target language proficiency over time, we compare target language proficiency across language backgrounds.

Method

STEX

The STEX data come from the state exam of Dutch as a Second Language (Nederlands als Tweede Taal, NT2). Passing this exam is a formal entry requirement for Dutch universities and for starting many higher-level education jobs. The Dutch Governmental Board of Examinations provided exam results collected over a period of 15 years. The full state exam consists of a speaking, writing, listening, and reading test. The exam aims at a B2 passing level, which is upper-intermediate according to the Common European Framework of Reference for Languages (Council of Europe, 2001). The exam is mostly taken by a heterogeneous group of newcomers (150 different countries of birth). The exam requires considerable personal investment of time and money, which ensures high motivation of the candidates. A second, alternative state exam aims at a B1 passing level and is meant for nonacademic contexts; the scores from that exam are not analyzed here.

Data Sample

The study includes proficiency scores (first administration only) for 39,300 multilingual candidates ($M_{\text{age}} = 30.2$ years; $Mdn = 29.0$; 26,225 females, 13,075 males) who reported both a mother tongue and a best additional language besides their mother tongue on the questionnaire; these candidates were selected from the original pool of 50,500 candidates' scores. All candidates participated in the speaking exam between 1995 and 2010 at various locations in the Netherlands. Candidates with a country of birth, L1, or L2 with fewer than 15 individuals available were excluded; also excluded were candidates who gave missing or invalid (e.g., unreadable) answers to the questionnaire that they filled out voluntarily before the exam. This resulted in the exclusion of participants who did not report age of arrival, country of birth, or mother tongue. It was possible to determine linguistic distance for most of the L1s of the remaining candidates. Out of the 50,500 original candidates, 11,200 (including the monolinguals) had to be excluded because L1 and/or L2 linguistic distance could not be determined for them (specifically, morphological distance, see below). However, we compared L3 Dutch scores for monolinguals with the scores for multilinguals to test for general effects of speaking a L2.

The multilingual candidates spoke 56 different L1s ($M = 701.8$ speakers per language; $Mdn = 256.5$) and 35 different L2s ($M = 1,122.8$ speakers per language; $Mdn = 64$, with English accounting for 68.0% of all L2s). Following WALs (Dryer & Haspelmath, 2011), the 56 L1s come from 32 different genera that belong to 14 language families. Of these languages, 27 are non-Indo-European and 29 are Indo-European. The candidates were born in 119 different countries around the world and represented a total of 536 L1–L2 combinations (161 combinations had at least 15 candidates); 25.2% of all L1 speakers had a L2 other than the most common L2 for that L1 (which was mostly English), illustrating the cross-classified nature of the data. When candidates with English as a L2 were excluded, this value increased to 38.0%.

Tasks

Candidates performed a mix of short and long speech tasks in 30 minutes (a typical exam consists of 14 tasks), in which they needed to provide or ask for information, give instructions, and so on. For example, in the 1997 exam, candidates needed to describe and give a motivated opinion about marketing campaigns in 2 minutes. The test assesses whether candidates can respond adequately to a given situation. The instructions were simultaneously provided through headphones and on paper. The performance tasks require candidates to produce functional language. Performance is assessed in a classroom setting with computer testing carrels.

Measures

The exams are designed to measure proficiency around the passing level of 500 points. Two independent, professional examiners evaluate the spoken language for both content and accuracy. The most important content criteria are the fit of the content to the task (about 30%) and the size of the vocabulary (about 18%). The most important formal linguistic criteria are word and sentence formation (word order, verbal inflection, tense, accounting for about 28%) and pronunciation (about 12%). The remaining 12% of the criteria are related to fluency, coherence, word choice, tempo, and register. These percentages are based on the speaking exam from 1998 but are representative of a typical exam. Almost all criteria are influenced by lexical and morphological characteristics of candidates' speech. The final score is calculated from the set of ratings given by both examiners. Each of the 14 tasks is rated on two to six criteria, depending on the duration of the task. About 40% of the ratings are two-way (insufficient or sufficient) ratings and 60% are four-way ratings (insufficient, almost sufficient, sufficient, and good). The final standardized L3 Dutch speaking scores, were

computed by the language testing institution CITO using an item-response theory model based on the distributions of scores on previous exams. This standardization ensures comparability of test scores over time.

We derived several background characteristics for all participants, based on the exam information and questionnaires completed by participants. These data included participants' country of birth, date of exam, date of birth, gender, years of education, date of arrival in the Netherlands (useful to infer length of residence and age at arrival), L1, and L2. The question for the L1 was "What is your mother tongue?" and the question for the L2 was "Do you speak another language besides Dutch and your mother tongue?" and, if the person answered yes, "Which other language do you speak? If you speak more than one, name the language that you know best" (translated from Dutch). Note that "best" is not quantified in our study and probably covers varying proficiency levels in the reported L2s. The correct interpretation of the Dutch form of the expression "Do you speak another language" is whether one knows how to express oneself orally in another language. In general, it can be assumed that learners can already understand this question at the A2 level in Dutch, which is below the passing level of the current exam. For example, the A2 level describes that a learner is able to "understand sentences and frequently used expressions related to areas of most immediate relevance."¹ Clearly, the A2 Dutch level requires candidates to have acquired at least part of the Dutch morphology. Figure 1 illustrates the distributions of speaking proficiency scores using the Chinese and German native speakers, including monolinguals. German is the language closest to Dutch (besides Afrikaans); Chinese has the greatest lexical and high morphological distance from Dutch. The graph shows that a native speaker of Chinese is unlikely to obtain higher scores than the average native speaker of German.

We added measures of schooling quality, lexical distance, and morphological distance in order to explain variation across L1s, L2s, and countries of birth. Schooling quality was measured as the gross secondary school enrolment in 2006 (United Nations Educational, Scientific, and Cultural Organization, 2011), which is the ratio of total enrolment in secondary education per country (see also Schepens et al., 2013b). The first measure of linguistic distance from the L1 and the L2 to Dutch was lexical. It was defined as the degree of evolutionary change based on shared cognates between Indo-European languages and measured as the sum of branch lengths that connect both languages to each other in a phylogenetic language family tree (Gray & Atkinson, 2003). This measure can be qualified as lexical distance. A maximum lexical distance as observed in the Indo-European language family tree was used for L1s from different language

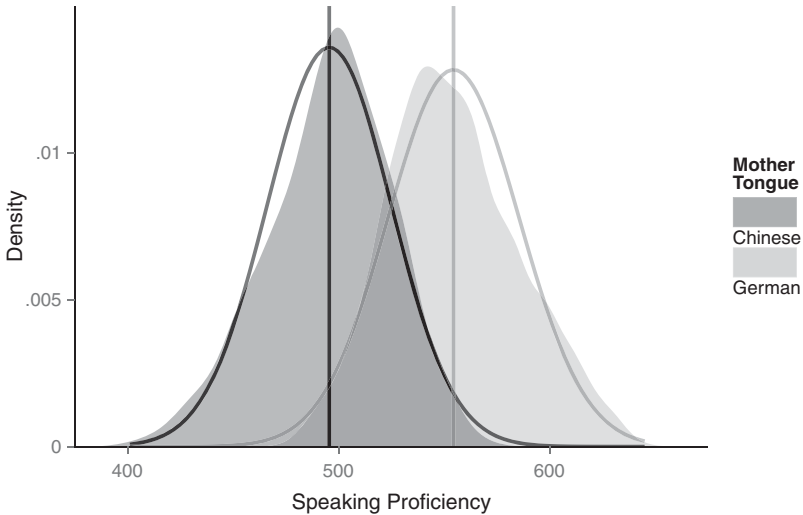


Figure 1 The distributions of German ($n = 4,773$) and Chinese ($n = 776$) speaking proficiency scores compared with normal distributions, showing differences across subgroups.

families. Lexical distances have higher values for larger distances: 0 means that the two languages share all words in the Swadesh list (Swadesh, 1952) and a higher value means that the languages share a longer branch length, effectively having a smaller lexical overlap. Lexical distances ranged from .0105 to .5950 ($M = .322$, $SD = .175$).

The second measure of linguistic distance was based on increasing morphological complexity. This measure of morphological distance was computed by comparing the complexity of Dutch to the L1 for 29 morphological features (Schepens et al., 2013a). For example, Dutch marks the feature “past tense” morphologically, whereas some languages do not (Dahl & Velupillai, 2013). The distance measure is a weighted sum of these feature differences. Among these feature differences were differences in verbal person and number marking, past tense, polar question coding, question particle, coding of negation, inflectional synthesis of the verb, degree of inflectional morphology, and so on. In contrast to the lexical distance measure, this measure also varies across non-Indo-European L1s. Due to missing feature values in the WALS data, which was used to develop morphological distance, some missing morphological distance values were set to the same distance as linguistic neighbors (e.g., Bosnian to Croatian, Ukrainian and Belarusian to Russian, Catalan to Spanish, Fulani to

Wolof, Malay to Indonesian). Morphological distances also have higher values for larger distances: 0 corresponds to no increase in morphological complexity from a given language to Dutch (e.g., for the Dutch–German language pair) and a higher value corresponds to an increase in morphological complexity (e.g., for the Dutch–Igbo language pair). Morphological distances ranged from $-.017$ to $.327$ ($M = .050$, $SD = .057$).

Several variables were tested but were eventually left out after model comparisons; these included geographical distance (from the capital of the country of birth to Amsterdam), the Greenberg diversity index (Greenberg, 1956),² whether or not the L1 or L2 could be classified as Germanic or Indo-European (Adsera & Pytlikova, 2012), peer group size (number of learners from country of birth taking part in the exam), total number of citizens in the country of birth, and gross domestic product per capita. These factors did not significantly influence L3 speaking proficiency testing scores.

Analysis

We used a mixed-effects regression approach to predict variation in L3 proficiency scores across L1s and L2s. The approach summarizes over individual differences, resulting in aggregate scores across L1s and L2s that are controlled for potentially moderating factors such as age, exposure, education, gender, differences across countries, and differences resulting from specific L1–L2 combinations. The speaking scores were analyzed by using cross-classified random effect models (CCREM) in R (R Core Team, 2013) and fitted with *lme4* (Bates, Mächler, Bolker, & Walker, 2014), which is supplementary to R. All candidates, irrespective of difference in language background, were included in one CRREM analysis by treating country of birth, L1, L2, and L1–L2 combinations as random effects (Schepens, Van der Slik, & Van Hout, in press). We keep this random-effect structure constant throughout the rest of the article (except for the final analysis).

Effects of country background on L3 learnability needed to be disentangled from language-specific influence (Fearon, 2003). As many countries in the world are to some extent fractionalized in terms of languages, the most frequently used language in a country is not necessarily the L1 or even the L2 of a candidate. Furthermore, intercountry linguistic differences can reach the level of completely different language families. A generalization from the country level to the language level is likely to neglect any existing linguistic diversity. We chose *lme4* instead of *nlme* (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2013) because *lme4* can fit models with partially crossed random effects

in large unbalanced data, which is necessary given that not every language is spoken in every country.

CCREMs assume independence between crossed random effects (Baayen, Davidson, & Bates, 2008). The inclusion of random effects for both languages and countries only associates variation in proficiency scores with either languages or countries if that variation is unambiguously associated with a language or country, assuming that by-country and by-language variation is independent. This random-effect structure results in conservative lower bound estimations of by-L1 variation and by-country variation (the minimal unexplained by-L1 and by-country variance that can be observed in the data).

It is often a mistake to assume that random effects in cross-classified models are completely mutually independent in unbalanced data. The interdependency between languages and countries is not further investigated here. However, the degree of interdependency between L1s and L2s is potentially important for establishing whether L1 and L2 distance effects are indeed additive and independent. This is the reason why we also included L1–L2 combinations as a L1-by-L2 random interaction effect. In order to further control for dependency across the random effects, we decided to include random slopes for L1 linguistic distance across L2s and L2 linguistic distance across L1s. However, such models did not converge, likely because of the sparse crossing between L1s and L2s. Random slopes for the distance effects across countries also did not result in converged models.

We tested fixed predictors using a semi-backward elimination procedure (Baayen et al., 2008), in which we always retested a predictor after another predictor had been removed. We performed model comparisons with likelihood ratio tests as well as Akaike's Information Criterion (AIC) comparisons for nested models. We computed evidence ratios based on the AIC for non-nested models (Spiess, 2013). The higher the evidence ratio, the more evidence for a particular model (in favor of the model being compared to another one). We removed 787 outliers by excluding multilingual candidates who had a standardized residual for the speaking proficiency measure higher than 2.5 standard deviations from 0, amounting to 2.0% of the data. Residuals were based on a mixed-effects model applied to the set of both mono- and multilinguals.

The intra-class correlation in a null model with no fixed effects (Goldstein, 2011) for the multilingual candidates (dividing the variance component of interest by the sum of variance components) indicated that 21.1% of the variation in proficiency scores was due to differences across L1s, 6.1% across L2s, 5.3% across L1–L2 combinations, and 18.0% across countries; a further 49.5% of the variation was due to individual differences. The null-level

variance components were the following: L1–L2 3.07 (95% confidence interval [CI] = 2.08, 4.17), countries 10.42 (95% CI = 8.85, 12.31), L1 12.31 (95% CI = 9.76, 15.46), L2 3.51 (95% CI = 2.41, 5.14), residual 28.75 (95% CI = 28.55, 28.96). We computed percentages of variance explained by subtracting the relevant variance component from the null-model variance component and dividing again by the null-model variance component (Kreft & Leeuw, 1998; Snijders & Bosker, 2011).

The variation across combinations may potentially be explained by interactions between L1 and L2 knowledge. Therefore, our goal was to reduce the language-related variances, while assuming that individual variation was homogeneous within each specifically observed language background. We already know that linguistic distance interacts with individual-level factors, such as age of arrival and length of residence from a previous study (Schepens et al., 2013b). The effect of distance was higher for older learners and for learners who had resided in the Netherlands for a longer period of time. Although these cross-level interaction effects increase model fit, we observed that they reduce individual-level variation rather than language-related variation.

Results

We started with testing L1 distance effects in all learners together (mono- and multilingual) in order to assess whether multilingual learners have higher L3 proficiency scores than monolingual learners. Next, we tested L2 language background effects on L3 performance in multilingual learners. We compared an additive L1 + L2 distance model and two nonadditive models. One nonadditive model was based on the lowest distance between the L1 and L2 (typological primacy), the other one was based on unique L1 × L2 combinations (the L1 × L2 interaction effect). Finally, we tested whether L1 and L2 distance effects are robust against including both monolingual and multilingual learners in the same model and examined how this inclusion can be carried out.

L1 Distance Effects in Monolingual and Multilingual Learners

In two previous studies (Schepens et al., 2013a, 2013b), we showed for both mono- and multilingual learners that lexical and morphological distances correlate with L2/L3 speaking proficiency scores. Our first goal here, before we advance to a L2 distance model of by-L2 speaking variation, was to assess whether the L1 distance measures should still be included in a model of by-L1 speaking variation for multilinguals. L1 lexical and morphological distances are different dimensions of the communicative competences of L3 learners of Dutch. Therefore, both measures might explain part of the variance, which

should become evident from likelihood ratio tests. According to Schepens et al. (2013a), an increased morphological complexity of a L2, compared with a L1, results in lower L2 learnability.

We first examined how much of the remaining variation in L3 speaking proficiency results from differences between monolinguals and multilinguals after accounting for L1 morphological and lexical distance effects. We started by testing whether there is an overall significant difference in proficiency between the monolingual learners of L2 Dutch (coded as 0) and the multilingual learners of L3 Dutch (coded as 1). This binomial variable was highly significant, $B = 8.97$, $SE = .42$, $t(47910) = 21.53$, $p < .0001$; $\chi^2(1) = 461.32$, $p < .0001$. The dummy variable shows that the L2 matters beyond the L1 and has an additional positive effect on the speaking proficiency score (almost nine points) in favor of multilinguals.

Before testing L2 lexical and morphological distances, we evaluated whether the L1 distance effects remain the same when we restrict our analysis to multilinguals only. As was the case for the whole group, we found most evidence for a joint model in which deleting both increasing morphological complexity, $\chi^2(1) = 13.21$, $p < .001$,³ and lexical distance, $\chi^2(1) = 21.99$, $p < .001$, decrease model fit significantly. This “L1 model” explains 47.7% of the by-L1 variance in speaking proficiency scores, which is more than either lexical distance (39.6%) or morphological complexity (30.8%) alone. Explained variance across countries of birth (43.7%), L2s (-4.1%), L1-L2s (2.4%), and the individual level residual variation (2.3%) remained constant when either lexical distance or morphological complexity were removed. Note that a negative value for the proportions of such predictor-specific R^2 values can arise as a side effect from subsequent updates to a model in which variance is relocated to other predictors (Nakagawa & Schielzeth, 2013; Raudenbush & Bryk, 2002; Snijders & Bosker, 2011).

The final L1 model included gender, age of arrival, length of residence, the interaction between years of education and educational quality, L1 lexical distance, and L1 morphological complexity. All fixed effects were significant at the .001 level (except years of education, although its interaction with educational quality was significant) using Satterthwaite approximations, which are used to determine the effective degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2014). The first four columns in Table 1 show the estimates and CIs for the factors in the L1 model for multilinguals. The last three columns show the “L1 + L2 model” for multilinguals, which will be discussed in the next section. Interestingly, the directions of the effects and the effect sizes in Table 1 did not change as compared to analyses that include both monolinguals

Table 1 Estimates and CIs for the random and fixed effects included in the L1 model and the L1 + L2 model fitted to the multilingual learner group

Effect	L1 Model		L1+ L2 Model	
	Estimate	95% CI	Estimate	95% CI
Random L1-L2 variance	3.01	2.11, 4.02	3.07	2.14, 4.07
Random C variance	5.87	4.80, 7.06	5.89	4.81, 7.07
Random L1 variance	6.45	4.61, 8.21	6.35	4.54, 8.12
Random L2 variance	3.65	2.54, 5.31	2.34	1.34, 3.63
Residual variance	28.10	27.93, 28.33	28.10	27.93, 28.33
Intercept	527.45	518.64, 536.09	534.26	525.1, 543.46
Gender (female = 1)	7.31	6.63, 8.00	7.31	6.64, 8.00
Age of arrival	-0.65	-0.70, -0.61	-0.66	-0.70, -0.61
Length of residence	0.53	0.46, 0.61	0.54	0.46, 0.61
Years of education	-0.09	-1.12, 1.00	-0.10	-1.16, 0.96
Education quality	0.15	0.09, 0.22	0.15	0.09, 0.21
Education years × quality	0.03	0.02, 0.04	0.03	0.02, 0.04
L1 morphological distance	-63.19	-96.23, -30.69	-62.20	-94.87, -29.73
L1 lexical distance	-39.51	-54.25, -24.54	-40.22	-54.94, -25.5
L2 morphological distance			-18.52	-51.46, 14.23
L2 lexical distance			-14.14	-23.56, -4.49

Note. Both distance measures are on continuous scales; lexical distances range from .0105 to 0.595 ($M = .322$, $SD = .175$), while morphological distances range from $-.017$ to $.327$ ($M = .050$, $SD = .057$). The 95% CIs are based on the profile likelihood.

and multilinguals. Indeed, the L1 distance effect is present in the subset of multilinguals as well as in the group that includes both mono- and multilingual learners.

We conclude that increasing morphological complexity can be used as a morphological distance measure jointly with L1 lexical distance and that they complement each other in explaining variation across L1s in multilinguals. The more morphologically complex and the more lexically distant Dutch is, compared to the L1 of the learner, the lower the L3 proficiency.

Adding L2 Effects for Multilingual Speakers

Similar to the L1 distance effect, Figure 2 shows the results of an exploratory analysis of the raw data that suggests that there seems to be a L2 distance as well. In Figure 2, three Indo-European languages are crossed as L1 and L2, and Farsi is added as the fourth Indo-European language to visualize the effect of

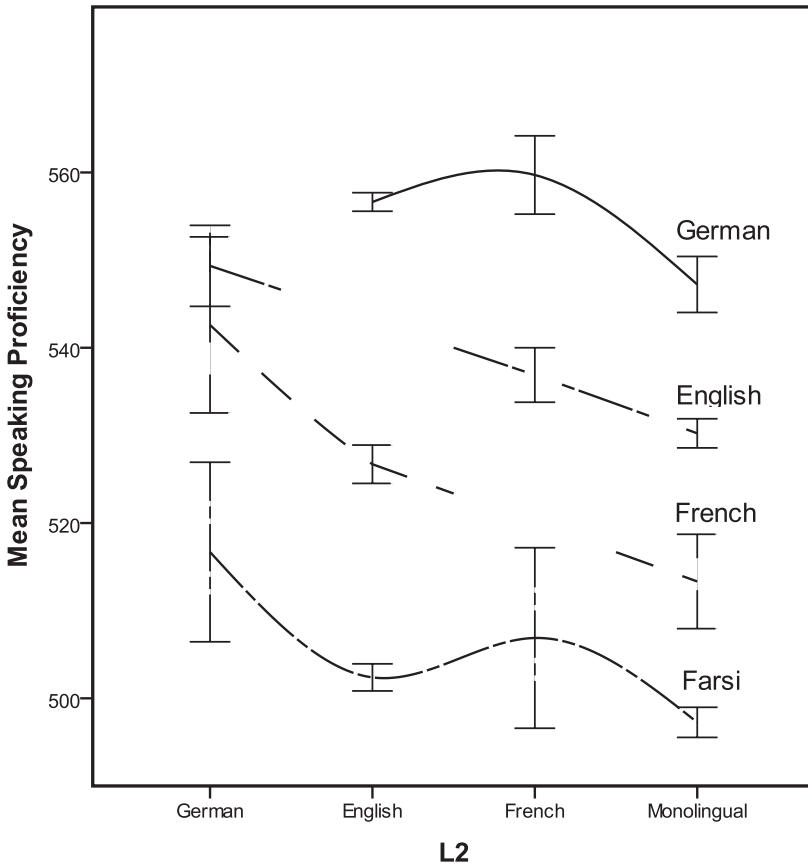


Figure 2 Mean speaking proficiency scores for 13 L1-L2 combinations with 95% CIs (all four L1s are Indo-European languages). The x-axis distinguishes between monolingual L1 and several multilingual L1-L2 combinations. The interpolations between mean speaking scores show downward trends, with monolinguals having the lowest scores.

these three languages as L2s. The figure explores differences in L3 proficiency across the L1s and L2s without controlling for gender, age, education, and exposure effects. The differences between the L1s seem to be larger, but the L2 seems to matter as well, with L2 German producing the highest outcomes. The monolinguals have the lowest scores.

The lexical and morphological distance measures are also suitable for measuring the lexical and morphological distances between L2s and Dutch. The question is whether lexical and morphological distance can explain variation

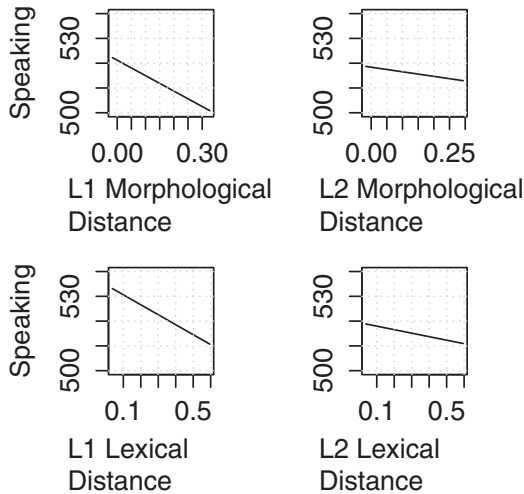


Figure 3 Partial effects for the L1 + L2 model. The L1 has a stronger influence on learning L3 Dutch, compared to the effect of the L2.

across L2s as well as across L1s. Likelihood ratio tests indicated that the updated model including both L2 lexical and morphological distances provides significant improvement of fit to the data, $\chi^2(2) = 15.32, p < .001$. Individually, lexical distance, $\chi^2(1) = 14.16, p < .001$, and morphological distance, $\chi^2(1) = 7.30, p < .001$, were both significant, although the improvement in fit for morphological distance was considerably smaller. The L1 + L2 model raises the explained-by-L2 variance from -4.1% to 32.1% . However, it was lexical distance that was largely responsible for the explained variance across L2s (32.4%), while morphological distance explained only 12.3% of the variation. Similarly, L2 morphological distance turned out to be nonsignificant in terms of the CI estimates (as shown in Table 1), although improvement in model fit was significant. Figure 3 shows the partial effects for each predictor, contingent on the remaining predictors in the model. The explained-by-L1 variance in speaking proficiency scores increased slightly (to 48.3%), while the individual variance and by-country variance remained the same. The explained variance across L1–L2 combinations was reduced from 2.4% to 0.7% , indicating that L2 distance did not increase the explained variance across L1–L2 combinations (which actually decreases slightly).⁴

Before the L2 distance effects were added to the model, there were significant correlations between by-L2 adjustments for all 35 L2s and both lexical

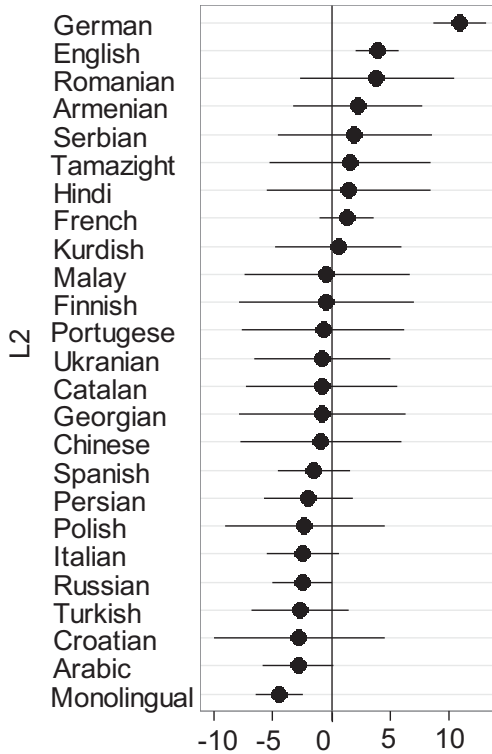


Figure 4 L2 score adjustments per L2 after L1 distances but before L2 distances are added to the model for the 25 most common L2s. The remaining random variance across L2s shows how monolinguals are hindered more than speakers of any L2, although the size of this difference depends on the L2.

($r = .58, p < .001$) and morphological ($r = .42, p < .05$) distance measures. The correlations vanished completely when distance was accounted for in the statistical model. The remaining L2-adjustments were most positive for speakers of L2 German, Hindi, and Armenian, who may experience additional benefit beyond linguistic distance, and lowest for speakers of L2 Italian, Russian, and English, who may experience impeding effects despite favorable linguistic distance (see also Figure 4).

We conclude that L2 lexical and morphological distance both play a role but that the role of L2 morphological complexity is less strong than that of lexical distance given the current distance measures. Most importantly, the L1 + L2 model predicts that there is an independent, constant L2 effect irrespective of

the L1. It means that the added value of L2 French, for instance, is constant, irrespective of whether the L1 is German or Spanish.

A Non-Additive L2 Effect: Typological Primacy

An outstanding problem in the literature is whether the status of having learned a language as a L1 or L2 determines whether L1 or L2 knowledge will be transferred to the L3. It has been proposed that knowledge from a typologically closer language is more likely to be transferred than knowledge from a more distant language, irrespective of having learned the language as a L1 or L2 (the status factor). If a L3 learner speaks a L2 that is typologically close to Dutch, the L2 may replace the primary L1 effect and the L1 effect may or may not still produce a secondary effect. We will refer to this new distinction as the “min model.” We can fit the min model by selecting the minimal distance (min) of both the L1–L3 and L2–L3 distances. The minimal distance is the distance to Dutch from the language least distant to Dutch. Its opposite, the maximal distance, is the distance to Dutch from the language most distant to Dutch.

The evidence for the min model in our data was more than one million times lower than the evidence for the L1 + L2 model (using an evidence ratio test based on AIC; Spiess, 2013). This indicates that the L1 and L2 effects were more stable across learners than minimal distance effects. The min model can be improved by adding the maximum distance to it, $\chi^2(2) = 31.62$, $p < .001$, indicating that the most distant language among the L1s and the L2s significantly influences L3 speaking proficiency. However, the evidence for the max + min model was still 100,277.2 times lower than the evidence for the L1 + L2 model. We conclude, therefore, that the data do not support a min model or a min + max model as strongly as they support a L1 + L2 model. The status of the L1 and the L2 is an important determiner of their influence, and the status or impact of the L1 is more important than the status of the L2.

L1 × L2 Interaction Effect

Although the additive L2 distances provide a significant improvement to a L1 distance model, it is possible that the learners’ L1 distance negatively interacts with their L2 distance. The unexplained variance across L1–L2 combinations in the null model was estimated at 5.3% of the total unexplained variance across learners. The L1 + L2 model only accounted for 0.7% of this percentage. Can an interaction effect between L1 and L2 distances explain a significant part of the random variance across L1–L2 combinations? We may expect that a highly

similar L2 will provide a relatively higher benefit to speakers of a particular L1 when that L1 is relatively distant from Dutch; the same L2 may provide less benefit if a L1 is close to Dutch already. A positive interaction effect (L1 multiplied by L2) may be able to capture this. The closer the L1 and the L2 are to Dutch, the more they will support each other; the more distant the L1 and L2 are to Dutch, the more they will diminish each other's added value. Cases in which either the L1 or the L2 is similar (and the other distant) will be relatively more beneficial than in the additive cases. For example, suppose that English and German are both five times more similar to Dutch than French and Spanish are (e.g., distance of .1 for English and German versus .5 for French and Spanish). Then, in the interactive case, L1 Spanish–L2 English is five times less similar to Dutch than L1 German–L2 English (.1 × .5 vs. .1 × .1) and five times more similar than L1 Spanish–L2 French (.1 × .5 vs. .5 × .5). In the additive case, L1 Spanish–L2 English is three times less similar to Dutch than L1 German–L2 English (.5 + .1 vs. .1 + .1), but only 1.66 times more similar than L1 Spanish–L2 French (.1 + .5 vs. .5 + .5).

To test for a L1 × L2 interaction effect, we added multiplicative terms between lexical and morphological L1 and L2 distance to the L1 + L2 model. Likelihood ratio tests indicated that the “L1 × L2 model” did not fit the data better than the L1 + L2 model, $\chi^2(2) = .87, p = .65$. Neither lexical distance, $\chi^2(1) = .40, p = .53$, nor morphological complexity, $\chi^2(1) = .34, p = .56$, were significant. The L1 × L2 model did not explain additional variance across L1–L2 combinations, and the L1 × L2 multiplicative effect did not fall within the 95% CI, nor was it significant according to Satterthwaite approximations. Therefore, we conclude that L1-by-L2 random interaction cannot be explained by a L1-by-L2 fixed multiplicative effect of either lexical or morphological distance.

It may be the case that uncommon L1–L2 combinations produce unstable effects and that they obstruct our chance of observing systematic multiplicative effects. To assess whether the interaction of L1–L2 combinations was also not significant in the most common L1–L2 combinations, we removed the L1–L2 combinations that appeared less than 15 times in the database. The resulting models showed that multiplicative L1 × L2 distance effects remained non-significant, regardless of the removal of uncommon L1–L2 combinations (see Table 2). Thus, multiplications between different distances cannot account for the patterns observed across combinations beyond individual additive distances, and the remaining variation across L1–L2 combinations is more complex than a simple multiplicative effect.

Table 2 Comparisons of additive L1 and L2 distance effect models to models including L1 \times L2 multiplicative effects

L1-L2 < 15	Monolinguals	$\chi^2(2)$	<i>p</i>	Evidence ratio
Included	Excluded	0.87	.65	<.001
Included	Included	1.60	.45	<.01
Excluded	Excluded	2.70	.26	<.0001
Excluded	Included	1.81	.40	<.001
Excluded	Included + dummy	1.85	.40	<.001

Note. Multiplicative effects for both lexical and morphological distance are included. The lower the evidence ratio, the less evidence there is for interactive effects.

Effects of L2 Distance or Multilingualism?

It now seems evident that an independent L2 distance effect operates alongside a L1 distance effect. Part of this explanation, however, depends on the assumption that the addition of a L2 to the language inventory of the speaker does not affect L1 behavior, as the two effects are independent. If that is true, monolingual speakers would have the distance advantage of their L1, but no profit of the L2 as there is no L2. We reexamined the L1 effect in the complete group of monolingual and multilingual speakers. Monolinguals were given the value “monolingual” for their L2. As L2 distance is not defined for monolingual speakers, we set L2 distance to the highest observed distance across all L2s (Albanian for lexical distance; Igbo for morphological distance). Setting the L2 distance to the L1 value would wrongly model that monolinguals benefit twice from their L1. We also added the dummy variable for being multilingual in order to test whether the general effect of being multilingual is still significant after L2 distance is accounted for.

We found that a dummy variable for multilingualism was nonsignificant when the L2 distance measure already accounted for by-L2 variation. However, the general effect of the multilingualism dummy variable was significant when random effects for the L2 and L1–L2 combinations were removed. This suggests that there is a gradual difference between monolinguals and multilinguals, depending on the type of L2. In a model including a random effect for the L2 instead of the dummy variable (and no L2 distance measures), the score adjustment for the monolinguals was the most negative one observed (see Figure 4), reflecting a (gradual) difference between monolinguals and multilinguals.

We also found that including monolinguals did not change the estimations of the L1 and L2 distance effects. Explained variances remained robust against the inclusion of monolinguals in the L1 + L2 model. Also, after adding monolinguals and setting their L2 to a maximum distance from Dutch, L1 \times L2

distances remained nonsignificant (see Table 2). We also compared the effects of including and excluding monolinguals and uncommon L1–L2 combinations on the distribution of variance across the random effects. This showed only small changes: The inclusion of monolinguals gives an increase of variation across L1–L2 combinations while excluding uncommon L1–L2 combinations results in a slight increase of explained variance across L1s, but not across L2s. We conclude that knowledge of a L2 can be helpful in general but also that its effect size mainly depends on the specific L2 to L3 distance.

Discussion

We investigated effects of L1 and L2 distance on L3 Dutch learnability in STEX state examination data. We started by hypothesizing that the lower the L2 distance, the higher the L3 learnability and that this effect is weaker than the effect of L1 distance. Accordingly, robust additive L1 and L2 distance effects were found in upper intermediate learners of L3 Dutch, where the L1 effect was stronger than the L2 effect. Cross-classified random effect models were used to decompose variance in speaking proficiency test scores into by-country, by-L1, by-L2, and L1-by-L2 variation. We first implemented a number of third factors into our model (country of origin, education, gender, age, and exposure) in order to see whether enough variation remains for a role of the L2. L1 and L2 distances were measured with lexical and morphological distance measures. There were five main findings in our study.

1. Both lexical distance (39.6% across L1s and 32.4% across L2s) and morphological distance (30.8% across L1s and 12.3% across L2s) explain a significant proportion of the variance in the group including multilinguals only. Adding monolinguals to this model does not change these effects significantly.
2. There is more variation in L3 proficiency across L1s than across L2s (21.1% vs. 6.1%). The L1 distance effect is stronger than the L2 distance effect (distances explain 47.7% of the variance across L1s vs. 32.4% across L2s).
3. An additive L1 + L2 distance model fits the data better than more complex L1 × L2 multiplicative distances.
4. Language status (whether a language is a L1 or L2) fits the data better than an alternative model that gives primacy to a language based on the smallest distance.
5. Being multilingual is generally better for learning a new language than being monolingual, provided the L1 is the same. This facilitative effect is generally smaller for larger distances, and there are L2s that are not always more beneficial than being monolingual.

Table 3 Predictions from previous research compared with current findings

Study	Model	Prediction	Finding
Flynn et al. (2004)	Increasingly cumulative	Bilingual > monolingual	L2 distance > monolingual
Bohnacker (2006)	L2 blocks L1 transfer	L2 > L1	L1 > L2
Bardel & Falk (2007)	L2 status factor	L2 status > L1 distance	L1 status > min distance (L1, L2)
Rothman (2011)	Typological primacy	min distance (L1, L2)	L1 + L2

The observed additive L1 and L2 distance effects provide evidence for a theory of L3 learnability in which knowledge of previously acquired languages has to be accounted for. We discuss how these findings relate to existing theories of L3 learning, validity issues in measurements of L3 performance, and issues in L1 by L2 mixing.

Relevance to Existing Theories of L3 Learnability

We found that the L1 is more important than the L2. This differs from a model that predicts no special importance to the L1 (Rothman, 2010, 2011) and a model that predicts a L2 blocking effect to the L1 (Bohnacker, 2006). Both models are summarized in Table 3. However, our results meet the predictions of the cumulative enhancement model (Flynn et al., 2004), which states that the effect of the L2 (a) is not absorbed by the L1, (b) is either neutral or positive, and (c) is more beneficial for learning a L3 than having no L2 at all. With respect to (c), we cannot reliably conclude that all the 35 L2s, including the distant ones, provided added benefit above and beyond having no L2 at all. Our outcomes add two additional conclusions, namely, that L1 influence is stronger than L2 influence and that the degree of L1 and L2 influence is related to the respective L1 and L2 distances. The degree of L1 and L2 influence is not affected by the distance of the other language involved, and the L2 distance effect does not prohibit, impede, or enhance the L1 distance effect. In fact, L2 distance can benefit the learner, and this effect adds to and does not substitute or change the L1 distance effect. What are the consequences for the other two theories mentioned in Table 3? Our findings align with a variation of the L2 status factor model (Bardel & Falk, 2007). We find that the status of a language has to be taken into account when accounting for the typological similarity in

L3 learning. However, typological factors do not change the importance of the L1 (Rothman, 2011), even if the L2, as compared to the L1, is closer to the L3.

Evaluation of L3 Performance Measures

We expect that additive L1 and L2 distances not only predict STEX performance, but that this finding bears on L3 learnability in general as well. Throughout this article, we used the term L3 learnability to refer to the relative L1 and L2 influence on the proficiency of L3 speakers of Dutch. These speakers had a variety of L1s and knowledge of one or more additional languages, representing a wide sample of languages across the world. Unlike English, French, Spanish, or German, Dutch is not a language often taught in secondary schools across the globe. This avoids a comparability problem because of L2/L3 exposure in country of birth. Although such problems may admittedly complicate the study of world languages like English or Spanish as L3s, we believe that the present study provides directions to conduct such research.

Linguists tend to be critical toward the accuracy and reliability of test scores (Munro, 2008). However, test scores are almost certainly more accurate and reliable than self-reported proficiency levels. If there is remaining noise in the STEX proficiency measures, the estimations of the distance effects reported here represent lower bound conservative estimates of the actual distance effects. However, care should still be taken in interpreting the importance of the L2 compared to the L1, as the L2s reflected subjective judgments of participants' best additional languages. We do not expect, however, that many participants would report L2s in which they are unable to express themselves. The self-assessment of knowing how to express oneself in a L2 is made in the formal setting of a B2 language exam, and self-reported proficiency is known to correlate weakly, but significantly, with more objective measures of proficiency. We cannot exclude the possibility that candidates reported a L2 from which they cannot transfer anything due to low proficiency in that language. Similarly, participants may have had competing candidate languages that can impede or facilitate learning. Theoretically, such noise, if present, only leads to an underestimation, rather than an overestimation, of L2 distance effects. More research is necessary to study whether higher L2 proficiency leads to stronger L2 distance effects.

L1 × L2 Mixing

Additive L1 and L2 distance effects appear to provide a better explanation for how L1 and L2 effects work independently. The model has the benefit of being straightforward and transparent. The data do not provide support for

multiplicative distance effects, but we need to be careful in reaching conclusions. This is because variance remains in our data in the mix of different L1 and L2 types that is not yet explained by additive L1 and L2 distance effects. It is unclear whether this variance in specific L1 \times L2 pairs is due to idiosyncrasies of the current data set, for instance, due to other variables not included in our database (such as social factors) or due to particular supportive or impeding L1 \times L2 pairs. For example, after looking specifically at individual L1–L2 combinations, we observed that the L1 + L2 model does not explain how L1 Polish interacts with L2 Italian, as observed performance was lower than the model's prediction. We found no evidence that the combination of the L1 and the L2 can explain such idiosyncrasies. This does not exclude the possibility that social factors, such as social class or the amount of working hours, may be important.

Our findings suggest that L1 influence does not change by adding a L2, implying that speakers efficiently combine L1 and L2 influences in L3 learning. It means that the L1 effect in monolinguals remains comparable to the L1 effect in bilinguals and that L1 knowledge remains intact after a L2 has been added. Bilinguals make use of an additive L2 distance effect, which leads to performance increases in learnability.

Linguistic Distance

Linguistic distance nicely captures the regularity in L1- and L2-dependent learning difficulty of a L3 (i.e., L3 learnability). The variation in L3 learnability cannot be explained by only one distance effect, because evidence for two additive effects of linguistic distance was found. We used lexical distance as a measure of similarity between L3 Dutch and Indo-European languages with a maximum distance to non-Indo-European languages, and morphological distance as a measure of similarity between L3 Dutch and both Indo-European and non-Indo-European languages. In addition, as learning difficulty is not necessarily symmetrical, we used lexical distance as a symmetric distance measure and morphological distance as an asymmetric distance measure. In combination, these two measures account for symmetric and asymmetric distances within Indo-European languages, and asymmetric distances only within non-Indo-European languages. In addition, because L1 distance, age, and speaking proficiency interacted in our data, future work could examine the non-linearity involved in the role of L2 distances. Furthermore, because the phonology of an additional language has a persistent influence on L2 learning difficulty as well, we also intend to investigate effects of phonological distance on L2 and L3 learnability in the near future.

Conclusion

This study investigated whether linguistic distances of previously acquired languages predict the learnability of an additional language (L3 Dutch in this case). The study shows that lexical distances within language families and morphological distances between languages from different language families can be used. The results demonstrate the importance and robustness of distance effects from L2s in learning a new L3. That is precisely what was predicted by our main hypothesis. We can conclude that the closer the L2 is to the L3, the higher is the learnability of the L3. Learning a L3 becomes more difficult when the L1 or the L2 are lexically distant and morphologically less complex. Does that imply that a L2 can have a negative or impeding effect on learning an additional language? Our second hypothesis was that the L2–L3 distance effect is facilitative as language distance decreases, because multilingual learners obtain higher proficiency scores in L3 Dutch than monolinguals. The results support this hypothesis, showing that the L2 distance effects are comparable to the L1 distance effects. No significant effects of being multilingual remain after the L2 distance effects are accounted for, which indicates that general cognitive effects of multilingualism on additional language learning can be deconstructed into independent L1 and L2 distance effects. The monolinguals have only L1 distance effects and do not profit nor are they hampered by additional L2 effects. Finally, the outcomes corroborate our third hypothesis that states that the L2 distance effect is weaker than the L1 distance effect.

Previous studies have not provided conclusive evidence for independent additive L2 distance effects, because of small-scale data or because of the lack of distance measures. A central step in our analyses was the definition of distances, both at the lexical and morphological levels. Evidence from large-scale comparative data across language backgrounds seems useful for the study of distance effects on L3 learnability. This approach can further benefit from studies of other languages besides Dutch. The frequency with which multilingualism occurs in society creates an opportunity for large-scale analyses of the persistent diversity in the learning of additional languages by adults. The present findings provide new empirical evidence for theories that predict L1 and L2 distance effects that cannot be neglected in the learning of additional languages by adults. Our study does not predict concrete transfer phenomena; instead, it shows how the degree of effort necessary to learn a L2 or L3 varies depending on L1 and L2 linguistic distances. This implies that absence of specific forms of transfer is not counterevidence against global L1 and L2 linguistic distance effects. In sum, we hope that the present findings contribute to

clarifying the challenges involved in additional language learning by monolingual and multilingual speakers.

Final revised version accepted 2 April 2015

Notes

- 1 http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf p.24
- 2 The degree of fractionalization between individuals across groups (languages) in an area (countries) affects political and economic developments (Desmet, Ortuño-Ortín, & Wacziarg, 2012). We tested for such effects by using Greenberg's measure of linguistic diversity (Greenberg, 1956) in the country of birth (number of languages per area).
- 3 As in our previous study (Schepens et al., 2013a), in which we compared three measures of morphological complexity, we found that L1 morphological similarity can be removed from a model that includes morphological similarity, increasing morphological complexity, and lexical distance without reducing model fit significantly, $\chi^2(1) = .31, p = .58$. Removing decreasing morphological complexity did not change the model fit significantly either, $\chi^2(1) = .02, p = .89$.
- 4 A number of different general R^2 estimates (in contrast to variable-specific measures) of goodness of fit are available for mixed-effects models (Nakagawa & Schielzeth, 2013); we report two of them here. The overall variance explained by all of the fixed effects is 58.1% (general R^2 marginalized over the random effects). The variance explained by the fixed effects, given that the random effects are known, is 91.7% (simple R^2 , conditional on the random effects). The AIC as estimated from an unrestricted maximum log likelihood fitted L1 + L2 model is 366,963.0 versus 366,974.4 for L1 only. This is a clear improvement in model fit (Baayen et al., 2008).

References

- Adsera, A., & Pytlikova, M. (2012). *The role of language in shaping international migration* (SSRN Scholarly Paper No. ID 2003666). Rochester, NY: Social Science Research Network. Retrieved from <http://papers.ssrn.com/abstract=2003666>
- Ahukanna, J. G. W., Lund, N. J., & Gentile, J. R. (1981). Inter- and intra-lingual interference effects in learning a third language. *The Modern Language Journal*, *65*, 281–287. doi:10.1111/j.1540-4781.1981.tb00985.x
- Ard, J., & Homburg, T. (1983). Verification of language transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 157–176). Rowley, MA: Newbury House.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005
- Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*, 459–484. doi:10.1177/0267658307080557

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. Retrieved from <http://arxiv.org/abs/1406.5823>
- Bohnacker, U. (2006). When swedes begin to learn german: From V2 to V2. *Second Language Research*, 22, 443–486. doi:10.1191/0267658306sr2750a
- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., et al. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, 337, 957–960. doi:10.1126/science.1219669
- Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 8–20). Clevedon, UK: Multilingual Matters.
- Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, 7, 71–87.
- Cenoz, J., Hufeisen, B., & Jessner, U. (Eds.). (2001). *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives*. Clevedon, UK: Multilingual Matters.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Dahl, Ö., & Velupillai, V. (2013). The past tense. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/66>
- De Angelis, G. (2005). Interlanguage transfer of function words. *Language Learning*, 55, 379–414. doi:10.1111/j.0023-8333.2005.00310.x
- Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics*, 97, 322–338. doi:10.1016/j.jdeveco.2011.02.003
- Desmet, K., Weber, S., & Ortuño-Ortín, I. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7, 1291–1318. doi:10.1162/JEEA.2009.7.6.1291
- Dryer, M. S., & Haspelmath, M. (Eds.). (2011). *The world atlas of language structures online*. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info>
- Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473, 79–82. doi:10.1038/nature09923
- Dunn, M., Terrill, A., Reesink, G., Foley, R. A., & Levinson, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science*, 309, 2072–2075. doi:10.1126/science.1114615
- Dyen, I., Kruskal, J. B., & Black, P. (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82, iii–132.

- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305–352. doi:10.1017/S027226310505014X
- Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, 27, 59–82. doi:10.1177/0267658310386647
- Falk, Y., Lindqvist, C., & Bardel, C. (2013). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition, FirstView*, 1–9. doi:10.1017/S1366728913000552
- Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8, 195–222. doi:10.1023/A:1024419522867
- Finnie, R., & Meng, R. (2005). Literacy and labour market outcomes: Self-assessment versus test score measures. *Applied Economics*, 37, 1935–1951.
- Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, 1, 3–16. doi:10.1080/14790710408668175
- Goldstein, H. (2011). *Multilevel statistical models*. Chichester, UK: Wiley.
- Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, 435–439. doi:10.1038/nature02029
- Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language*, 32, 109–115.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, 14, 31–38. doi:10.1111/1467-9280.01415
- Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 21–41). Bristol, UK: Multilingual Matters.
- Holman, E. W., Brown, C. H., Wichmann, S., Müller, A., Velupillai, V., Hammarström, H., et al. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52, 841–875.
- Hulstijn, J. H., Schoonen, R., de Jong, N. H., Steinel, M. P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing*, 29, 203–221. doi:10.1177/0265532211419826
- Inon, T., & Wexler, K. (2002). Why is “is” easier than “-s”? Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research*, 18, 95–136. doi:10.1191/0267658302sr195oa

- Isphording, I. E., & Otten, S. (2011). *Linguistic distance and the language fluency of immigrants* (No. 274). Ruhr Economic Papers. Retrieved from <http://www.econstor.eu/handle/10419/61444>
- Isphording, I. E., & Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior & Organization*, *105*, 30–50. doi:10.1016/j.jebo.2014.03.027
- Jaensch, C. (2013). Third language acquisition: Where are we now? *Linguistic Approaches to Bilingualism*, *3*, 73–93. doi:10.1075/lab.3.1.04jae
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.
- Jessner, U. (2012). Dynamics of multilingualism. In Carol A. Chappelle et al. (Eds.), *The encyclopedia of applied linguistics* (pp. 1798–1805). Malden, MA: Blackwell.
- Jessner, U. (2014). On multilingual awareness or why the multilingual learner is a specific language learner. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism* (pp. 175–184). Berlin, Germany: Springer.
- Kellerman, E., & Sharwood Smith, M. (1986). *Crosslinguistic influence in second language acquisition*. New York: Pearson.
- Kreft, I. G. G., & Leeuw, J. D. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: SAGE.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). (Version 2.0-11). Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325–343. doi:10.3758/s13428-011-0146-0
- Lewis, P., Simons, G., & Fennig, C. (Eds.). (2013). *Ethnologue: Languages of the world* (17th ed.). Dallas, TX: SIL International. Retrieved from <http://www.ethnologue.com>
- Lindqvist, C. (2010). Inter- and intralingual lexical influences in advanced learners' French L3 oral production. *International Review of Applied Linguistics in Language Teaching*, *48*, 131–157. doi:10.1515/iral.2010.007
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1), e8559. doi:10.1371/journal.pone.0008559
- McArthur, E. K., & Siegel, P. M. (1983). Developments in the measurement of English language proficiency. In *Proceedings of the American Statistical Association* (pp. 373–378).
- Moran, S., & Blasi, D. (2014). Cross-linguistic comparison of complexity measures in phonological systems. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity* (pp. 217–240). Oxford, UK: Oxford University Press.

- Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–219). Amsterdam: John Benjamins.
- Murphy, S. (2005). Second language transfer during third language acquisition. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 3. Retrieved from <http://journal.tc-library.org/index.php/tesol/article/view/23>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142. doi:10.1111/j.2041-210x.2012.00261.x
- Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge, UK: Cambridge University Press.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2013). *nlme: Linear and nonlinear mixed effects models*. (version R package version 3.1-122), 2015. <http://CRAN.R-project.org/package=nlme>.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Retrieved from <http://www.r-project.org>
- Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon, UK: Multilingual Matters.
- Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *International Review of Applied Linguistics in Language Teaching*, 48, 245–273.
- Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, 27, 107–127. doi:10.1177/0267658310386439
- Ruhlen, M. (1991). *A guide to the world's languages: Classification*. Stanford, CA: Stanford University Press.
- Schepens, J., Van der Slik, F., & Van Hout, R. (in press). The L2 impact on acquiring Dutch as a L3: The L2 distance effect. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed effects regression models in linguistics*. Berlin, Germany: Springer.
- Schepens, J., Van der Slik, F., & Van Hout, R. (2013a). Learning complex features: A morphological account of L2 learnability. *Language Dynamics and Change*, 3, 218–244. doi:10.1163/22105832-13030203
- Schepens, J., Van der Slik, F., & Van Hout, R. (2013b). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In

- L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 199–230). Berlin, Germany: De Gruyter.
- Siegel, P., Martin, E., & Bruno, R. (2001, February 12). *Language use and linguistic isolation: Historical data and methodological issues*. Washington, DC: U.S. Census Bureau. Retrieved from <https://www.census.gov/content/dam/Census/library/working-papers/2001/demo/li-final.pdf>
- Singleton, D. (1987). Mother and other tongue influence on learner French. *Studies in Second Language Acquisition*, 9, 327–345. doi:10.1017/S0272263100006719
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: SAGE.
- Spieß, A. N. (2013). qpcR: Modelling and analysis of real-time PCR data (Version 1.3-7.1). Retrieved from <http://cran.r-project.org/web/packages/qpcR/index.html>
- Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society*, 452–463.
- Trask, R. L. (2000). *Time depth in historical linguistics*. Cambridge, UK: The McDonald Institute for Archaeological Research.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford, UK: Oxford University Press.
- United Nations Educational, Scientific, and Cultural Organization. (2011). The World Bank, Institute for Statistics, School enrollment, secondary (% gross). Retrieved January 3, 2013, from <http://data.worldbank.org/indicator/SE.SEC.ENRR>
- Van der Slik, F. (2010). Acquisition of Dutch as a second language. *Studies in Second Language Acquisition*, 32, 401–432. doi:10.1017/S0272263110000021
- Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics*, 19, 295–333.

Supporting Information

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

Appendix S1. Aggregated Data.