Jan Vanhove* and Raphael Berthele

Interactions between formal distance and participant-related variables in receptive multilingualism

Abstract: Recognizing cognates in a related but unknown language (Lx) is of key importance in receptive multilingualism. Many studies have consequently investigated the impact of both item-related characteristics (most notably the cognates' formal distance to their L1/L2 counterparts) and participant-related variables (e. g., the make-up of the participants' linguistic repertoires) on Lx cognate recognition. However, little is known about how these two factors interact with one another. Using data from a lifespan study on Lx (Swedish) cognate recognition in German-speaking participants, we investigate how the effect of the Lx cognates' formal distance to their L1/L2 counterparts varies as a function of the participants' richness of linguistic experience and their ability to deal with abstract patterns flexibly ('fluid intelligence'). We do so for both written and spoken stimuli. The results underscore that the relationship between formal distance and recognition in receptive multilingualism, and cross-linguistic influence more generally, may vary systematically as a function of participant-related variables.

Keywords: receptive multilingualism, Levenshtein distance, perceived similarities, cognates

1 Introduction

When languages are closely related, speakers of one language can often make use of cross-linguistic similarities to partially understand written or spoken texts in the other language (Lx) without having properly learned or acquired it. This is known as *receptive multilingualism*, examples of which can be found to a greater or lesser extent in Scandinavia, between German and Dutch, and in the Romance language family. A particularly rich and crucially important source of cross-linguistic similarities in receptive multilingualism is provided by

*Corresponding author: Jan Vanhove, Department of Multilingualism, University of Fribourg, Rue de Rome 1, CH-1700 Fribourg, Switzerland, E-mail: jan.vanhove@unifr.ch

Raphael Berthele, Department of Multilingualism, University of Fribourg, Rue de Rome 1, CH-1700 Fribourg, Switzerland, E-mail: raphael.berthele@unifr.ch

cognates, i. e., genealogically related words in different language varieties (e. g., Möller and Zeevaert 2010; Heuven 2008). However, the mere presence of a cognate relationship does not automatically mean that readers or listeners can make use of them. To borrow the central metaphor of Scott Jarvis's (2014) plenary address, some steps in learning and using additional languages are easier to climb than others and some people are better at climbing steps than others. Similarly, some cognates are easier to recognize than others, and some readers or listeners are better able to recognize cognates than others. Consequently, much research on receptive multilingualism takes a deliberately reductionist approach by focusing on factors that may affect readers' or listeners' success in guessing the meaning of individual Lx cognates.

One of the main factors that is often considered in such studies is the degree of formal overlap between the Lx words and their cognates in a known language, usually the L1. This degree of formal overlap is typically measured by means of the Levenshtein algorithm, which computes the smallest number of insertions, deletions, and substitutions necessary to transform one string into another. By way of example, Figure 1 shows how the Swedish orthographic string *avskaffa* 'to abolish' can be transformed into German *abschaffen* using two insertions ('I') and three substitutions ('S') for a total operation cost of five. For spoken stimuli, phonetic transcriptions can be used as input strings. The operation cost is then normalized for the length of the alignment (in Figure 1: ten slots), yielding the Levenshtein distance (in Figure 1: 5/10 = 0.5). Lx stimuli with high Levenshtein distances to their L1 counterparts have been shown to be more difficult to understand in receptive multilingualism, both at the text level (Beijering et al. 2008; Gooskens 2007) and at the individual word level (e. g., Berthele and Lambelet 2009; Doetjes and Gooskens 2009; Gooskens et al. 2011; Kürschner et al. 2008; Bezooijen and Gooskens 2005; Vanhove et al. 2015b; but see Berthele 2011).[1]

However, applied linguists have long recognized that the existence of formal cross-linguistic similarities does not, in and by itself, guarantee that foreign-language learners can make use of them; what is important is whether and how these cross-linguistic similarities are perceived (e. g., Kellerman 1977, 1983). This leaves open the possibility that the precise relationship between formal distance

---

1 Levenshtein distances can further be refined by weighting substitutions by phonologically or phonetically similar phones (e. g., a fortis consonant such as [t] by its lenis counterpart [d]) less than substitutions by more different phones (e. g., [t] by [w]), as was done by Beijering et al. (2008). To our knowledge, however, no study reports that doing so actually improves the Levenshtein distance's power to predict the comprehensibility of spoken or written words (see also Möller 2012).

1	2	3	4	5	6	7	8	9	10	
a	v	s	k		a	f	f	a		
a	b	s	c	h	a	f	f	e	n	
	S		S	I				S	I	= 5

Figure 1: Levenshtein alignment of Swedish *avskaffa* and German *abschaffen*.

and perceived similarity varies from learner to learner. By the same token, the effect of formal distance on Lx comprehension may well differ from one participant to the next and may do so as a function of certain participant-related factors. To our knowledge, however, the effect of formal distance (or its complement, formal similarity) on Lx comprehension has always been studied averaged over all participants in a given sample. In this contribution, we therefore consider the possibility that the effect of formal distance on the comprehension of individual Lx words varies systematically as a function of (a) the breadth of the participants' linguistic repertoire and experience and (b) their ability to deal with abstract (non-linguistic) patterns flexibly. To stick to Jarvis's (2014) metaphor, what interests us here is whether some kinds of steps are comparatively easier to take for some multilinguals than for others.

With respect to (a), findings by Berthele (2008, 2011) and Vanhove and Berthele (2015a) suggest that participants with larger linguistic repertoires (including more as well as more varied L1 experience) are better at receptive multilingualism and Lx word comprehension. Their finding that inter-individual differences, even in L1 vocabulary knowledge, predict receptive multilingualism success ties in with an observation that Teleman (1981) made in a Scandinavian setting: infrequent L1 lexemes (e. g., Swedish *begynna* instead of the more common *börja* 'to start') sometimes correspond to unmarked Lx forms (Norwegian *begynne*). Knowledge of such rare L1 forms can therefore be advantageous in receptive multilingualism. Berthele (2008), however, submitted an additional explanation. Participants with richer linguistic repertoires may show a greater degree of perceptual tolerance in receptive multilingualism, i. e., a greater degree of flexibility in dealing with linguistic input that deviates from their own L1 (L2, …, Ln) norm. This suggestion ties in with findings about 'wider grammars' in bilinguals (Zobl 1992): compared to monolinguals, bilinguals are often more tolerant of discrepant syntax in reception, presumably because their search spaces are wider when analyzing syntactic structures. Similarly, wider

search spaces may be advantageous when coping with phonological or ortho-graphic interlingual discrepancies as well: multilinguals who master repertoires with a great number of interlingual correspondences can be expected to be better at finding potential interlingual connections in an unknown target lan-guage, too, and thus be less strongly affected by the sheer formal distance between Lx materials (morphemes, words, phrases, texts) and their counterparts in known languages. Greater flexibility with respect to dealing with Lx input that deviates from the corresponding form in the L1 (L2, ..., Ln) should be reflected in a weaker link between the formal distance between the Lx and L1 material and Lx comprehension.

With respect to (b), Vanhove and Berthele (2015a) found that participants who can deal more flexibly with abstract, non-linguistic patterns ('fluid intelli-gence') are better at guessing the meaning of spoken Lx cognates. Here, it can be expected that participants who are better at dealing flexibly with abstract patterns can apply those skills to receptive multilingualism contexts so that formal discrepancies between Lx stimuli and their L1 (L2, ..., Ln) cognates would not affect them as much.

In sum, the expectation is not only that higher levels of linguistic experience and fluid intelligence are associated with an *overall* better performance in Lx cognate guessing (see Vanhove et al. 2015a), but also that participants with higher levels of linguistic experience and fluid intelligence depend to a lesser degree on formal matches when guessing the meaning of Lx cognates (i. e., the slope linking formal distance and cognate guessing success should be less steep in these participants). It is the latter prediction that this paper addresses.

2 Method

The data analyzed in this paper stem from a cross-sectional lifespan study on receptive multilingualism (Vanhove 2014; Vanhove et al. 2015a) in which 159 German-speaking Swiss participants aged ten to 86 were asked to translate 50 written and 50 spoken Swedish words into German. Ninety of the words had translation-equivalent cognates in German and/or the foreign languages com-mon to most participants, viz. French and English, and their meaning could therefore in principle be inferred even without a context. All participants were recruited to provide data for three parallel projects (see Berthele and Kaiser 2014) and took part in a battery of cognitive and linguistic tasks that lasted up to three hours. For want of space, this Method section only discusses the tasks that are directly relevant to this contribution. For additional details on the exact

procedures as well as a list of the stimuli used and their cognates, we refer to Vanhove (2014) and Vanhove and Berthele (2015a). Similarly, the cross-sectional aspects of the data, including the independent effect of age as well as its correlations and interactions with the variables considered in this paper, are amply discussed in the publications cited.

2.1 Participants

We recruited 167 German-speaking Swiss participants (76 men) who were roughly uniformly distributed between the ages of 10 and 86 years. Four participants did not complete the cognate translation task (see below) due to computer malfunctions. Four further participants had to be excluded due to experimenter error, leaving a sample of 159 participants. With the exception of some of the youngest participants, all of them had at least some knowledge of French and English in addition to Standard German and their Swiss German dialects. Participants with self-reported knowledge of Germanic languages other than Standard German, Swiss German, and English were screened out before the data collection.

2.2 Tasks

2.2.1 Cognate translation task

This computer-run task consisted of two blocks featuring 50 different isolated Swedish words each. One block consisted of written words presented on the computer screen; the other block consisted of spoken words presented through headphones. The order of the blocks as well as the order of the stimuli within each block was randomized for each participant. 2 × 45 words had translation-equivalent cognates in German, English and/or French and could thus in principle be understood by the participants ('target words'). 2 × 5 words did not have any German, English or French translation-equivalent cognates and were included in order to verify that the participants did not have any prior lexical knowledge of Swedish. Participants were asked to try to translate the stimuli into German but were given the option to pass in case no translation came to mind. For full technical details as well as a list of the stimuli with their German, English and French cognates, see Vanhove (2014) and Vanhove and Berthele (2015a).

2.2.1.1 Scoring

All translations were checked manually and coded for their correctness (binary variable). German capitalization errors were disregarded. Other misspellings were not penalized either as long as they did not give rise to another existing word. A detailed scoring protocol is provided in Vanhove (2014: Section 4.2.6) and in Vanhove and Berthele (2015a: Appendix).

2.2.1.2 Computation of formal distances

The formal distances between the Swedish stimuli and their cognates were computed by means of the Levenshtein algorithm (see Introduction). The Levenshtein computations were based on orthographic strings for written items and on phonetic strings for spoken items. Analyses by Vanhove (2014; see also Vanhove et al. 2015b) on these data showed that, as a group, participants were sensitive to formal distances with respect to both German and English in the written modality but only with respect to German in the spoken modality. Neither in the written nor in the spoken modality did formal distance with respect to French emerge as a statistically relevant factor. The Levenshtein variable used for the written stimuli was therefore a combination of the Levenshtein distances with respect to German and those with respect to English: we used whichever of the two was the smaller one for each stimulus ('Germanic Levenshtein distance'). The Levenshtein variable used for the spoken stimuli was the Levenshtein distance with respect to German. When a stimulus did not have a translation-equivalent cognate in a relevant source language, the Levenshtein distance was set to 1 (i. e., the maximum).

Figure 2 shows the 2 × 45 Swedish target words and their Levenshtein values. Most of the items are situated in the middle of the Levenshtein continuum. This is the result of a deliberate choice on our part to include a relatively large number of items of presumably intermediate difficulty in order to avoid ceiling and floor effects as well as frustration in our participants. However, a more uniform distribution of items across the Levenshtein continuum might be preferable to avoid a scarcity of data points for some combinations of predictor variables.

2.2.2 Advanced German vocabulary test (Schmidt and Metzler 1992)

This test is commonly used to measure verbal ('crystallized') intelligence levels in natively German-speaking participants. It serves as a general proxy of our participants' L1 linguistic experience (see, e. g., Kuperman and Van Dyke 2013,
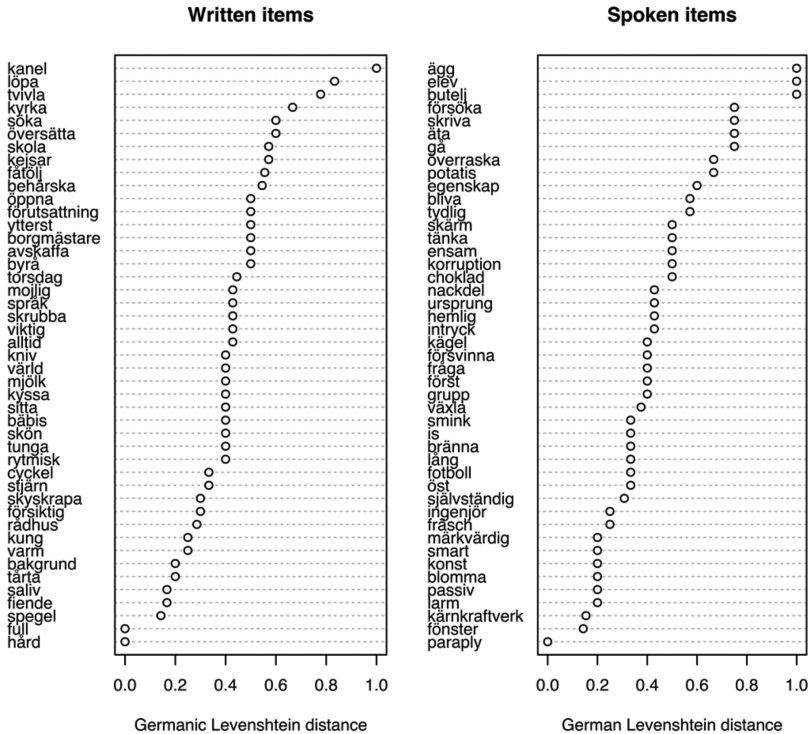
Figure 2: Dot plots of the 2 × 45 target words' Levenshtein values.

on the link between linguistic experience and vocabulary size). The test consists of 42 series of words and nonwords. The participants' task is to tick the existing German word presented alongside five orthographically and phonotactically permissible nonwords. The target words range from the educated but common (e. g., *Ironie* 'irony') to the highly arcane (e. g., *Heddur*, a type of aluminum alloy). One point was awarded for every correctly ticked target word.

2.2.3 Raven's advanced progressive matrices (Raven 1962)

This test of fluid intelligence contains 36 abstract puzzles. Each puzzle presents eight patterns in a 3-by-3 grid. The participants' task is to select the logically fitting ninth pattern from a list of eight possible patterns presented underneath the grid. One point was awarded for each correctly selected pattern.
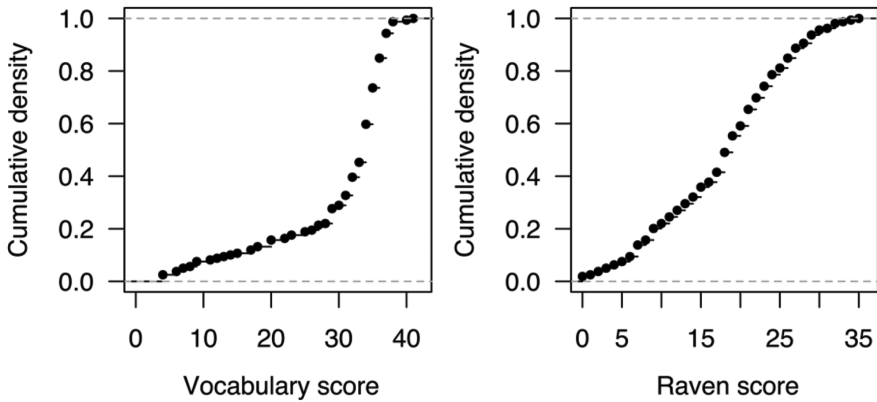
Figure 3: Cumulative density plots of the participants' vocabulary and Raven scores.

Figure 3 shows the distribution of the vocabulary test and Raven task scores. While the Raven scores are approximately normally distributed in our sample, the vocabulary scores are negatively skewed: Most participants fall into the 30-to-40 bracket. The vocabulary and Raven scores were not strongly correlated with each other ($r_s = 0.12$).

2.2.4 English proficiency test

Proficiency in additional languages that are closely related to both the L1 and the Lx often contributes positively to receptive multilingualism skills (e. g., Berthele 2011; Berthele and Lambelet 2009; Vanhove et al. 2015a; but see Bezooijen et al. 2012). A measure of proficiency in English, the only Germanic language beside Standard German and Swiss German that our participants knew, was therefore included in the analyses as a control covariate. The proficiency test consisted of a 20-item multiple-choice test and a 25-item C-test. The scores on both parts were z-normalized (i. e., expressed in standard deviations) and the two normalized scores were added together for each participant.

A measure of proficiency in the other additional language shared by almost all participants, i. e. French, was not predictive of cognate translation accuracy and will not further be discussed here for reasons of space. For a discussion, see Vanhove (2014: Section 5.2.1) and Vanhove and Berthele (2015b); for similar findings, see Vanhove and Berthele (2013: 188).

2.3 Modeling approach

2.3.1 Generalized additive models

To investigate the interactions between fluid intelligence and the vocabulary test scores on the one hand and formal distance on the other hand, we adopted a multivariate approach in which the binary correctness variable was modeled in terms of continuous predictor variables (Raven score, German vocabulary score, English proficiency, and Levenshtein distance). The 'traditional' vehicle for fitting binary dependent variables is logistic regression, which can further be furnished with random effect terms in order to specify the dependency structure of the data ('mixed-effects modeling', specifically generalized linear mixed models, or GLMMs; see Jaeger 2008). Logistic regression assumes a linear relationship between the dependent and the predictor variables, however. This assumption is relaxed in *generalized additive models* (GAMs). Like logistic regression, GAMs can cope with binary dependent variables by modeling the data in terms of log-odds, but they lend themselves better to the fitting of nonlinear relationships between the dependent and an independent variable. In general terms, this is accomplished by fitting higher-order polynomial regressions to subsets of the data, gluing the pieces together, and then applying a cross-validation procedure (or an algebraic approximation thereof) to prune back excessive nonlinearities (see Wood 2006; for an accessible introduction geared towards subject-matter researchers, see Zuur et al. 2009: Ch. 3). GAMs can also incorporate random effects up to a certain extent, giving rise to generalized additive mixed-effects models (GAMMs).[2]

The principal reason for our using GAMMs is that, in addition to modeling nonlinear main effects, they can model nonlinear interactions between two (or several) continuous independent variables. This is accomplished through the fitting of 'tensor products', the mathematical details of which need not concern us here as end users (see Wood 2006: 162–167): in essence, these tensor products generalize the fitting approach outlined above to higher dimensions. GLMMs, like traditional regression models, can accommodate linear interactions between

---

2 GAMMs can straightforwardly accommodate random intercepts (e. g., by-participant or by-item adjustments to the intercept) as well as random slopes for linear terms (e. g., by-item adjustments to the slope of the English proficiency variable). In principle, GAMMs can also accommodate the nonlinear counterpart of random slopes ('factor-smooth interactions'), but we have not been able to fit those in models with crossed dependency structures (i. e., items and participants). Research on linear mixed models has highlighted the importance of including random slopes (e. g., Barr et al. 2013; Jaeger et al. 2011; Schielzeth and Forstmeier 2009), and similar considerations may well apply to GAMMs.

continuous variables, but fitting nonlinear interactions with them is a nontrivial exercise. Their ability to fit nonlinear (as opposed to linear) interactions between continuous variables make GAMMs eminently suited for our present purpose: while we expect that the effect of formal distance on cognate guessing success varies according to vocabulary test and Raven task performance, these effects need not be linear.

2.3.2 Software and model specification

We fitted separate GAMMs for spoken and written stimuli using the bam() function in the mgcv package (version 1.8-2; Wood 2014) for R (R Core Team 2014). Only responses to target words were analyzed. The GAMMs featured (a) nonlinear main effects for Raven task score, vocabulary test score, and Levenshtein distance, (b) nonlinear interactions between Raven task score and Levenshtein distance as well as between vocabulary test score and Levenshtein distance, (c) a linear main effect for English proficiency (as a control variable), and (d) random intercepts for participants and for items. The nonlinear terms were fitted with cubic regression splines using mgcv's ti() function. Readers interested in full technical details can inspect our script (see below).

2.3.3 Interpretation

The mgcv package provides numerical information about the fitted nonlinear terms (e. g., degrees of freedom, χ^2-values, and p-values). These, however, are intended as approximations rather than perfectly accurate numbers (e. g., Wood 2006: xvii; see also Note 1). As a result, blind reliance on the traditional boundary between significant and nonsignificant p-values at $\alpha = 0.05$ is even less likely than usual (see, e. g., Cohen 1994; Gigerenzer 2004) to be a useful heuristic for gauging the generalizability of GAMM-fitted trends.

Furthermore, these numerical summaries reveal little about the shapes of the nonlinear relationships, which is why they need to be inspected visually.[3]

---

3 Readers familiar with traditional linear models will note that we do not report fit indices such as the R^2 coefficient of determination for the GAMM fits. The reason is that such fit indices (and their generalizations for non-Gaussian models) translate poorly to mixed models. We know of no R^2 pendant that is straightforwardly implementable and in widespread use for indicating the fit of models with crossed random effects (i. e., participants and items).

Nonlinear interactions can be graphed in contour plots, which are two-dimensional representations of three-dimensional surfaces in which points at the same 'height' (i. e., with the same fitted values) are connected by contour lines. Reading a contour plot of a nonlinear interaction is thus essentially the same as reading a topographic map of hilly terrain.

2.3.4 Code and data availability

A script containing the R commands that we used as well as the dataset are available from http://dx.doi.org/10.6084/m9.figshare.1172058, allowing interested readers to fully reproduce our results or carry out their own analyses. The software used is in constant development, however, and later program versions may yield somewhat different, presumably more accurate results. For unabridged data and code, see Vanhove (2014).

3 Results

3.1 Written items

We fitted a generalized additive mixed-effects model on the correctness of the translations offered by 159 participants to 45 target items (total *n*: 7,155). This GAMM featured nonlinear interactions between a measure of formal distance between the stimuli and their German and English cognates (Germanic Levenshtein distance) on the one hand and the participants' scores on an advanced German vocabulary test and Raven's advanced progressive matrices on the other hand. The fitted interactions are presented by means of contour plots in the right-hand panel of Figure 4. For comparison, the left-hand panel shows the corresponding contour plots of a GAMM without the interactions.

 The contour plots on the left-hand side can be read as follows. Going from left to right along the *x*-axis in the upper plot (irrespective of the position on the *y*-axis), the fitted values increase by somewhat more than 2 log-odds (note that the values in the upper left corner are negative: the minus sign is not part of the contour line). This represents the bare main effect of the participants' vocabulary test result. In the bottom plot, the corresponding change is less than 0.5 log-odds. This change represents the bare main effect of the participants' Raven task performance. Similarly, moving vertically along the *y*-axis in either plot, the fitted values change by slightly more than 5 log-odds, which represents the bare
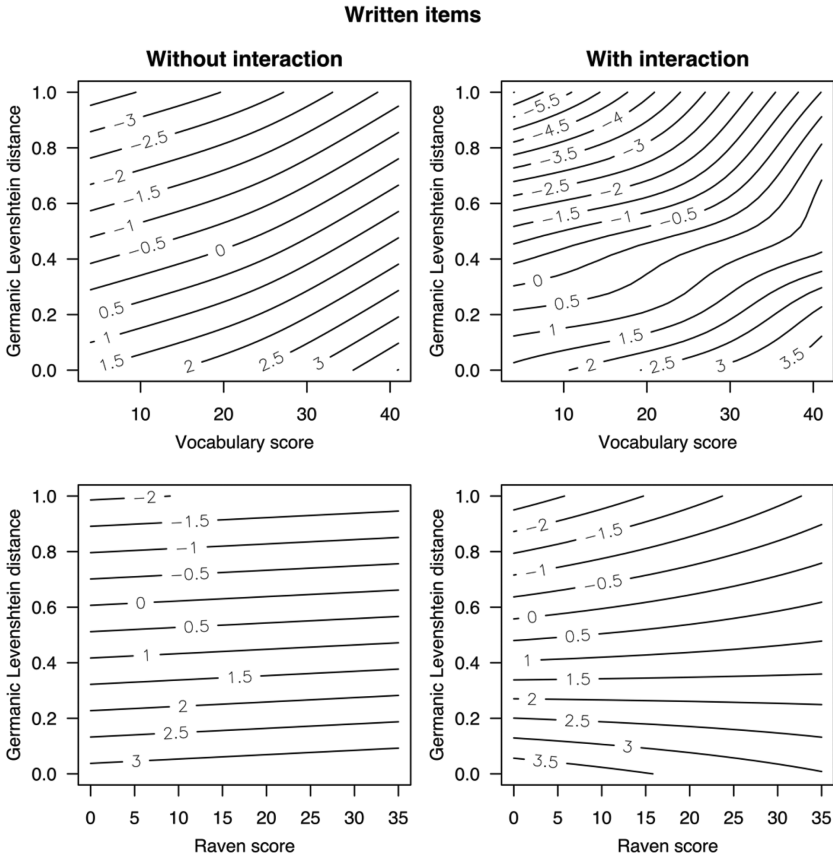
Written items

Figure 4: GAMM-fitted contour plots (written modality). The plots on the left show the effects of Levenshtein distance and vocabulary score (upper row) and of Levenshtein distance and Raven score (bottom row) without interaction terms; the plots on the right allow for interaction effects.

main effect of Germanic Levenshtein distance. Unsurprisingly, all three of these effects largely correspond to the effects reported by Vanhove (2014) and Vanhove and Berthele (2015b): They were derived from the same dataset but using a different statistical model (GAMM instead of GLMM).

For the model plotted on the right-hand side of Figure 4, the effect of Levenshtein distance and the participant-related effects were allowed to vary with respect to each other. In the upper right plot, the effect of Levenshtein distance decreases from about 7.5 log-odds to 4 log-odds as the vocabulary test results increase from 10 to 40. By the same token, going right along the x-axis, one finds an effect of the vocabulary test results of about 2 log-odds for smaller

Levenshtein distances (up till about .4), but this increase gets larger for larger Levenshtein distances: For Levenshtein distances around .8, this increase is 4.5 log-odds. This interaction between Levenshtein distance and vocabulary test results is statistically significant (χ^2(reference df: 7.0) = 15.9, p = 0.03), but we reiterate that this statistical test is an approximation.

The bottom right plot shows the interaction between Levenshtein distance and Raven task performance. The fitted effect of Levenshtein distance (plotted on the *y*-axis) decreases from more than 6 log-odds (Raven score of 0) to about 4 log-odds (Raven score of 35). The plot suggests that there may be a cross-over interaction between Raven score and Levenshtein distance: For written items with small Levenshtein distances (e. g., 0.1), the fitted values actually *decrease* with increasing Raven task performance, whereas for items with large Levenshtein distances (e. g., 0.8), one finds a roughly one-unit increase on the log-odds scale. This interaction, too, was statistically significant (χ^2(reference df: 1.0) = 10.1, p = 0.001).

3.2 Spoken items

All 159 participants responded to 45 spoken target stimuli (total *n*: 7,155). The right-hand side of Figure 5 presents the interactions between German Levenshtein distance on the one hand and vocabulary score and Raven score on the other hand. The left-hand side presents the respective bare main effects for comparison. The upper right panel of Figure 5 suggests that the effect of Levenshtein distance is smaller in participants with low vocabulary test results than in participants with high vocabulary test results – contrary to our expectations. This interaction seems to be driven primarily by items with small Levenshtein distances: Participants with vocabulary test results of about 25 and lower hardly show a Levenshtein effect for items with Levenshtein distances between 0.0 and 0.2, whereas participants with higher vocabulary test scores do show a Levenshtein effect for these stimuli. This interaction is not statistically significant, however (χ^2(reference df: 7.7) = 13.2, p = 0.10).

The interaction between Levenshtein distance and Raven score is plotted in the bottom right panel of Figure 5. This plot suggests that the effect of Levenshtein distance is roughly stable across the range of Raven scores (about 5 log-odds). A subtle interaction may be present in that the Raven score effect may differ somewhat for different Levenshtein values: Items with large (>0.7) or small (<0.1) Levenshtein distances show a small Levenshtein effect of about 0.5 log-odds. In the mid-range, i. e., for Levenshtein distances between about 0.2 and 0.6, the Raven variable has an effect of about 1.5 log-odds. This interaction, however, is not statistically significant (χ^2(reference df: 8.2) = 14.9, p = 0.07).
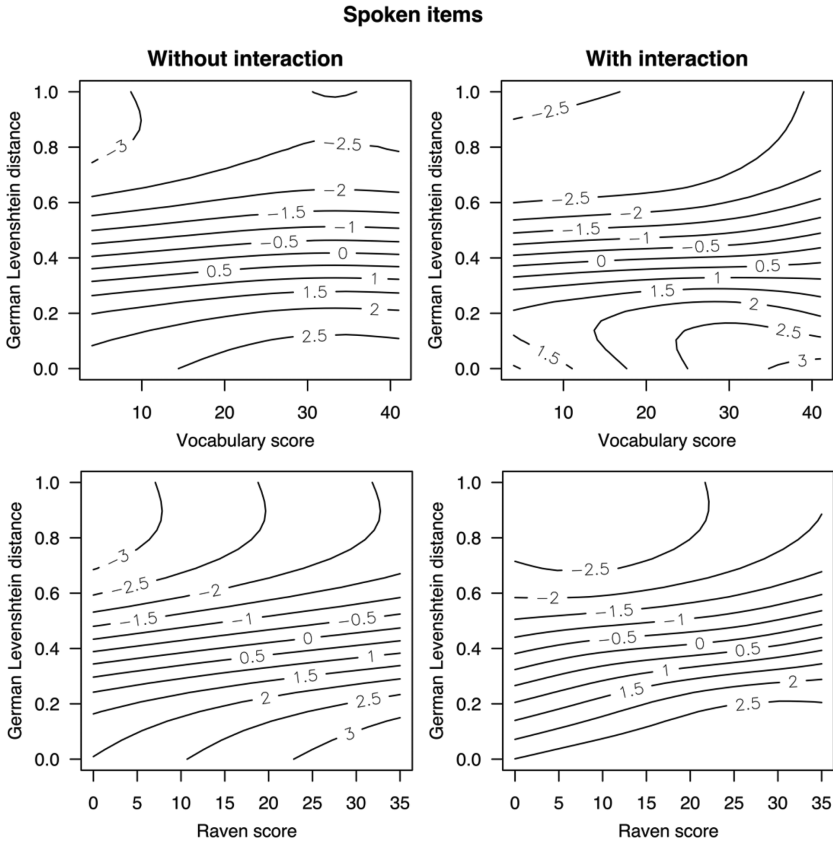
Figure 5: GAMM-fitted contour plots (spoken modality). The plots on the left show the effects of Levenshtein distance and vocabulary score (upper row) and of Levenshtein distance and Raven score (bottom row) without interaction terms; the plots on the right allow for interaction effects.

4 Discussion and conclusions

This paper investigated whether the effect of the formal distance between Lx words and their cognates in known languages on how well the words can be understood in a receptive multilingualism context varies as a function of readers' or listeners' linguistic experience and their fluid intelligence. It was predicted that formal distance would be less important in participants with higher levels of linguistic experience and fluid intelligence. These hypotheses were addressed separately for the written and for the spoken modality using data collected from 159 German-speaking participants who translated 90 Swedish words with cognates in German, English, or French.

In the written modality, nonlinear interactions fitted in generalized additive mixed-effects models yielded results in line with the expectations. First, the slope of the function relating formal distance (Germanic Levenshtein distance) to cognate translation accuracy was less steep when readers performed better on the advanced L1 vocabulary test. This interaction is particularly pronounced for Lx words showing a large orthographic distance to their German or English counterpart. This may suggest that readers with a wealth of linguistic experience show a larger degree of perceptual flexibility and are less hampered by formal differences when guessing the meaning of written Lx cognates, as suggested by Berthele (2008, 2011).

Second, readers with high fluid intelligence levels – as measured by means of the Raven task – showed a weaker effect of Germanic Levenshtein distance than did participants with low fluid intelligence levels. However, the fluid intelligence × Levenshtein distance interaction also revealed an unexpected cross-over interaction: While higher Raven scores were associated with more accurate cognate translations when the stimuli had large Levenshtein values, for stimuli with low Levenshtein values, higher Raven scores were actually associated with slightly *lower* translation accuracy. This crossover interaction may explain why Vanhove and Berthele (2015a) did not report a main effect of fluid intelligence since the negative effect partially cancels out the positive effect. It does raise the question of why higher Raven scores are associated with (slightly) lower cognate guessing performance for items with low Levenshtein values, however. One possible explanation goes as follows. Participants with high fluid intelligence levels are arguably more adept at coping with obscured cognate relationships as they can treat the stimuli's actual forms more flexibly by abstracting away from formal discrepancies and instead establishing similarities. When cognate relationships are obscured, this is usually advantageous, but when they are not, this flexibility may backfire. For instance, showing some flexibility with regard to formal discrepancies is necessary when decoding the stimulus *förutsättning* (Gm. *Voraussetzung* 'requirement'). When decoding a stimulus like *hård* 'hard', not much flexibility is required; in fact, decoders who are too flexible can sometimes come up with incorrect translations such as *Herz* 'heart' or *Herd* 'stove'. That said, the negative effect of fluid intelligence is relatively small.

With respect to the spoken modality, our expectations were not borne out. First, participants with high vocabulary test scores actually showed a stronger Levenshtein effect in our sample. Second, the effect of Levenshtein distance was more or less the same for different Raven scores, and the two variables only seemed to interact in specific regions of the contour plot. Neither of the two interactions was statistically robust.

In general terms, our findings at the very least suggest that the effect of formal distance – a classic variable in research on receptive multilingualism (see Introduction) – is not merely a matter of the stimuli used in the cognate translation task but also one of the participants who take the task. Moreover, they indicate that some of the between-participant variability in the formal distance effect is systematically related to their levels of linguistic experience and fluid intelligence – at least in the written modality. In this respect, this study not only underscores Kellerman's (1977, 1983) classic distinction between purely formal and perceived similarities but also that the relationship between the two may vary systematically between participants with the same L1 background.

Two caveats are in order, however. First, as discussed in the Method section, borderline significant results (e. g., $p = 0.03$) are to be treated with caution – even more so than usual – given the approximate nature of the test. Second, the distribution of both the Levenshtein and the participant-related variables was such that the number of data points was low in some regions of the contour plots (see Figures 2 and 3). For both of these reasons, the true test of the findings reported here should be a more targeted study in which more items from the whole formal distance continuum are used. In conclusion, however, we maintain that research on receptive multilingualism can gain more nuanced insights by jointly considering both item- and participant-related variables and their interplay.

References

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

Beijering, Karin, Charlotte Gooskens & Wilbert Heeringa. 2008. Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. *Linguistics in the Netherlands* 25. 13–24.

Berthele, Raphael. 2008. Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. Erkenntnisse zum interlingualen Potenzial des Provinzlerdaseins. *Sociolinguistica* 22. 87–101.

Berthele, Raphael. 2011. On abduction in receptive multilingualism: Evidence from cognate guessing tasks. *Applied Linguistics Review* 2. 191–220.

Berthele, Raphael & Irmtraud Kaiser (eds.). 2014. Multilingualism across the lifespan [Special issue]. *Bulletin suisse de linquistique appliquée* 99.

Berthele, Raphael & Amelia Lambelet. 2009. Approche empirique de l'intercompréhension: répertoires, processus et résultats. *Lidil. Revue de linguistique et de didactique des langues* 39. 151–162.

Bezooijen, Renée van & Charlotte Gooskens. 2005. How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why? *Linguistics in the Netherlands* 22. 13–24.

Bezooijen, Renée van, Charlotte Gooskens & Sebastian Kürschner. 2012. Deens is makkelijker voor Friezen dan voor Nederlanders – feit of fabel? [Danish is easier for Frisians than for the Dutch – truth or tale?] In Piter Boersma, Goffe T. Jensma & Reinier Salverda (eds.), *Philologia Frisica anno 2008. Lêzings fan it achttjinde Frysk Filologekongres fan de Fryske Akademy op 10, 11 en 12 desimber 2008*, 286–298. Leeuwarden: Afûk/Fryske Akademy.

Cohen, Jacob. 1994. The Earth is round (*p* < .05). *American Psychologist* 49(12). 997–1003.

Doetjes, Gerard & Charlotte Gooskens. 2009. Skriftsprogets rolle i den dansk-svenske talesprogsforståelse [The role of the written language in Danish-Swedish speech comprehension]. *Språk och stil* 19. 105–123.

Gigerenzer, Gerd. 2004. Mindless statistics. *Journal of Socio-Economics* 33(5). 587–606.

Gooskens, Charlotte. 2007. The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development* 28(6). 445–467.

Gooskens, Charlotte, Sebastian Kürschner & Renée van Bezooijen. 2011. Intelligibility of Standard German and Low German to speakers of Dutch. *Dialectologia* Special issue, II. 35–63. http://www.publicacions.ub.edu/revistes/dialectologia/

Heuven, Vincent J. van. 2008. Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing* 2. 39–62.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446.

Jaeger, T. Florian, Peter Graff, William Croft & Daniel Pontillo. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15(2). 281–320.

Jarvis, Scott. 2014. Lexical processing and production in an L3. Plenary address at the *9th International Conference on Third Language Acquisition and Multilingualism*, Uppsala, 12–14 June.

Kellerman, Eric. 1977. Towards a characterisation of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin* 2. 58–145.

Kellerman, Eric. 1983. Now you see it, now you don't. In Susan Gass & Larry Selinker (eds.), *Language transfer in language learning*, 112–134. Rowley, MA: Newbury House.

Kuperman, Victor & Julie A. Van Dyke. 2013. Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance* 39(3). 802–823.

Kürschner, Sebastian, Charlotte Gooskens & Renée van Bezooijen. 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal of Humanities and Arts Computing* 2. 83–100.

Möller, Robert. 2012. Germanic sound correspondences in didactic presentation and in reality. Paper presented at *Sociolinguistics Symposium* 19, Berlin, 21–24 August.

Möller, Robert & Ludger Zeevaert. 2010. "Da denke ich spontan an Tafel": Zur Worterkennung in verwandten germanischen Sprachen. *Zeitschrift für Fremdsprachenforschung* 21. 217–248.

Raven, John C. 1962. *Advanced progressive matrices, set II*. London: H. K. Lewis.

R Core Team. 2014. R: A language and environment for statistical computing. Software, version 3.1.1. http://www.r-project.org/

Schielzeth, Holger & Wolfgang Forstmeier. 2009. Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology* 20(2). 416–420.

Schmidt, K.-H. & P. Metzler. 1992. *Wortschatztest (WST)*. Weinheim: Beltz Test.

Teleman, Ulf. 1981. Om förståelse i allmänhet och om förståelse via grannspråken i synnerhet. [On comprehension in general and on comprehension via the neighbouring languages in specific.]. In Claes-Christian Elert (ed.), *Internordisk språkförståelse. Föredrag och diskussioner vid ett symposium på Rungstegaard utanför Köpenhamn den 24–26 mars 1980*, 102–120. Umeå: Umeå University.

Vanhove, Jan. 2014. *Receptive multilingualism across the lifespan. Cognitive and linguistic factors in cognate guessing*. Fribourg: University of Fribourg dissertation. http://doc.rero.ch/record/210293

Vanhove, Jan & Raphael Berthele. 2013. Factoren bij het herkennen van cognaten in onbekende talen: algemeen of taalspecifiek? [Factors in recognizing cognates in unknown languages: General or language-specific?] *Taal en Tongval* 65. 171–210.

Vanhove, Jan & Raphael Berthele. 2015a. The lifespan development of cognate guessing skills in an unknown related language. *International Review of Applied Linguistics in Language Teaching* 53(1). 1–38.

Vanhove, Jan & Raphael Berthele. 2015b. Item-related determinants of cognate guessing in multilinguals. In Gessica De Angelis, Ulrike Jessner & Marijana Kresić (eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning*, 95–118. London: Bloomsbury.

Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.

Wood, Simon N. 2014. mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. R package, version 1.8-2. http://cran.r-project.org/package = mgcv

Zobl, Helmut. 1992. Prior linguistic knowledge and the conversation of the learning procedure: Grammaticality judgments of unilingual and multilingual learners. In Susan M. Gass & Larry Selinker (eds.), *Language transfer in language learning* (Rev. edn.), 176–196. Amsterdam: John Benjamins.

Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.