Title: A multi-metric analysis of 50,000 linguistic profiles provides sparse evidence that language distance modulates bilingual cognition

Short title: Language distance in bilingualism

Authors and Affiliations: Evelina Leivada,[1,2] Fritz Günther,[3] Camilla Masullo,[4] Jon Andoni Duñabeitia,[5] Marit Westergaard[6,7] & Jason Rothman[5,6,8]

[1]Universitat Autònoma de Barcelona
[2]Institució Catalana de Recerca i Estudis Avançats (ICREA)
[3]Humboldt-Universitat zu Berlin
[4]Università degli Studi di Milano-Bicocca
[5]Universidad Nebrija
[6]UiT-The Arctic University of Norway
[7]Norwegian University of Science and Technology (NTNU)
[8]Lancaster University

Abstract:

Similarity in mental, linguistic representations modulates the degree of recruitment of cognitive control mechanisms, which have been linked to neurocognitive adaptations in bilingual populations. While ample evidence exists for this claim, its coverage is limited, as testing is geared towards WEIRD communities that use sizeable, Indo-European languages, thus potentially providing a biased view of bilingual cognition. We assess the role of distance as a key moderator of bilingual adaptations through a large-scale aggregation analysis of 510 experiments. To measure distance, we develop a multi-metric approach, using state-of-the-art databases, such as Grambank. Analyzing data from 56,122 participants who speak 79 different languages, spanning 11 language families and a language isolate, we find sparse evidence for a distance effect. Our results suggest that moderators such as language distance can shed light on the cognitive divide between language and dialects in a way that addresses the perennial question of what makes bilinguals distinct.

Teaser:

Language distance affects the bilingual brain in subtle, variable, and so far, undetermined ways

A multi-metric analysis of 50,000 linguistic profiles provides sparse evidence that language distance modulates bilingual cognition

Introduction

Many studies provide evidence that significant amounts of bilingual/multilingual experience calibrate to enhanced performance in certain cognitive measures, especially in attention and executive functioning, and adaptations in brain structures and neural networks where language and executive control overlap (1-6). At the same time, some studies, mostly behavioral ones, find null or negative results (7, 8). This variation has led to claims that bilingualism *per se* is unlikely to be a bona fide moderator of neurocognition (7, 9, 10). Under such a view, the precise conditions, if any at all, under which neurocognitive adaptations to bilingualism occur are anticipated to be rare and are yet to be determined. Can dual/multiple language experience affect how one processes information, a phenomenon often referred to in the literature as the *bilingual advantage hypothesis* (11)? The conflicting evidence in the empirical record entails no current position has yet succeeded in providing a universally accepted, ecologically valid theory predicting and explaining the overall body of evidence in a way that accounts for the *source* of bilingual adaptations (12). Consequently, the evidence for the type and the origin of bilingual adaptations —whether they have been labelled as advantages, disadvantages, or, more neutrally, as bilingual effects or trade-offs— has been called into doubt as “discrepant” (8), “contradictory” (13), and “vexingly inconsistent” (9), while the explanations offered for it have been described as “deceptively simple” (14), “showing a lack of theoretical specification” (15), and leading to debate even over the essential assumptions (16).

What potentially makes bilinguals different? We refer to this question as the *Source Problem*: The biggest challenge for the field of bilingualism boils down to a lack of consensus about the source of bilingual effects (3, 12, 15, 17, 18). Undoubtedly, the systematic use of more than one language entails an additional level of complexity that should in principle influence processing in multiple ways. Although this claim is generally accepted as true, it generates more questions than answers. What *exactly* is it about being bilingual that drives this influence? What exactly is affected and what are the implicated mechanisms? Do bilingual effects stem from handling two grammatical and lexical systems? If so, do (dis)similarities in the systems matter? If they do, what is the directionality of the effect, in terms of increasing system overlap mitigating or intensifying bilingual adaptations?

Almost all theories about the bilingual mind rely on one important yet unclear assumption: *language distance* (also referred to as *language similarity* or *proximity* in the literature) affects processing and modulates the degree of recruitment of cognitive control mechanisms (19-23). According to models of bilingual lexical access (e.g., BIA/BIA+; 24), increased lexical form similarity yields higher levels of activation for the two words in the mental lexicon. At the sublexical level too, contrasting differences between a bilingual’s languages may result in reduced cross-language activation (25). Syntactic processing accounts further highlight the role of similar representations. According to the Language Distance Hypothesis (26), unshared grammatical properties pose a challenge for native-like syntactic processing, predicting differences for bilinguals whose languages are separated by a greater degree of grammatical differences. At the phonological level too, languages that have non-overlapping inventories show a greater degree of separation (27).

While the role of language distance in bilingualism is frequently evoked, our understanding of how it works is extremely limited. In recent years studies have specifically sought to test the role language distance brings to bear in modulating neurocognitive adaptations to bilingualism, especially in inhibition and other executive function measures (23, 28, 29). While many studies argue that language distance matters, reports about the *directionality* of the effects are mixed: Some studies find that bilingual adaptations are more pronounced in bilinguals that use similar languages (30-32), while others find this for bilinguals that use distant ones (33-35).

This approach of measuring the impact of distance on bilingual cognition through comparing bilingual groups using language pairs that can be classified as similar (e.g., Spanish-Catalan) or dissimilar (e.g., Spanish-Basque) suffers from at least three limitations. First, it typically involves a small pair (i.e. two or three) of bilingual/multilingual groups, usually speakers of Indo-European languages. Thus it lacks ecologically valid, wide population coverage. Cognitive science is predominantly geared towards testing Western, Educated, Industrialized, Rich, and Democratic (WEIRD) communities (36). Since WEIRD populations are privileged, working exclusively with them as targets for experimental testing provides a biased view of human language and cognition (37-40).

Second, this approach often relies on crude phylogenetic relationships in order to attribute the labels 'similar' or 'distant' to different groups. For instance, Spanish and Catalan are perceived as similar because they are both Romance languages. However, using language family as a proxy for language similarity may lead to unwarranted conclusions (41). For example, based on phylogenetic assumptions alone, one could expect that Irish, which belongs to the Indo-European language family, is closer to Spanish (Indo-European family) than to Arabic (Afroasiatic family). Yet according to the Ceolin et al. (42) database that charts universal morphosyntactic differences in the nominal domain, Irish and Spanish have a Jaccard Distance index of 0.37 (see section Methods for details about the formula), which is identical to the index of Irish and Arabic. Overall, while various studies compare the magnitude of bilingual adaptations in bilinguals that speak pairs of languages that differ in terms of phylogenetic proximity (e.g., English-French/Farsi in 23, Serbian-Slovak/Hungarian in 35), they usually do not measure how similar or distant the language systems actually are. Moreover, no study has tapped into the impact of language distance on bilingual adaptations in a way that goes beyond comparing a small number of bilingual groups.

Third, the few studies that do measure distance using a specific metric often measure it in a highly constrained way, through tapping into one level of linguistic analysis (e.g., phonology; 41). The challenge is that two languages that are very similar at the phonological level (e.g., Spanish and Basque) are not necessarily grammatically similar or globally similar (i.e. across multiple other levels of linguistic analysis). In the absence of a metric that covers different levels of linguistic analysis and that can be readily applied to different populations providing empirical coverage that goes beyond WEIRD samples, how language distance is to be conceptualized "is a fundamental question that has yet to be discussed seriously in the realm of bilingualism" (43:4).

All in all, while the answer to what drives bilingual adaptations is unlikely to be monolithic (44), among the potential variables that may modulate such adaptations, language distance holds a particularly unclear role. While an increasing number of (recent) studies find evidence for distance effects in bilingualism, any directionality of such effects cannot be deciphered in the absence of adequate cross-linguistic coverage that spans different language pairings and families. The aim of the present work is to address this important gap. Taking advantage of the development of state-of-the-art databases that chart the global distribution of language universals, grammatical features, functional dependencies, and interactions between language, cognition, and culture (e.g., Grambank; 45), we aggregate different meta-analyses that track bilingual adaptations in different populations, and then map their results to four language metrics that measure language distance in either morphosyntax or the lexicon. This amounts to the first systematic, large-scale, multi-metric analysis of the role of language distance in bilingual cognition.

The Research Questions (RQs) are the following:

(RQ1). Is there evidence for a modulating role of language distance in cognitive adaptations to bilingualism when one performs large-scale mapping of different bilingual populations to precise measures of language distance?

(RQ2). Do different databases of lexical inventories or morphosyntactic features correlate when used as metrics of language distance?

Results

The complete dataset consists of 510 datapoints representing 295 different studies on bilingual adaptations, comprising 79 different language pairs. Figure 1 shows the geographical distribution of the tested bilingual populations. Reflecting the imbalance of the published literature record, the distribution of language pairs differs substantially; there are 148 datapoints on the English-Spanish pair, 49 on English-Mandarin, 42 on English-French, 23 on Catalan-Spanish, and 16 on English-Welsh. For all other language pairs, we have fewer than 15 datapoints. All pairs were measured in terms of distance, using four different databases: Grambank (45), the Parametric Comparison Method (PCM; 42), EZ Glot (46), and the Similarity Database of Modern Lexicons (SDML; 47). Grambank and PCM tap into universal grammatical features, and EZ Glot and SDML into the lexicon. In the complete dataset, we have similarity scores based on Grambank for 50 language pairs (440 datapoints), based on PCM for 39 language pairs (416 datapoints), based on EZ Glot for 55 language pairs (456 datapoints), and based on SDML for 66 language pairs (490 datapoints).

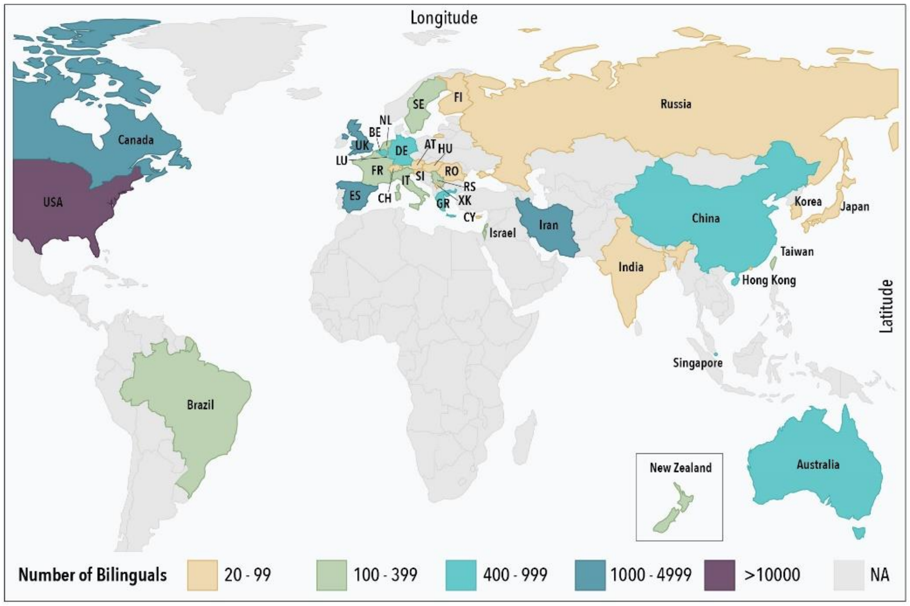

Figure 1: The geographical distribution of the bilingual population size (number of individuals) included in our dataset

Each datapoint corresponds to a study that measures bilingual adaptations. For every study, we coded, as binary variables, the presence or absence of a finding that could be framed as a bilingual adaptation argued to be positive (often referred to as bilingual advantage), the presence or absence of a negative adaptation (bilingual disadvantage), or the absence of an effect. In what follows, we always report two types of analyses: First, we test whether the language similarity scores predict the likelihood of a positive adaptation. In a second step, we test whether they predict any bilingual effect at all, either positive or adverse. The reason for calculating separately the likelihood of observing bilingual adaptations that have been framed as advantages has to do with possible biases that favor the publication of positive results (see Methods). Figure 2, panel A, provides an overall picture of the language pairs and their likelihood of being linked to bilingual adaptations, whereas panels B-E show the distance of the language pairs that are measurable in all metrics.

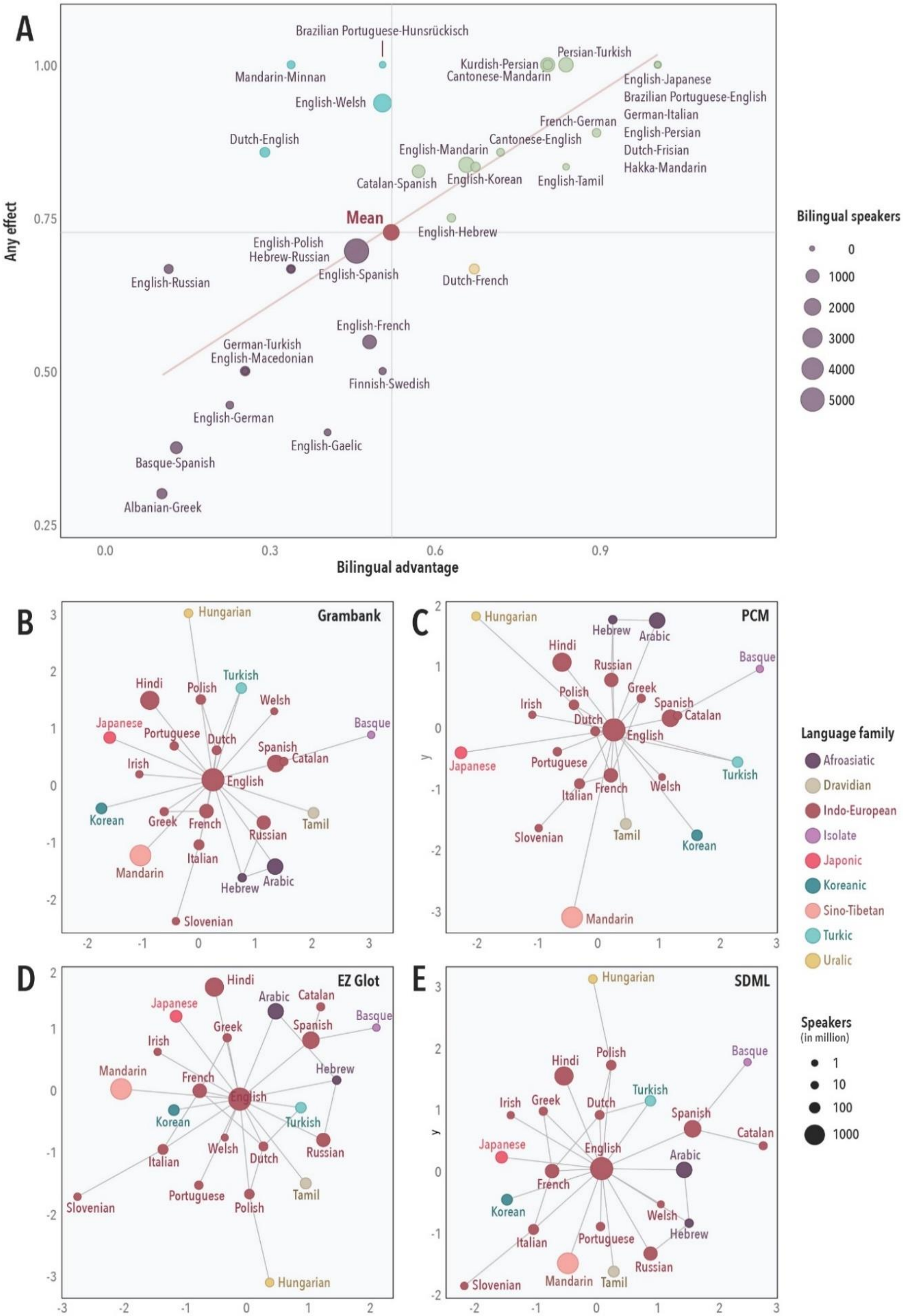
A
Brazilian Portuguese-Hunsrückisch
Mandarin-Minnan
Kurdish-Persian
Persian-Turkish
Cantonese-Mandarin
English-Japanese
Brazilian Portuguese-English
German-Italian
English-Persian
Dutch-Frisian
Hakka-Mandarin
English-Welsh
French-German
Dutch-English
Cantonese-English
English-Mandarin
Catalan-Spanish
English-Korean
English-Tamil
Mean
English-Hebrew
English-Polish
Hebrew-Russian
English-Spanish
English-Russian
Dutch-French
English-French
German-Turkish
English-Macedonian
Finnish-Swedish
English-German
English-Gaelic
Basque-Spanish
Albanian-Greek
Any effect
Bilingual advantage
Bilingual speakers
0
1000
2000
3000
4000
5000
B
Grambank
C
PCM
D
EZ Glot
E
SDML
Hungarian
Hindi
Polish
Turkish
Welsh
Japanese
Portuguese
Dutch
Basque
Spanish
Catalan
Irish
English
Korean
Greek
French
Russian
Tamil
Italian
Mandarin
Hebrew
Arabic
Slovenian
y
Language family
Afroasiatic
Dravidian
Indo-European
Isolate
Japonic
Koreanic
Sino-Tibetan
Turkic
Uralic
Speakers
(in million)
1
10
100
1000

Figure 2: (A) The classification of language pairs in terms of their findings. If a pair is left of the vertical line, it is less likely than average to show positive adaptations. If a pair is right of the vertical line, it is more likely than average to show this. If a pair is below the horizontal line, it is less likely than average to show any effect. If a pair is above the horizontal line, it is more likely than average to show any effect. If a pair deviates from the diagonal, this entails absence of balance between possible outcomes: “advantageous” vs. “any effect”. For example, the English-Russian pair has been linked to a lower rate of positive bilingual adaptations compared to other pairs. This must mean it produces more than the expected rate of adverse effects (“disadvantages”). For this scatterplot, only language pairs with at least 3 datapoints have been considered. The number next to each pair denotes the number of analyzed datapoints. (B)-(E) Language distance per metric for all the language pairs that can be measured in all 4 metrics. The networks are two-dimensional projections of linguistic system dissimilarity. The length of the lines denotes distance.

Individual analyses of the different metrics. In the first set of analyses, we tested the effect of each of the four metrics individually. To this end, we employed Generalized Linear Mixed Effect Models (48), using the *R* packages *lme4* (49) and *lmerTest* (50). The models predict the likelihood of observing a bilingual effect (whether positive or any effect at all; see below) from a fixed effect of the metric in question, as well as random effects for the language pair, in a logistic regression. When predicting the likelihood of observing a *bilingual advantage*, we find no effect of (language similarity measured through) Grambank (β = 0.43, z = 0.29, p = .768), no effect of PCM (β = -0.94, z = -0.90, p = .368), no effect of EZ Glot (β = -2.31, z = -1.57, p = .116), and no effect of SDML (β = -1.24, z = -0.27, p = .788). When predicting the likelihood of observing *any bilingual effect*, we again find no effect of Grambank (β = 0.67, z = 0.48, p = .629), no effect of PCM (β = -0.14, z = -0.13, p = .896), no effect of EZ Glot (β = -0.43, z = -0.34, p = .729), and no effect of SDML (β = -0.30, z = -0.07, p = .944).

Simultaneous analyses of the different metrics. Having tested the effects of the metrics individually, we then tested them in combination to reveal whether their synergistic interplay modulates bilingual adaptations. To do so, we selected the subset of our dataset for which all four metrics are available (i.e. some databases do not classify certain languages, hence not all language pairs can be captured by all metrics). This subset includes 373 datapoints and comprises 29 different language pairs. For these 29 language pairs, our four metrics show medium to (very) high correlations (Figure 3). This means that, when similarity in one of the metrics increases, the other three similarity metrics tend to linearly increase alongside it, which will impact the interpretation of parameter estimates due to high multicollinearity.

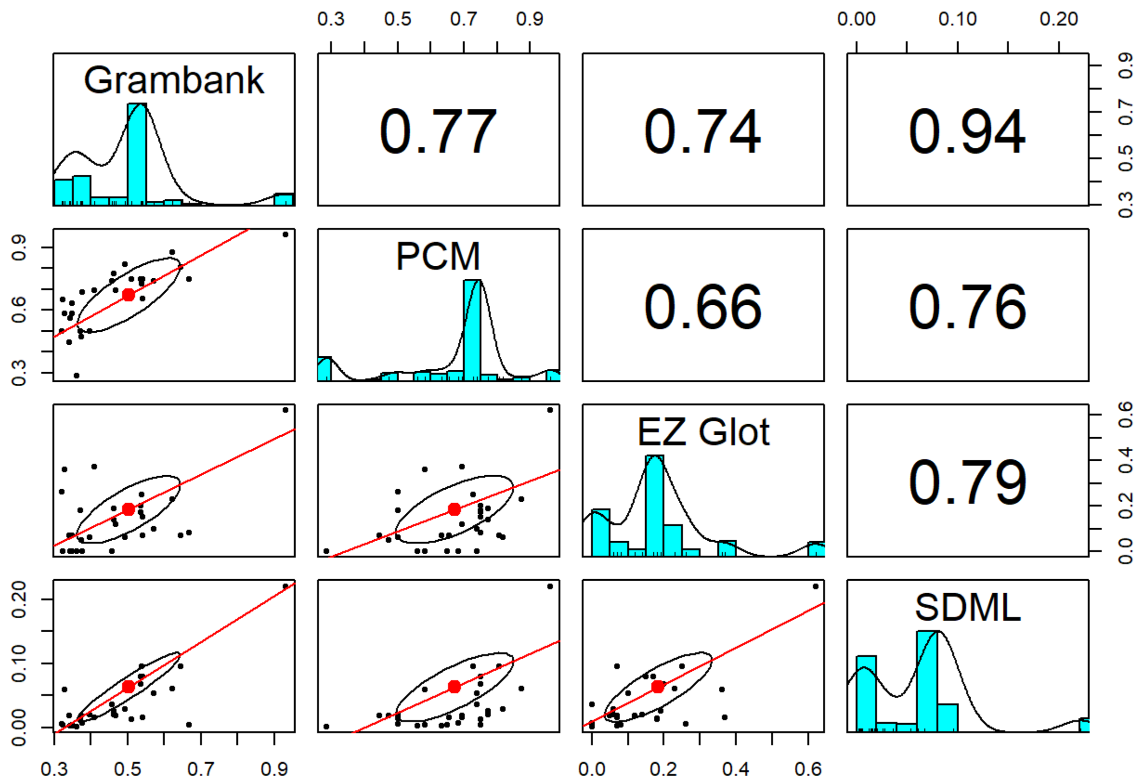

Figure 3: Distributions and correlations for the 29 language pairs which are measured in all four metrics, with the diagonal showing the individual distributions, the lower triangle showing scatterplots of the relation between the metrics in the respective rows and columns, and the upper triangle showing the Pearson correlation between the metrics in the respective rows and columns.

Since the subset of data considered here is smaller than the dataset on which we tested the individual effects, we will first re-test those individual effects, before conducting combined tests on all four metrics. Starting with the likelihood of observing a positive adaptation, the effect of Grambank (β = -0.12, z = -0.11, p = .916), EZ Glot (β = -1.16, z = -1.06, p = .288), and SDML (β = -1.54, z = -0.54, p = .592) remain non-significant. However, in this dataset the effect of PCM is significant (β = -1.44 z = -2.38, p = .017), with fewer studies finding positive adaptations for larger values of PCM. In other words, the higher the similarity value (as measured by the PCM), the less likely it is for the populations that speak these similar languages to be linked to positive adaptations that can be framed as bilingual advantages.

For testing the simultaneous effects of all four metrics, we first fit a GLMM containing parameters for all four metrics as fixed effects, as well as random effects for the language pair. Then, using step-wise backwards selection, we remove non-significant fixed effects from this model until only significant effects remain. This results in a GLMM containing only a significant negative effect for PCM.

Applying the same procedure for the likelihood to observe any bilingual effect, we again find no effect of Grambank (β = 0.49, z = 0.40, p = .689), no effect of PCM (β = -0.40, z = -0.36, p = .719), no effect of EZ Glot (β = 0.15, z = 0.13, p = .893), and no effect of SDML (β = -0.94, z = -0.29, p = .771) individually in the reduced dataset. As the result of the stepwise selection, we however arrive at a GLMM that contains a significant positive effect for Grambank (β = 6.80, z = 2.48, p = .013) alongside a significant negative effect for SDML (β = -18.97, z = -2.64, p = .008). The positive effect of Grambank means that the higher the similarity measured through Grambank, the higher the likelihood of finding a bilingual effect. The negative effect of SDML means that the higher the similarity measured through SDML, the more unlikely it becomes to find a bilingual effect. However, due to the very high positive correlation of r =.94 between these two metrics (Figure 3), in most cases, these effects cancel each other out (Figure 4 where almost all data points lie in the same band of predicted values). To be more precise, the effects play out in a few cases where one similarity value is very high while another is very low for the same language pair. For example, the Spanish-Basque pair shows a low similarity value in morphosyntax, measured through Grambank, and a high degree of cognates, measured through SDML. The model links this language pair with a lower rate of bilingual adaptations. Conversely, a high Grambank similarity value and a low SDML similarity value (e.g., English-Greek) predicts a higher rate of bilingual adaptations.

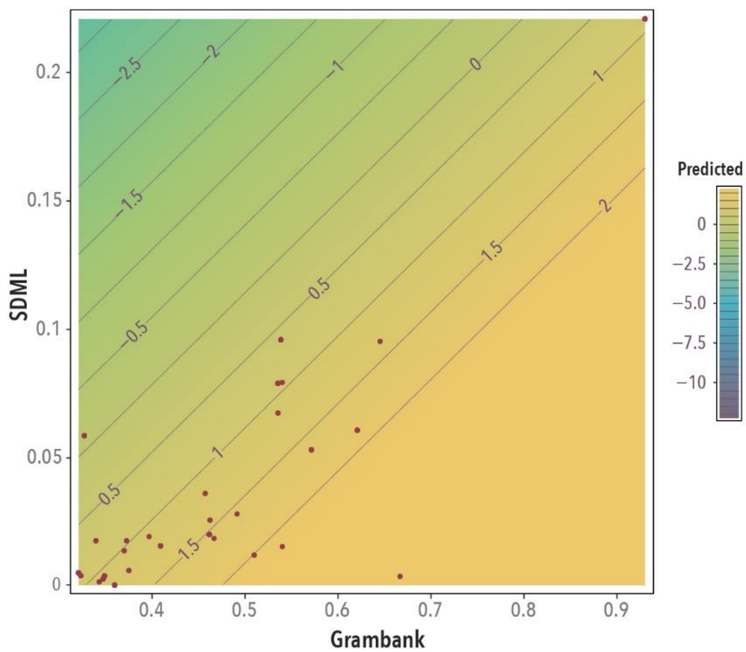

Figure 4: Simultaneous effects of Grambank (x-axis) and SDML (y-axis), with the predicted likelihood of observing any bilingual adaptation on the z-axis (encoded by the color, expressed on a logit scale). Each red point represents the Grambank score and the SDML score of one individual datapoint. As can be seen, most points lie within the same bands of predicted values (between 0.5 and 1.5), with only a few exceptions.

Interactions with sample size. In the next set of analyses, we test for interaction effects between sample size and the four metrics individually, to examine if there might be metrics that only show an effect in studies employing larger sample sizes, which provide a more reliable estimate of bilingual adaptations. To this end, we again employ the datasets described at the beginning of this section. For this analysis, 7 datapoints had to be excluded since they do not clearly mention a sample size of bilinguals tested.

Starting by only considering the likelihood of observing a positive adaptation, and again adopting the same GLMM structure as described above, we observed no effect of sample size *per se* (β = -0.003, z = -1.66, p = .095). To test for interactions, we then estimated for each metric individually GLMMs including as a fixed effect the interaction between the similarity metric and sample size, as well as random effects for the language pairs. We observed no significant interaction between sample size and Grambank (β = 0.02, z = 0.96, p = .334), EZ Glot (β = 0.02, z = 1.10, p = .270), or SDML (β = 0.04, z = 0.70, p = .484). On the other hand, we do observe a positive interaction term between PCM and the number of bilinguals (β = 0.08, z = 2.53, p = .011), with both main effects in the model being negative (β = -3.60, z = -2.61, p = .008 for PCM; β = -0.06, z = -2.66, p = .008 for sample size). These effects are displayed in Figure 5 (panel A). As the figure shows, for the sample size range of the vast majority of studies considered here (between 0 and about 100), these main effects and the interaction effect essentially manifest in the way that they cancel each other out: According to the main effects, an increase in similarity as measured through the PCM is associated with a lower likelihood of observing positive adaptations (i.e. bilingual advantages), as is an increase in sample size; however, when *both* increase at the same time, this is offset by a simultaneous and additive increase in likelihood. In practice, this leads to very similar GLMM predictions across the range of PCM values and sample sizes included in the dataset (as can be seen in Figure 5, panel A, nearly all points representing the available studies lie within the same band of predicted likelihoods). Thus, while a few datapoints indicate that positive adaptations are less likely for low values of similarity as per the PCM database and larger sample sizes (the points in the bluer areas in the middle to upper left of panel A), and more likely for high values of PCM and larger sample sizes (the points in the more yellow areas in the middle to upper right), the very small number of studies with these predictor values only allows for highly tentative interpretations of this finding.

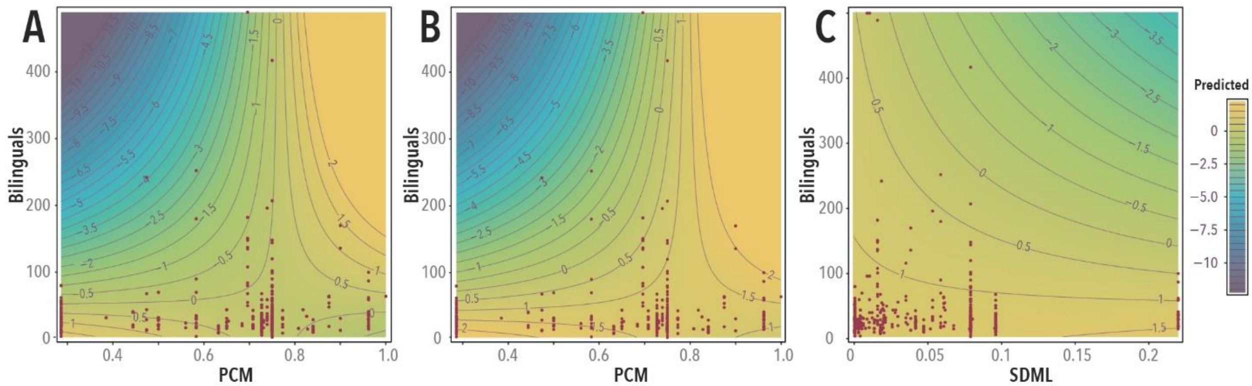

Figure 5: (A) Interaction effect between PCM (x-axis) and sample size (y-axis), with the predicted likelihood of observing a positive adaptation on the z-axis (encoded by the color, expressed on a logit scale). Each red point represents the PCM similarity score and the sample size of one individual datapoint. (B) Interaction effect between PCM (x-axis) and sample size (y-axis), with the predicted likelihood of observing any bilingual effect on the z-axis. Each red point represents the PCM similarity score and the sample size of one individual datapoint. (C) Interaction effect between SDML (x-axis) and the sample size (y-axis), with the predicted likelihood of observing any bilingual effect on the z-axis. Each red point represents the SDML similarity score and the sample size of one individual datapoint.

When repeating the same analysis for the likelihood of observing any bilingual effect, we again find no effect of the sample size *per se* (β = -0.002, z = -1.14, p = .254). We also observe no significant interaction between sample size and Grambank (β = 0.01, z = 0.45, p = .650) or EZ Glot (β = -0.01, z = -0.84, p = .402). As in the previous analysis, we do find a significant interaction between PCM and sample size (β = 0.07, z = 2.19, p = .028), with a significant negative main effect of sample size (β = -0.05, z = -2.41, p = .016), but no main effect of PCM (β = -2.52, z = -1.57, p = .115). According to these effects, when sample size increases, the likelihood of finding a bilingual effect decreases, but this is offset by a simultaneous (additive) increase in likelihood when both sample size and PCM-measured similarity increase at the same time. These effects are displayed in Figure 5 (panels A and B). As can be seen, when comparing the two panels, this interaction plays out in the same way as described above for positive adaptations, with the same limitations.

Unlike in the previous analysis, for the likelihood of observing any effect, we also observe a significant interaction between SDML and sample size (β = -0.05, z = -26.27, p = < .001; see Figure 5, panel C. Similar to the interaction effect for PCM in panels A and B, we again find that the vast majority of datapoints lie within the same band of predicted likelihoods (lower left). The pattern of the remaining points suggests that the likelihood of observing any effect gets slightly lower for studies on language pairs with lower SDML similarity that employ a larger number of participants (the upper and middle left in panel C), and conversely slightly higher for studies on language pairs with high SDML similarities that employ a smaller number of participants (the lower right). However, as for the PCM interaction, this finding again is based on a small number of datapoints.

Analyses of language family similarity. In the final set of analyses, we approach language distance from a phylogenetic perspective, by testing whether bilingual adaptations are more likely for pairs of languages belonging to the same language family or branch. To this end, we estimated a GLMM including as predictors a fixed effect for a three-level factor "phylogenetic relationship", with the levels "different family", "different branch", "same branch", as well as random effects for the language pair. Being part of the same branch entails being part of the same family, so "different branch" and "same branch" imply "same family". To test for the factor-level effect of "relationship", we compared this model to a model that only included an intercept and the random effects using a likelihood ratio test. These models were estimated on all 510 datapoints in the dataset.

When predicting the likelihood of observing a positive adaptation, the model including the "relationship" effect did not perform better than the model without this effect ($X^2(2) = 0.49$, $p = .783$). Accordingly, neither the "different branch" nor the "same branch" condition differed from the "different family" condition ($\beta = -0.27$, $z = -0.60$, $p = .548$ and $\beta = -0.001$, $z = -0.003$, $p = .996$, respectively). The same was the case when predicting the likelihood of observing any bilingual effect ($X^2(2) = 1.26$, $p = .532$ for the likelihood-ratio test; with non-significant differences between "different family" and "different branch", $\beta = 0.09$, $z = 0.21$, $p = .835$, or "same branch", $\beta = 0.54$, $z = 1.03$, $p = .302$).

Discussion

Returning to the RQs that guide this work, RQ1 asks whether large-scale, multi-metric investigations of language distance corroborate the popular claim that linguistic system similarity modulates the degree of recruitment of cognitive control mechanisms (e.g., 20, 21). Recently, there has been a significant upsurge in the number of studies that provide evidence in favor of this claim (23, 28, 31-35, 51). While reaching different conclusions concerning the *directionality* of the observed distance effects in terms of being more pronounced in similar vs. distant language bilinguals, these studies share two important properties. First, they find evidence for a distance effect. Second, they tap into distance through comparing the performance of a small number of bilingual populations, often not more than two: one bilingual group that speaks "similar" languages is compared to another bilingual group that speaks "less similar" languages, using the same set of tasks. Not least because such an approach lacks ecological coverage, to remedy this and begin to understand any degree of generalizability one could apply from the findings of individual studies, we performed a large-scale mapping of two recent meta-analyses to four different datasets that chart the global distribution of language features and thus serve as universal metrics for language distance.

The present results suggest that some evidence for the role of language distance in modulating bilingual adaptations exists, but this is sparse. Thus, caution should be applied when interpreting the claims made on the basis of any given singular study, even when such an effect seems to replicate in other singular studies. Indeed, while our main finding is ostensibly at odds with all the literature that reports evidence for a robust effect of language distance on bilingual cognition, it accurately reflects (i) the fact that studies that do find an effect pull to opposite directions (i.e. finding that bilingual effects exist in similar vs. distant language bilinguals) and (ii) the many studies that obtain null, mixed or spurious results which do not amount to reliable evidence for a distance effect in bilingualism (52-55). Interpreting our findings, we argue that the evidence we obtain in favor of a distance effect in bilingual cognition is modest for different methodological reasons. The first reason has to do with the databases we used as metrics. We found that the effect of PCM is significant in the reduced dataset that consists of language pairs that are measurable in all four metrics: when the L(anguage)1-L2 similarity in the morphosyntax of the nominal domain (which is what the PCM measures) increases, the likelihood of observing a bilingual advantage in these similar language bilinguals decreases. However, Grambank measures morphosyntactic similarity too (although not restricted to the nominal domain), and yet we did not find a significant association of this metric with the occurrence of bilingual advantages.

One explanation for this variation could be that very large databases such as Grambank cover many grammatical domains in a vast number of languages, and to achieve universal coverage they may need to sacrifice finer points of variation. To be more precise, for most language pairs, we relied on 195 morphosyntactic values when using Grambank and 94 values when using the PCM. This difference in detail is considerable if we take into account that the PCM covers only one domain of grammar (i.e. nominal morphosyntax), whereas Grambank is much broader, including not only the nominal domain, but also values on tense, aspect, mood, negation, auxiliary support, and even more discourse-oriented functions of certain grammatical markers such as politeness and inclusion. If some of these grammatical markers are aggregates that can be further split to capture finer points of variation (i.e. micro-cues; 56, 57), the level of detail (or the lack thereof) of the database used to measure distance will affect the results, possibly leading to sparse evidence, even if distance indeed modulates bilingual adaptations. When measuring the impact of language distance on the likelihood to observe any bilingual effect, we find no statistically robust effect of distance for any of the metrics when taken individually, but we do observe a significant positive effect for Grambank and a significant negative effect for SDML when both are considered at the same time. However, this model is on the verge of significance ($X^2(1) = 3.96$, $p = .046$) when compared to a model that only includes a Grambank effect (which is not significant by itself) and thus any conclusions drawn from these results are not based on firm evidence. Certain language pairs (e.g., English-Greek, Spanish-Basque) seem to stand out in terms of hosting bilingual adaptations, however the datapoints are so few that their role cannot be meaningfully interpreted. It is also possible that what we argue in terms of finer details of one metric over others highlights important considerations for determining what should be a preferred metric to use for distance calculations, depending on individual study specific questions, hypotheses, and argumentation for mechanistic links (e.g., what domains of language are predicted to be most relevant for particular adaptations to distinct areas of neurocognition, and why). If what we argued above is on the right track, the finer detail of a particular metric could meaningfully align with the predictions of individual studies targeting particular cognitive functions. For example, if (dis)similarity in the nominal domain can be linked to a specific monitoring or working memory task, then the PCM would be the best metric. Conversely, if a study can sustain from its outset a position that the nominal domain is not the right focus *a priori* for determining the relevant distance, then other metrics would obviously be preferred, potentially a more global one like Grambank. Of course, this underscores the need for cross-disciplinary collaboration to hone in on the most informed choices in terms of how to assess *a priori* language distance, that is prior to undertaking any given study.

Second, the adoption of datasets with universal coverage in terms of cross-linguistic differences is not sufficient when the aim is to decipher the full effect of distance. The reason is that these datasets need to be subsequently mapped to cognitive studies that provide evidence for bilingual adaptations, and these studies include predominantly WEIRD samples (figure 1). While specific language combinations are abundant in our study (e.g., English-Spanish, Catalan-Spanish, English-German, English-French, French-German), several other combinations are featured just once, and far too many that in principle could exist given their real world distributions are simply missing. In the absence of diversity in the language combinations that can elucidate the effect of distance, our understanding of it remains incomplete. In this context, the sparse evidence in favor of a distance effect may mostly be reflecting the sparse representation of languages that do not belong to the Indo-European family.

Third, this limitation is complemented by limitations in measuring distance. While we employed four different metrics, they all tap into either morphosyntax or the lexicon. Semantic similarities are left in the margins due to the absence of universal metrics, but undoubtedly, they play a role when it comes to determining the overall distance between two languages: while words with proximate meanings tend to cluster together cross-linguistically, languages diverge in how those clusters relate across the semantic space (58). Of course, there are many other domains that are also largely unrepresented, such as discourse pragmatics, that also likely bring significance to bear in the relevant sense.

While these explanations can help interpret our results, they are not the only possible ones. For instance, our approach does not delve into possible interactions of language distance with other factors of the bilingual experience, such as age of acquisition, degree of bilingual engagement, quality and quantities of input, proficiency/dominance, linguistic attitudes, linguistic landscapes, and degree of code-switching. It is likely that the modulatory effect of distance interacts with and possibly gets mitigated by such variables (28). Even if language distance does modulate bilingual adaptations, the answer to the Source Problem is unlikely to boil down to a single variable. Since the empirical evidence that we mapped to metrics of language distance includes very heterogeneous bilingual/multilingual groups of different sizes that come from different studies and cover different developmental trajectories, the sparse evidence we obtain could be the outcome of attempting to trace an effect at the *population* level. Put another way, while causality may exist at the level of the *individual* (i.e. language distance causing modulations to the degree of recruitment of cognitive control mechanisms), distance, as a moderator of bilingual adaptations, interacts with other variables of the bilingual experience, hence at the population level, we can only observe diluted statistical correlations that take the form of sparse evidence.

Moreover, there is no control across studies related to other aspects of life that contribute to individual differences in neurocognitive adaptations that could interact with whatever contribution bilingualism makes independently. A perennial issue of both sampling in individual studies and aggregating across studies for meta-analytic purposes, when a dynamic variable such as bilingualism can be moderated in combination with other potentially co-occurring factors for any given outcome, concerns the potential for missing a few trees in the forest, even when keeping one's eye firmly fixated on the forest itself. In the present case, we know other life-style factors such as (higher) socio-economic status and degree of educational and occupational attainment also contribute to neurocognitive development and cognitive adaptations. As these known quantities vary significantly across bilinguals, it should come as no surprise that such factors are relevant interactors with degree of bilingual engagement when predicting individual neurocognitive adaptations (59). As this cannot be teased out in the present analysis, not least as the vast majority of papers on bilingual cognition have historically not measured these other factors or report them, we leave this part of the discussion to inspire future work that considers linguistic distance with increased sophistication to also factor in these and other potential interactions.

RQ2 asks whether different databases correlate when used as metrics of language distance. This question is important because it sheds light on whether distance can be approached as composite measure or not. This directly relates to the Source Problem, which asks what makes bilinguals different. One challenge that research in cognitive science faces is the inability to define in an ecologically accepted way what counts as similar when we are referring to mental representations that pertain to language. Succinctly put, a language may be conceived as similar to another language, but there is no consensus about what this means and how it is to be measured. Is it the percentage of shared cognates what determines similarity or is it the phonemic inventory? Is it a matter of overlapping syntax or of organizing concepts in a similar way in the semantic space? Do all these factors matter to exactly equal degrees or does one take precedence over the others? Does the answer to this last question depend, perhaps, on the domain of cognition one is trying to make predictions for? From this perspective, understanding whether different metrics correlate when measuring language distance is important because it can enhance our understanding of why cross-linguistic variation looks the way it does, and whether there are functional trends and trade-offs that lead languages to self-organize in specific ways (e.g., semantically similar words tend to be phonologically similar too; 60).

Our analysis suggests that the answer to RQ2 is positive. We used two databases that chart morphosyntactic variation (PCM, Grambank) and two databases that tap into the lexicon (SDML, EZ Glot). Figure 3 shows medium to (very) high correlations across the four databases we used as metrics of cross-linguistic similarity. However, contrary to what would perhaps be deemed as the default expectation, the strongest correlation is not found between the two morphosyntactic or the two lexical databases, but between a lexical and a morphosyntactic one: Grambank and SDML. Since Grambank and PCM both chart variation in the nominal domain (although using different phrasing of the relevant features), one could expect to find the strongest correlation between these two datasets. Interestingly, the pair with the lowest correlation also involves a lexical and a morphosyntactic database: PCM and EZ Glot. These are the models with the smaller coverage in each domain: 94 morphosyntactic features for PCM (vs. 195 for Grambank) and roughly 1.5m words for EZ Glot (vs. 8m cognates for SDML). This finding suggests that cross-linguistic distance may be unevenly distributed across domains (lexicon, semantics, grammar, etc.), such that languages that share a large number of cognates may not necessarily show a great overlap in, for instance, nominal morphology. Therefore, the use of precise metrics across levels of linguistic analysis is necessary in order to establish an accurate calculation of L1-L2 similarity.

Returning to the Source Problem, one could ask whether our investigation of whether distance predicts the likelihood of observing bilingual adaptations can truly address this issue. Essentially, the question of what makes bilinguals different entails some level of comparison: Different to whom? The default answer has been monolinguals, typically understood as people that know or use —depending on the definition of bilingualism one endorses — only one language, although this answer is not unproblematic (61, 62). Thus, one could question whether language distance can provide an answer to the Source Problem, on the grounds that, by definition, distance entails at least two language systems, and monolinguals have only one. We argue that our approach is not simply relevant but uniquely suited to capture the bigger picture, by addressing one of the fundamental debates in bilingual cognition: Are bilingual adaptations a difference in *kind* or *degree*? A difference in kind entails the deployment of *different* neurocognitive resources/mechanisms (e.g., bilinguals benefit from the use of an inhibitory mechanism for selecting among competing alternatives, and this mechanism is qualitatively different from the one monolinguals use when choosing between synonyms), but a difference in degree entails stronger or weaker use of the *same* resources/mechanisms (41). While some scholars have argued in favor of qualitative differences across monolinguals and bilinguals, others have suggested that the differences are purely quantitative, thus supporting the degree interpretation (15). This is where our results become relevant: If language distance indeed modulates bilingual adaptations, the answer to the Source Problem can likely be framed as a matter of degree. We can think of bilingualism as a continuum, not only in the sense of dual language engagement, exposure, proficiency and the like (40, 63-65), but also related to degree of language similarity, with monolinguals being at one end. Extremely similar dialect speakers follow; these are speakers of languages with high mutual intelligibility and possibly 100% similarity in certain metrics (see Methods), who do not need to code-switch for effective communication, hence are very similar to monolinguals from a functional point of view. Somewhat similar language speakers come next, leading all the way to the other end of the continuum, which hosts speakers of languages that show (close to) zero overlap. To confirm this hypothesis, one would need to combine our distance metrics with targeted testing of carefully selected bilingual groups, manipulating the variables of language similarity and intelligibility accordingly. Overall, language distance can help answer the question of what makes (some) bilinguals different in a way that sheds light on the cognitive implications of the divide between languages and dialects.

Materials and Methods

To create a dataset of bilingual adaptations, we used two recent systematic reviews (66 and 67), with the former being an aggregate of two big meta-analyses (68 and 69) of bilingual effects. The dataset, the cross-linguistic similarities, and the code we used to run the analyses are available at https://osf.io/kwmrp/. The total number of articles included in our dataset is 286: 237 from Masullo et al. (67) and 49 from Grundy (66). The total number of studies reviewed in Masullo et al. (67) is 368, but 131 articles had to be removed for the purposes of our analysis for the following reasons: 127 studies included bilingual populations whose other language was not specified (i.e. they presented a bilingual population as speaking English or Spanish plus another language, without identifying the other language(s)), and four studies were removed because they included bilingual signers, and our metrics do not cover sign languages. Grundy (66) consists of 186 articles. For our analysis, 137 studies were removed: 90 studies were duplicates, already included in our sample through Masullo et al. (67), 29 studies included bilingual populations whose other language was not specified, 12 studies were not included because the full text was not available, four studies were removed because they were not written in English, and two studies were removed because they did not include a comparison between at least one monolingual and one bilingual group. Therefore, our final sample consists of 286 studies that test bilingual adaptations, and 56,122 participants, understood here as "testing units", as some studies ran more than one experiment without always specifying whether the same participants have been tested or not. For every study, we coded as binary variables the presence or absence of a finding that could be framed as a so-called bilingual advantage, the presence or absence of a bilingual disadvantage, or the absence of any effect (i.e. null result). The reason for looking separately into advantages vs. overall bilingual effects has to do with publication biases: It has been suggested that results that show positive bilingual adaptations are more likely to be published compared to null or negative results (70), although more recent estimates have shown that such claims may have been inflated by sampling biases (71). Still, to provide the most comprehensive picture possible, we analyzed our results from both perspectives, that is, first by separately testing the effect of distance in relation to the allegedly most plentiful category (i.e. positive adaptations) and then by repeating the analysis, merging all bilingual effects together. Depending on how many experiments were run, one study could give rise to more than one entry if it provided evidence for different results (e.g., an enhanced bilingual performance in naming in one experiment, combined with a null result in monitoring in another experiment). Thus, the 286 studies gave rise to 510 datapoints that cover 79 different language pairs, spanning 11 language families (i.e. Indo-European, Sino-Tibetan, Koreanic, Dravidian, Afroasiatic, Turkic, Austroasiatic, Japonic, Uralic, Tai-Kadai and Niger-Congo) and a language isolate (Basque).

All language pairs were measured using the following metrics: Grambank (45), PCM (42), EZ Glot (46), and SDML (47). Grambank charts morphosyntactic variation in 2,467 languages, in the form of 195 binary features that cover different domains of grammar. Manhattan Distances are used to calculate the sum total of the number of differences between the represented languages. In this context, 'language' is understood as a set of 195 features for which each language shows a value of either 0 (feature absent) or 1 (feature present). For example, if there are ten features, and languages A and B share the same values for all, this gives a Manhattan Distance of 0, meaning that A and B are identical, when their similarity is measured through Grambank. In a few cases, the value in some language is unclear, hence instead of 0 or 1, the feature is marked as unknown (?). To determine distance, value identicalness is necessary. This means that if a language has a feature marked with a value of 1, and another language has the same feature with anything but 1 (i.e. 0 or ?), this counts as a difference for the purposes of our analyses. In a small number of cases, Grambank allows for multiple realizations of a feature. For example, feature GB065 concerns the pragmatically unmarked order of the adnominal possessor noun and the possessed noun. If in a language the possessor nominal or independent possessor pronoun precedes the possessum, this is coded as 1. If it follows, it is coded as 2. If both options are possible, this is coded as 3. In this case, for calculating distance, the pairs of values 1-1, 2-2, 3-3, 3-1, and 3-2 count as identical, and only 1-2 counts as a difference.

The PCM measures Jaccard Distance in 94 morphosyntactic features from the nominal domain. It covers 58 modern languages, mostly from Eurasia. Its empirical coverage is thus more constrained compared to Grambank both in terms of languages but also in terms of morphosyntactic variation. However, its focus on one domain likely entails a more fine-grained representation of the actual variation space within this domain. Similarly to Grambank, languages in the PCM are sets of binary features, represented as – (feature absent) or + (feature present). For example, if there are ten features, a Jaccard Distance of 0 means that for all features the languages A and B have exactly the same values. The Jaccard Distance is calculated as $D(X,Y) = 1 – J(X,Y)$. $J(X,Y)$ denotes Jaccard Similarity, which is computed as the ratio of the length of the set intersection to the length of the set union.

EZ Glot measures lexical similarity, while also considering the number of words a language shares with other languages. In this sense, this metric provides a comparative perspective into the lexical similarity between the two languages in relation to lexical similarities in all other languages (46). It is represented as: $S(L1|L2) = |S|(L1|L2) * (N(L1) / D(L1))$ with $|S|$ = similarity, S = scaled similarity, N = Number of words shared with other languages, D = Number of words analyzed. A lexical similarity of 100 entails that there is a perfect overlap for all the words in D. This metric consists of 1.5 million contemporary dictionary words that come from 93 languages, mined from resources such as Wiktionary, OmegaWiki, FreeDict, or Apertium (47).

The SDML measures lexical similarity based on cognate words, using as input the CogNet database (72) that contains over 8 million cognate pairs. Similarity is measured by comparing a set of standardized wordlists and counting those forms that are similar in both form and meaning. The online database contains over 27 thousand language pairs, covering 331 languages. A similarity value and a confidence value are given for each language pair. Similarity is a value between 0 and 100. Confidence is ranked as low, medium, or high. The ranking depends on the sizes of the lexicons over which the similarity value was computed: the smaller the lexicon sizes, the lower the confidence (47).

While the three of the four metrics (i.e. PCM, EZ Glot, SDML) come with calculated distances in their respective domains and language pairs, to obtain reliable results we had to (i) make their measurements uniform and (ii) enhance their empirical coverage whenever possible. For example, with respect to the empirical coverage, we had to add values for Catalan in the PCM and calculate its similarity with other languages. Similarly, Spanish is missing from Grambank, so we had to include it. The reason is that Spanish and Catalan are very prominent in our sample of bilingual adaptations (see figure 2), and not having metrics for them would entail leaving in the margins a significant portion of our dataset. Another issue has to do with mixed groups. For instance, some studies test bilingual populations that are presented as speakers of Chinese. However, Chinese is not a language with its own code in Glottolog or in other databases of the world's languages; it is a set of languages. When calculating distance, such nuances matter because Mandarin, Cantonese, and other varieties of the Sino-Tibetan language family correspond to different sets of features. In such cases of uncertainty or mixed identity (e.g., mixed Mandarin and Cantonese speakers subsumed under the label 'Chinese'), we interpreted Chinese as Mandarin, thus analyzing it as the set of features that correspond to the Glottolog/Grambank code mand1415.

With respect to making the measurements uniform, two metrics (EZ Glot, SDML) report similarity while PCM reports distance. We thus translated Jaccard Distance in PCM to Jaccard Similarity. The fourth metric, Grambank, represents languages as sets of features, but does not involve calculated distances or similarities between these sets at its current state of development. To calculate similarity, we used the database 'Values' (45, Supplementary Material). We identified the languages that are featured in our dataset and calculated Jaccard Similarity on an individual pair basis. For the purposes of our re-coding of Grambank values, when a language pair involved an unknown status (?), we treated this comparison as null. When computing similarity, binary strings of features marked as absent in both languages were also removed, following previous practice with redundant parameters (42). Table 1 provides a summary of our coding. Jaccard Similarity was calculated for each language pair using the following formula $J(A, B) = |A \cap B| / |A \cup B|$.

Language A	Language B	Coding
0	1	Difference
1	0	Difference
1	1	Similarity
0	0	Null/Removed
?	?	Unknown status /Removed

?	0	Unknown status /Removed

Table 1: Coding Grambank values

Last, we removed from our dataset one language pair that showed 0% distance. Greek and Cypriot Greek have a Jaccard Distance of 0 according to the PCM. This is the only metric that includes Cypriot Greek. If this pair was included in the analysis, the two sets of features would be indistinguishable due to the very limited coverage of Cypriot Greek. This high similarity brings along mutual intelligibility, and in certain cases, it entails populations that are functionally like monolinguals, as they do not switch between different varieties, but instead incorporate elements from them into one fused lect (73).

Acknowledgments

EL acknowledges funding by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033), project number PID2021-124399NA-I00. FG acknowledges funding by the German Research Foundation (DFG), Emmy-Noether grant "What's in a name?" (project number 459717703). MW and JR acknowledge funding by the Trond Mohn Foundation, grant number TMS2023UiT01.

Competing Interests Statement

The authors declare that they have no competing interests.

Data Availability Statement

All data needed to evaluate the conclusions in the paper are present in the paper and the following repository https://osf.io/kwmrp/

References

[1]. E. Bialystok, Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *Int. J. Biling. Educ. Biling*. **10**(3), 210–223 (2007).

[2]. E. Bialystok, The bilingual adaptation: How minds accommodate experience. *Psychol. Bull*. **143**(3), 233 (2017).

[3]. E. Bialystok, Bilingualism modifies cognition through adaptation, not transfer. *TiCS* (2024).

[4]. A. Costa, M. Hernández, N. Sebastián-Gallés, Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*. **106**, 59–86 (2008).

[5]. V. DeLuca, J. Rothman, E. Bialystok, C. Pliatsikas, Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *PNAS*, **116**(15), 7565-7574 (2019).

[6]. V. DeLuca, J. Rothman, E. Bialystok, C. Pliatsikas, Duration and extent of bilingual experience modulate neurocognitive outcomes. *Neuroimage*. **204**, 116222 (2020).

[7]. K. R. Paap, H. A. Jonhson, O. Sawi, Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*. **69**, 265-278 (2015).

[8]. K. R. Paap, L. Mason, B. Zimiga, Y. Ayala-Silva, M. Frost, The alchemy of confirmation bias transmutes expectations into bilingual advantages: A tale of two new meta-analyses. *Q. J. Exp. Psychol.* **73**, 1290-1299 (2020).

[9]. K. R. Paap, Z. I. Greenberg, There is no coherent evidence for a bilingual advantage in executive processing. *Cogn. Psychol.* **66**(2), 232-258 (2013).

[10]. J. A. Duñabeitia, J. A. Hernández, E. Antón, P. Macizo, A. Estévez, L. J. Fuentes, M. Carreiras, The inhibitory advantage in bilingual children revisited: Myth or reality? *Exp. Psychol.* **61**, 234-251 (2014).

[11]. M. Lehtonen, V. Fyndanis, J. Jylkkä, The relationship between bilingual language use and executive functions. *Nat. Rev. Psychol*. **2**, 360–373 (2023).

[12]. E. Leivada, V. Dentella, C. Masullo, J. Rothman, On trade-offs in bilingualism and moving beyond the stacking the deck fallacy. *Bilingualism.* **26**(3), 550–555 (2022).

[13]. N. Planckaert, W. Duyck, E. Woumans, Is there a cognitive advantage in inhibition and switching for bilingual children? A systematic review. *Front. Psychol.* **14**, 1191816 (2023).

[14]. R. J. Hartsuiker, Why it is pointless to ask under which specific circumstances the bilingual advantage occurs. *Cortex.* **73**, 336–337 (2015).

[15]. E. Blanco-Elorrieta, A. Caramazza, On the need for theoretically guided appromaches to possible bilingual advantages: An evaluation of the potential loci in the language and executive control systems. *Neurobiol. Lang*. **2**(4), 452–463 (2021).

[16]. E. Bialystok, F. I. M. Craik, How does bilingualism modify cognitive function? Attention to the mechanism. *Psychon. Bull. Rev*. **29**, 1246–1269 (2022).

[17]. B. Treccani, C. Mulatti, No matter who, no matter how… and no matter whether the white matter matters. Why theories of bilingual advantage in executive functioning are so difficult to falsify. *Cortex*. **73**, 349-351 (2015).

[18]. A. de Bruin, A. S. Dick, M. Carreiras, Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiol. Lang*. **2**(4), 433-445 (2021).

[19]. T. Dijkstra, J. Grainger, W. J. B. van Heuven, Recognition of Cognates and Interlingual homographs: The neglected role of phonology. *J. Mem. Lang.* **41**, 496–518 (1999).

[20]. A. Costa, M. Santesteban, I. Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *J. Exp. Psychol.: Learn. Mem. Cogn*. **32**(5), 1057–1074 (2006).

[21]. A. Rodriguez-Fornells, R. de Diego Balaguer, T. F. Münte, Executive Control in Bilingual Language Processing. *Lang. Learn*. **56**, 133–190 (2006).

[22]. J. Rothman, J. G. Alonso, E. Puig-Mayenco, *Third Language Acquisition and Linguistic Transfer* (Cambridge University Press, Cambridge, 2019).

[23]. N. Radman, L. Jost, S. Dorood, C. Mancini, J. M. Annoni, Language distance modulates cognitive control in bilinguals. *Sci. Rep.* **11**, 24131 (2021).
[24]. T. Dijkstra, W. J. B. van Heuven, The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism.* **5**(3), 175–197 (2002).
[25]. A. Casaponsa, J. A. Duñabeitia, Lexical organization of language-ambiguous and language-specific words in bilinguals. *Q. J. Exp. Psychol*. **69**, 589–604 (2016).
[26]. A. Zawiszewski, I. Laka, Bilinguals processing noun morphology: evidence for the language distance hypothesis from event-related potentials. *J. Neurolinguist.* **55**, 100908 (2020).
[27]. A. Shook, V. Marian, The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism*. **16**, 304-324 (2013).
[28]. F. Gallo, A. Myachykov, M. Nelyubina, Y. Shtyrov, J. Kubiak, L. Terekhina, J. Abutalebi, Linguistic distance dynamically modulates the effects of bilingualism on executive performance in aging. *Bilingualism*. 1–10 (2023).
[29]. M. T. Carthery-Goulart, A. J. Privitera, B. S. Weekes, Does Language Distance Modulate the Contribution of Bilingualism to Cognitive Reserve in Seniors? A Systematic Review. *Am. J. Alzheimers. Dis. Other. Demen*. **38**, 15333175231167223 (2023).
[30]. E. L. Coderre, W. J. van Heuven, The effect of script similarity on executive control in bilinguals. *Front. Psychol*. **5**, 1070 (2014).
[31]. C. Morrison, V. Taler, ERP differences between monolinguals and bilinguals: The role of linguistic distance. *Bilingualism.* **26**(2), 293-306 (2023).
[32]. S. Malik-Moraleda, O. Jouravlev, M. Taliaferro, Z. Mineroff, T. Cucu, K. Mahowald, I. A. Blank, E. Fedorenko, Functional characterization of the language network of polyglots and hyperpolyglots with precision fMRI. *Cereb. Cortex*. **34**, bhae049 (2024).
[33]. R. G. Bosworth, E. M. Binder, S. C. Tyler, J. P. Morford, Automaticity of lexical access in deaf and hearing bilinguals: Cross-linguistic evidence from the color Stroop task across five languages. *Cognition*. **212**, 104659 (2021).
[34]. A. Lu, S. Liu, J. Zhang, M. Zhang, T. Song, L. Wang, X. Wang, The effect of phonetic similarity on domain-general executive control in color-shape task: Evidence from Cantonese-Mandarin and Beijing-Dialect-Mandarin bidialectals. *J. Psycholinguist. Res*. **52**(5), 1855-1874 (2023).
[35]. A. Perovic, D. Filipović Đurđević, S. Halupka-Rešetar, The effect of bilingualism on executive functions when languages are similar: A comparison between Hungarian–Serbian and Slovak–Serbian young adult bilinguals. *Mem. Cogn.* **51**(3), 561–581 (2023).
[36]. J. Henrich, S. J. Heine, A. Norenzayan. The weirdest people in the world? *Behav. Brain. Sci*. **33**, 61-83 (2010).
[37]. E. Adamou, *The Adaptive Bilingual Mind. Insights from Endangered Languages* (Cambridge University Press, Cambridge, 2021).
[38]. D. Blasi, J. Henrich, E. Adamou, D. Kemmerer, A. Majid, Over-reliance on English hinders cognitive science. *TicS.* **26**(12), 1153–1170 (2022).

[39]. E. Leivada, I. Rodríguez-Ordóñez, M. C. Parafita-Couto, S. Perpiñán, Bilingualism with minority languages: Why searching for unicorn language users does not move us forward. *Appl. Psycholinguist*. **44**(3), 384-399 (2023).
[40]. G. Luk, Justice and equity for whom? Reframing research on the "bilingual (dis) advantage". *App. Psycholinguist*. **44**(3), 301-315 (2023).
[41]. E. Eden, "Measuring Phonological Distance between Languages", Doctoral dissertation, University College London, UK (2018).
[42]. A. Ceolin, C. Guardiano, G. Longobardi, M. A. Irimia, L. Bortolussi, A. Sgarro, At the boundaries of syntactic prehistory. *Philos. Trans. R. Soc. B*. **376**, 20200197 (2021).
[43]. Y. -Y. Lee, A conceptual analysis of typological distance and its potential consequences on the bilingual brain. *Int. J. Biling. Educ. Biling.* **9**, 3333-3346 (2022).
[44]. E. Leivada, M. Westergaard, J. A. Duñabeitia, J. Rothman, On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed?. *Bilingualism*. **24**, 197-210 (2021).
[45]. H. Skirgård, H. J. Haynie, H. Hammarström, D. E. Blasi, J. Collins, J. Latarche, J. Lesage, T. Weber, A. Witzlack-Makarevich, M. Dunn, G. Reesink, R. Singer, C. Bowern, P. Epps, J. Hill, O. Vesakoski, N. Karolin Abbas, S. Ananth, D. Auer, N. A. Bakker, G. Barbos, A. Bolls, R. D. Borges, M. Browen, L. Chevallier, S. Danielsen, S. Dohlen, L. Dorenbusch, E. Dorn, M. Duhamel, F. El Haj Ali, J. Elliott, G. Falcone, A. -M. Fehn, J. Fischer, Y. Ghanggo Ate, H. Gibson, H -P. Göbel, J. A. Goodall, V. Gruner, A. Harvey, R. Hayes, L. Heer, R. E. Herrera Miranda, N. Hübler, B. H. Huntington-Rainey, G. Inglese, J. K. Ivani, M. Johns, E. Just, I. Kapitonov, E. Kashima, C. Kipf, J. V. Klingenberg, N. König, A. Koti, R. G. A. Kowalik, O. Krasnoukhova, K. Lynn Lindsey, N. L. M. Lindvall, M. Lorenzen, H. Lutzenberger, A. Marley, T. R. A. Martins, C. Mata German, S. van der Meer, J. Montoya, M. Müller, S. Muradoglu, H. Gatherer, D. Nash, K. Neely, J. Nickel, M. Norvik, B. Olsson, C. Akinyi Oluoch, D. Osgarby, J. Peacock, I. O.C. Pearey, N. Peck, J. Peter, S. Petit, S. Pieper, M. Poblete, D. Prestipino, L. Raabe, A. Raja, J. Reimringer, S. C. Rey, J. Rizaew, E. Ruppert, K. K. Salmon, J. Sammet, R. Schembri, L. Schlabbach, F. W. P. Schmidt, D. Schokkin, J. Siegel, A. Skilton, H. de Sousa. K. Sverredal, D. Valle, J. Vera, J. Voß, D. Wikalier Smith, T. Witte, H. Wu, S. Yam, J. Ye 葉婧婷, M. Yong, T. Yuditha, R. Zariquiey, R. Forkel, N. Evans, S. C. Levinson, M. Haspelmath, S. J. Greenhill, Q. D. Atkinson, R. D. Gray, Grambank (v1.0) [Data set]. Zenodo (2023) https://doi.org/10.5281/zenodo.7740140
[46]. L. Kovacevic, V. Bradic, G. de Melo, S. Zdravkovic, Ezglot [Data set] (2021), https://www.ezglot.com/
[47]. G. Bella, K. Batsuren, F. Giunchiglia, "A Database and Visualization of the Similarity of Contemporary Lexicons". In *Text, Speech, and Dialogue. TSD 2021. Lecture Notes in Computer Science*, K. Ekštein, F. Pártl, M. Konopík, Eds. (Springer, Cham, 2021).

[48]. B. M Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, J. S. S. White, Generalized linear mixed models: a practical guide for ecology and evolution. *Trends. Ecol. Evol*. **24**(3), 127-135 (2009).
[49]. D. Bates, M. Maechler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft*. **67**(1), 1-48 (2015).
[50]. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Soft*. **82**(13), 1-26 (2017).
[51]. A. Olguin, M. Cekic, T. A. Bekinschtein, N. Katsos, M. Bozic, Bilingualism and language similarity modify the neural mechanisms of selective attention. *Sci. Rep.* **9**(1), 8204 (2019).
[52]. J. Oschwald, A. Schättin, C. C. von Bastian, A. S. Souza, Bidialectalism and Bilingualism: Exploring the Role of Language Similarity as a Link Between Linguistic Ability and Executive Control. *Front. Psychol*. **9**, 1997 (2018).
[53]. D. E. Sörman, P. Hansson, J. K. Ljungberg, J. K. (2019). Different Features of Bilingualism in Relation to Executive Functioning. *Front. Psychol*. **10**, 269 (2019).
[54]. Laketa, A., Studenica, A., Chrysochoou, E., Blakey, E., & Vivas, A. B. (2021). Biculturalism, linguistic distance, and bilingual profile effects on the bilingual influence on cognition: A comprehensive multipopulation approach. Journal of experimental psychology. General, 150(11), 2273–2292.
[55]. N. Alrwaita, C. Houston-Price, L. Meteyard, T. Voits, C. Pliatsikas, Executive functions are modulated by the context of dual language use: diglossic, bilingual and monolingual older adults. *Bilingualism*. **27**(1), 178–203 (2024).
[56]. M. Westergaard, *The Acquisition of Word Order: Micro-cues, Information Structure and Economy* (John Benjamins, Amsterdam, 2009).
[57]. M. Westergaard, Linguistic variation and micro-cues in first language acquisition. *Linguist. Var*. **14**, 26-45 (2014).
[58]. M. Lewis, A. Cahill, N. Madnani, J. Evans, Local similarity and global variability characterize the semantic space of human languages. *PNAS*, **120**(51), e2300986120 (2023).
[59]. J. Rothman, Harnessing the bilingual descent down the mountain of life: Charting novel paths for Cognitive and Brain Reserves research. *Bilingualism*. 1-9 (2024).
[60]. I. Dautriche, K. Mahowald, E. Gibson, S. T. Piantadosi, Wordform Similarity Increases with Semantic Similarity: An Analysis of 100 Languages. *Cog. Sci*. **41**(8), 2149-2169 (2017).
[61]. J. Rothman, F. Bayram, V. DeLuca, G. Di Pisa, J. A. Duñabeitia, K. Gharibi, J. Hao, N. Kolb, M. Kubota, T. Kupisch, T. Laméris, A. Luque, B. van Osch, S. M. Pereira Soares, D. Tat, A. Tomić, T. Voits, S. Wulff, Monolingual comparative normativity in bilingualism research is out of "control": Arguments and alternatives. *App. Psycholinguist.* **44**(3), 316-329 (2023a).
[62]. A. De Houwer, danger of bilingual–monolingual comparisons in applied psycholinguistic research. *App. Psycholinguist.* **44**(3), 343-357 (2023).

[63]. G. Luk, E. Bialystok, Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *J. Cogn. Psychol*. **25**(5), 605-621 (2013).

[64]. J. Rothman, F. Bayram, V. DeLuca, J. G. Alonso, M. Kubota, E. Puig-Mayenco, Defining bilingualism as a continuum. In *Understanding Language and Cognition through Bilingualism: In honor of Ellen Bialystok*, G. Luk, J. A. E. Anderson, J. G. Grundy, Eds. (John Benjamins, Amsterdam, 2023b).

[65]. L. V. Kremin, K. Byers-Heinlein, Why not both? Rethinking categorical and continuous approaches to bilingualism. *Int. J. Biling*. **25**(6), 1560-1575 (2021).

[66]. J. G. Grundy, JThe effects of bilingualism on executive functions: An updated quantitative analysis. *J. Cult. Cogn. Sci*. **4**(2), 177–199 (2020).

[67]. C. Masullo, V. Dentella, E. Leivada, 73% of the observed bilingual (dis)advantageous effects on cognition stem from sociolinguistic factors: A systematic review. *Bilingualism*. **27**(3), 480–494 (2023).

[68]. M. Lehtonen, A. Soveri, A. Laine, J. Järvenpää, A. de Bruin, J. Antfolk, Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychol. Bull*. **144**(4), 394–425 (2018).

[69]. S. Donnelly, P. J. Brooks, B. D. Homer, Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychon. Bull. Rev*. **26**(4), 1122–1147 (2019).

[70]. A. de Bruin, B. Treccani, S. Della Sala, Cognitive advantage in bilingualism: An example of publication bias? *Psychol. Sci*. **26**, 99–107 (2015).

[71]. E. Leivada, A classification bias and an exclusion bias jointly overinflated the estimation of publication biases in bilingualism research. *Behav. Sci*. **13**, 812 (2023).

[72]. K. Batsuren, G. Bella, F. Giunchiglia, F. (2019). CogNet: A large-scale cognate database. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, L. Màrquez, Eds. (Association for Computational Linguistics, Florence, 2019), pp. 3136–3145.

[73]. P. Auer, From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *Int. J. Biling*. **3**(4), 309-332 (1999).