nature human behaviour

Registered Report https://doi.org/10.1038/s41562-025-02254-x

Measuring the semantic priming effect across many languages

Received: 12 December 2021

Accepted: 27 May 2025

Published online: 24 September 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Semantic priming has been studied for nearly 50 years across various experimental manipulations and theoretical frameworks. Although previous studies provide insight into the cognitive underpinnings of semantic representations, they have suffered from small sample sizes and a lack of linguistic and cultural diversity. In this Registered Report, we measured the size and the variability of the semantic priming effect across 19 languages (n = 25,163 participants analysed) by creating the largest available database of semantic priming values using an adaptive sampling procedure. We found evidence for semantic priming in terms of differences in response latencies between related word-pair conditions and unrelated word-pair conditions. Model comparisons showed that the inclusion of a random intercept for language improved model fit, providing support for variability in semantic priming across languages. This study highlights the robustness and variability of semantic priming across languages and provides a rich, linguistically diverse dataset for further analysis. The Stage 1 protocol for this Registered Report was accepted in principle on 15 July 2022. The protocol, as accepted by the journal, can be found at https://osf.io/u5bp6 (registration) or https://osf.io/q4fjy (preprint version 6, 31 May 2022).

Semantic priming is a well-studied cognitive phenomenon whereby participants are shown a cue word (for example, DOG) followed by either a semantically related (for example, CAT) or unrelated (for example, BUS) target word[1]. Semantic priming is defined as the decrease in response latency (that is, reduced linguistic processing or facilitation) for a single target word that is semantically related to the cue word in comparison to an unrelated cue word[1]. Semantic priming research spans nearly 50 years of study as a tool to investigate cognitive processes, such as word recognition, and to elucidate the structure and organization of knowledge representation[2], often by using results from these studies to develop theoretical and computational models that capture empirical effects[3–6]. Priming has also been used in studies on attention[7,8], bi/multilingual people[9,10] and neurodivergent individuals such as those affected by Parkinson's disease, aphasia or schizophrenia, as well as in a large body of neuroscience studies[11–13]. The purpose of this study is to leverage the power and network of the Psychological Science Accelerator (PSA)[14] to create a cross-linguistic normed dataset of semantic priming, paired with other useful psycholinguistic variables (for example, frequency, familiarity and concreteness). The PSA is a large network of research laboratories committed to large-scale data collection and open-scholarship principles.

Experimental psychologists have long understood that the stimuli in research studies are of great importance, and that controlled sets of normed information hold notable value for study control and allow for precision in measurement of effects. Often, stimuli are created in small pilot studies and then reused in many subsequent projects. However, both Lucas[15] and Hutchison[16] provided evidence that these small pilot data should be carefully interpreted given larger, more reliable datasets. In recent years, researchers have begun to more frequently publish large datasets with experimental stimuli for reuse in future work[17]. These datasets include lexical frequency[18,19], large collections of text (for example, corpora)[20], response latencies[21–23] and subjective ratings from participants on semantic dimensions such as emotion[24–26], concreteness[27] or familiarity[28]. Recent advances in computational capability, the growth of large-scale online data collection and the focus on replication and reproducibility may advance this research area.

e-mail: ebuchanan@harrisburgu.edu

Table 1 | Preregistered design table

Question	Hypothesis	Sampling plan	Analysis plan	Interpretation given to different outcomes
Is semantic priming a non-zero effect?	H_A: Response latencies will be faster for related word-pairs than for unrelated word pairs. H_0: Response latencies for related word-pairs will be slower than or equal to those for unrelated word-pairs.	We sampled participants on items until they reached a desired accuracy in parameter estimation CI width (s.e. = 0.09).	We calculated the mean and 95% CI for the priming effect subtracting related word conditions from unrelated word conditions at the item level by using an intercept-only regression model. These calculations were repeated for the data with 2.5 *Z*-score outlier trials excluded and with 3.0 *Z*-score outlier trials excluded.	The results support H_A when the lower limit of the CI is positive and non-zero, >0.0001. The results are inconclusive when the lower limit of the CI is negative or zero, ≤0.0001.
Does the semantic priming effect vary across languages?	H_A: Priming response latencies will be variable between languages (that is, heterogeneous). H_0: Priming response latencies will not be variable between languages (that is, homogenous).	We sampled participants on items until they reached a desired accuracy in parameter estimation CI width (s.e. = 0.09).	We added a random intercept of language to the previous intercept-only model to assess overall heterogeneity. These calculations were repeated for the data with 2.5 *Z*-score outlier trials excluded and with 3.0 *Z*-score outlier trials excluded.	The results support H_A when the ΔAIC (intercept-only minus random-intercept) is ≥2 points. The results are inconclusive when the ΔAIC (intercept-only minus random-intercept) is <2 points.

The importance of normed stimuli for research cannot be overstated. Not only do they provide methodological standardization for studies using the stimuli, but the stimuli themselves can also be studied to gain insight into cognitive architecture and processes, such as attention, memory, perception and language comprehension or production.

Normed datasets provide a wealth of information for studies on semantic priming. Facilitation in priming is based chiefly on semantic similarity or the related word-pair condition as contrasted to the unrelated word-pair condition. Traditionally, word pairs were simply grouped into pairs that were face-value similar (for example, DOG–CAT) and unrelated (for example, BUS–CAT), which was determined through pilot studies where word pairs provided the expected statistical results. However, for reproducibility and methodological control, semantic similarity values should be defined before the results are known[29]. Semantic similarity has various conceptual and computational definitions that all generally describe the shared meaning between two words or texts[5]. The most common forms of similarity are feature-based similarity (that is, the number of shared features between words)[30–32], association strength (that is, the probability of one word eliciting a second word when participants are simply shown the first word)[33,34] and text co-occurrence (that is, the words are similar because they frequently appear in similar contexts)[35–37]. Each of these computational definitions of similarity can be calculated from normed datasets or text corpora to provide a continuous measure of similarity from 0 (unrelated) to 1 (perfectly related).

The Semantic Priming Project comprised both a large-scale database collection and a semantic priming study that used defined stimuli to create related word pairs[21]. This project provided data for lexical decision and naming tasks for 1,661 English words and non-words, along with other psycholinguistic measures for future research. The results of the Semantic Priming Project showed 23-ms to 25-ms decreases in word response latencies (that is, lexical decision or naming speed) for the related word-pair conditions compared with unrelated word-pair conditions. Our study seeks to expand this dataset and address three key limitations of the Semantic Priming Project: reliability of item-level effects, small sample sizes per item, and the focus on English words and English-speaking participants.

First, Heyman et al.[38] explored the split-half reliability of item-level priming effects from the Semantic Priming Project, finding low reliability for the effects. This result corresponds with a study by Hutchison et al.[39] showing low reliability for priming effects; however, they demonstrated that priming effects can still be predicted at the item level, albeit with a smaller dataset. Relatedly, for the second limitation, Heyman et al.[40] noted that the required sample size necessary for reliable priming effects was much larger than the sample size used in the study, potentially explaining the differences between results as well as demonstrating the need for a larger dataset.

Last, the Semantic Priming Project contains only English data. If semantic priming provides a window into the structure of knowledge, the dominant focus on specific languages, such as English, has limited our understanding of the influence of linguistic variation on representation. Languages differ in script, syllables, morphology and semantics, as well as the cultural variations that occur across language users. Related concepts that one may consider universal, such as LEFT and RIGHT, are not coded into all languages. Studies with more than one language within the same study often focus on bi/multilingual individuals to elucidate the potential shared structure of knowledge across languages[41,42]. Therefore, claims about human language are often based on a small set of languages, limiting the generalizability of these claims[43]. Even with the increase in publication of normed datasets in non-English languages[17], conducting cross-linguistic studies on the same concepts is difficult, as large-scale data in this area are sparse.

Although it is challenging, newer computational techniques[44,45] and recently published corpora[20,46] enable the collection of a broader-coverage dataset in up to 43 languages. This study therefore aims to provide data that complement and extend the published data, which will encourage research on methodology, item characteristics, models, cross-linguistic consistency in priming and other theoretical areas that semantic priming has been applied to previously. The data address the proposed limitations by increasing sample size to improve reliability and expanding beyond the English language within the same target stimuli. From these openly shared data, two research questions are assessed, as detailed in Table 1:

(1) Is semantic priming a non-zero effect? To assess this research question, we examined the confidence interval (CI) of the semantic priming effect to determine if the lower limit of the CI is greater than zero using an intercept-only regression model estimating across all languages. We therefore predicted semantic facilitation with lower response latencies for related word-pair conditions than for unrelated word-pair conditions.

(2) Does the semantic priming effect vary across languages when examining the same target stimuli? We added a random intercept of language to the model estimated in Hypothesis 1 to estimate the variability of priming across languages. We concluded there is variability in priming effects across languages when the Akaike information criterion (AIC) for the random-intercept model is two or more points less than the AIC for the model in Hypothesis 1 (ref. 47). To contextualize these results, we provide a forest plot of the priming effects for languages to demonstrate the pattern of variability. For Hypothesis 2, we did not specify predicted directions for the effects but did expect potential variability in priming effects across languages. It is logical to expect

https://doi.org/10.1038/s41562-025-02254-x

differences in language due to culture, orthography, alphabet and so on, and empirical data suggest meaningful differences between languages[48,49].

This research crucially supplements the literature outlined above by focusing on several key components of psycholinguistic research. For sampling, we used accuracy in parameter estimation to ensure precision in our estimates[50,51] to address the known reliability issues in item-level responding[38,40] to support Hypothesis 1. The items were selected using new computational techniques for addressing semantic similarity[44,45] with recently available large corpora of movie subtitles[20] to appropriately match comparable items across languages. As noted in Buchanan et al.[17], research in non-English languages is expanding; however, stimulus matching is still sparse across published databases. By using large corpora, items are matched not only in their similarity levels but also for their frequency of use. Differences in priming can thus be attributed to differences in linguistic structure or culture rather than translation or poor item matching, supporting Hypothesis 2.

Results

In this section, we detail all languages included in the data collection and identify the languages that reached the preregistered minimum sample size. Next, we discuss the research labs and ethics involved in the project. We then detail the exclusion criteria from the preregistered plan, followed by the number of participants included in the available data. Descriptive statistics of the data are provided for participants, trials, items and priming. The final section covers the hypothesis testing from Table 1. To reduce redundancy, we provide an overview of the descriptive results and all preregistered descriptives in the Supplementary Information.

Languages

We originally identified 43 languages for possible data collection on the basis of the information available from the OpenSubtitles[20] and subs2vec[46] projects. We translated stimuli and collected data from at least one participant in the following 30 languages/dialects (languages with asterisks were included in our preregistered minimum data collection plan): Arabic, Brazilian Portuguese, Czech*, Danish, Dutch, English*, Farsi, French, German*, Greek, Hebrew, Hindi, Hungarian, Italian, Japanese*, Korean*, Norwegian, Polish, Portuguese (European)*, Romanian, Russian*, Serbian, Simplified Chinese*, Slovak, Slovenian, Spanish*, Thai, Traditional Chinese, Turkish* and Urdu. Table 2 provides a summary of the data collection for each language, including the number of included participants (based on the preregistered data inclusion rules), the number of participants excluded, the proportion of correct answers for the included participants (that is, participant accuracy scores were calculated, and then the average of participant accuracy scores for each language was calculated) and the median completion time for the included participants in minutes (https://osf.io/bqpk2). A complete breakdown of gender, education, age and stimulus completion can be found in the Supplementary Information (https://osf.io/y3dk7). The following 19 languages met the minimum data collection requirements and are analysed in this paper: Brazilian Portuguese, Czech, Danish, German, Greek, English, French, Hungarian, Italian, Japanese, Korean, Polish, Portuguese (European), Romanian, Russian, Serbian, Simplified Chinese, Spanish and Turkish. The stimuli for European and Brazilian Portuguese overlapped by 90%; the data were combined such that each unique target (unrelated and related trials) obtained the minimum number of participant answers. We present the combined results when discussing trials or global information but separate them when examining item- or priming-level effects. All data are available online, including those languages that did not meet the preregistered minimum data collection criteria for analysis (https://github.com/SemanticPriming/SPAML/tree/v1.0.2). For each language, we also provide data checks and a summary of the number of participants, trials, items and priming trials during data processing (summary: https://osf.io/zye59; 05_Data includes all processing files).

Table 2 | Language data collection sample sizes, accuracy and median study completion time in minutes

Language	*n* included	*n* excluded	Proportion correct	Median time (minutes)
Arabic	133	102	0.92	18.67
Czech	1,074	362	0.94	19.76
Danish	829	167	0.93	18.70
Dutch	184	25	0.93	17.60
English	5,122	1,607	0.92	17.64
Farsi	192	110	0.95	17.71
French	869	142	0.95	17.68
German	2,628	469	0.94	19.02
Greek	689	130	0.94	18.48
Hebrew	247	74	0.92	16.63
Hindi	1	2	0.82	27.39
Hungarian	718	180	0.94	17.94
Italian	1,085	142	0.95	18.10
Japanese	1,165	680	0.94	18.69
Korean	975	601	0.91	17.59
Norwegian	85	17	0.93	20.08
Polish	1,188	318	0.94	19.15
Portuguese (combined)	1,178	332	0.93	18.25
Romanian	741	174	0.94	19.65
Russian	1,806	956	0.94	19.68
Serbian	681	109	0.94	21.01
Simplified Chinese	729	291	0.93	17.75
Slovak	381	391	0.94	18.68
Slovenian	31	10	0.95	18.89
Spanish	1,468	284	0.94	18.04
Thai	65	20	0.95	18.34
Traditional Chinese	174	67	0.92	18.05
Turkish	2,218	790	0.93	17.83
Urdu	315	381	0.88	22.15

Ethics and research labs

A total of 133 labs completed ethics documentation for data collection, and 126 labs in 41 geopolitical regions collected data for the study. Each of the final data collection labs obtained local ethical review (81), relied on the ethical review provided by Harrisburg University (31) or provided evidence that no ethical review was required (14). The Supplementary Information provides links to the institutional review board approvals hosted on the Open Science Framework (OSF; https://osf.io/ycn7z/) and a table of participating labs with their data collection information, which includes languages sampled, geopolitical region of the team, compensation procedure and amount, online versus in-person testing, and testing type (individual participants or classroom-type settings; https://osf.io/ty4hp). This information can be matched to study data using the lab code that is present in the participant and trial-level files. See Fig. 1 for a visualization of the entire sample during data collection.

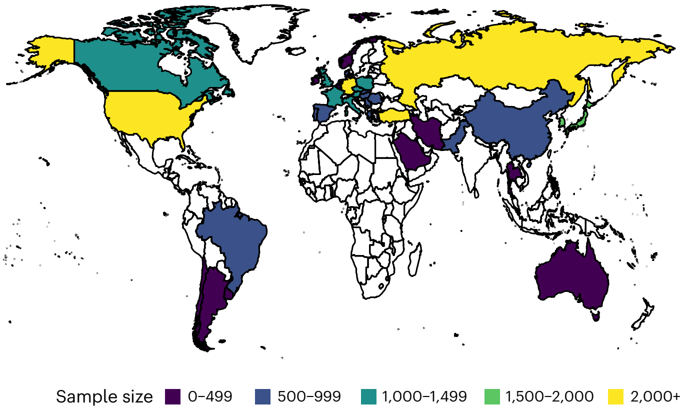

Fig. 1 | Sample sizes by region and language. Binned sample sizes based on research lab geopolitical region (or data collection language) demonstrating the full data available for reuse from the project.

Exclusion summary

Data were excluded for the following reasons in this order (per the preregistered plan):

(1) Participant-level data: all the participant's data were removed from the analyses if:
   a. The participant did not indicate at least 18 years of age.
   b. The participant did not complete at least 100 trials.
   c. The participant did not achieve 80% correct.

(2) Trial-level data: individual trials were removed from the analyses in the following instances:
   a. Timeout trials (that is, no response given in 3-s window). This value was chosen to ensure that the experiment was completed in under 30 min on average, while giving an appropriate amount of time in a lexical decision study to answer (using the Semantic Priming Project as a rubric for general trial length).
   b. Incorrectly answered trials.
   c. Response latencies shorter than 160 ms[52].

(3) Trial-level exclusions dependent on test: Participant sessions were *Z*-scored as described below, and trials were marked for exclusion in the dataset. Each analysis was tested with the full data and then without these values:
   a. Response latencies over the absolute value of $Z = 2.5$.
   b. Response latencies over the absolute value of $Z = 3.0$.

Participants

In this section, we describe both the full sample available for download and the analysed dataset. A total of 35,904 participants opened the study link, and 31,645 participants proceeded to complete at least one study trial (that is, past the practice trials). Of these participants, 26,971 were retained for analysis because they met our three participant-level inclusion criteria. The preregistered plan calculated accuracy as $\frac{n_{\text{Correct}}}{n_{\text{TrialsSeen}}}$ in the planned scripts; however, an administrative team discussion revealed that the preregistered report's definition of accuracy could alternatively be interpreted as $\frac{n_{\text{Correct}}}{n_{\text{Answered}}}$. If accuracy were defined using this alternative formula, 28,162 participants would have been included for analysis. This report uses the stricter criterion of accuracy $\frac{n_{\text{Correct}}}{n_{\text{TrialsSeen}}}$ for analysis, while an analysis using the rescored accuracy $\frac{n_{\text{Correct}}}{n_{\text{Answered}}}$ can be found in the Supplementary Information. The analyses reported below examine only those languages that met the minimum data criteria, which includes 32,897 total participants, of whom 29,155 completed at least one trial, 25,163 met the strict inclusion criteria and 26,197 met the rescored version of the inclusion criterion for accuracy.

The descriptive statistics of the participant data are provided below for the 25,163 participants who met the strict inclusion criteria.

Descriptive statistics

Participant (session)-level data. The following statistics are calculated by session, which generally represents one participant; however, participants could have taken the study multiple times. We describe these sessions as participants for ease of reading. Please see Fig. 2 for study procedure. We present the full sample information and the analysed sample information to demonstrate that the data analysed are similar to the full dataset. The participants predominantly self-identified as female (55.49%) or male (37.39%), with the rest missing data, not wanting to indicate their gender or indicating other. We used 'female', 'male', 'other' and 'prefer not to say' because these were the English labels on the survey. We asked participants to indicate their gender. Current norms suggest we should have used 'woman' and 'man' instead. We report the labels that were on the survey. In the analysed sample, the participants predominantly self-identified as female (60.95%) or male (37.44%). Looking at the entire sample, participants indicated they had completed high school (42.77%), some college (7.63%), college (30.47%), a master's degree (9.30%) or other options (less than high school, doctorate or missing). Participants included in the analysis also followed this pattern: high school (46.02%), some college (8.34%), college (31.97%) or a master's degree (9.61%). College was used to indicate university-type experience (community college or otherwise). 'Some college' indicated that they had not completed a degree but had completed some credits. We use the terms here that were listed on the survey, but the terminology for education was localized to the data collection area. Please see https://osf.io/vdgkr for the full participant information.

Full language percentage tables can be found in the Supplementary Information (https://osf.io/ta6wf, https://osf.io/652h8, Supplementary Table 1). The data indicate that the pattern of native languages was similar in the full data and the data used for analysis. The average self-reported age for all participants was 31.4 years (s.d. = 15.0), ranging from 18 to 104 years (median, 24; interquartile range, 20–39). In the demographic questions, we asked the participants to enter their year of birth, and the high maximum values probably belonged to participants who entered the minimum possible year allowable in the data collection form. The data of the participants included in the analysis showed the same age pattern: a mean of 30.4 years (s.d. = 14.2), ranging from 18 to 104 years (median, 24; interquartile range, 20–37).

The majority of participants used a Windows-based operating system (76.91%), followed by Mac OS (18.45%) and Linux (1.80%), with some missing data (2.85%) on the basis of browser metadata. The distribution of operating systems was similar for the participants used in the analysis: Windows (76.82%), Mac (18.70%), Linux (1.86%) and missing (2.61%). Web browsers were grouped into the largest categories for reporting as the data provided include specific version numbers. Most of the participants used Chrome (58.96%), followed by Edge (14.92%), Safari (8.88%), Firefox (8.18%), Opera (3.09%), Yandex (2.37%) and other web browsers (3.60%). The results were similar when examining only the participants who were included in the analysis: Chrome (59.81%), Edge (14.23%), Firefox (8.43%), Safari (9.22%), Opera (2.99%), Yandex (2.03%) and other browsers (3.29%). The full tables of browser languages can be found in the Supplementary Information (https://osf.io/93kep, https://osf.io/3yab7, Supplementary Table 1). Generally, this pattern matched the demographics of the study, as well as the targeted languages, except that more participants had their browser set in English than in the indicated native language.

Participants' overall proportion of correct answers was calculated, and participants who did not correctly answer at least 80% of the trials or saw fewer than 100 trials were marked for exclusion within the participant and trial-level datasets (see below). The average percentage of incorrect responses in the Semantic Priming Project was between 4%

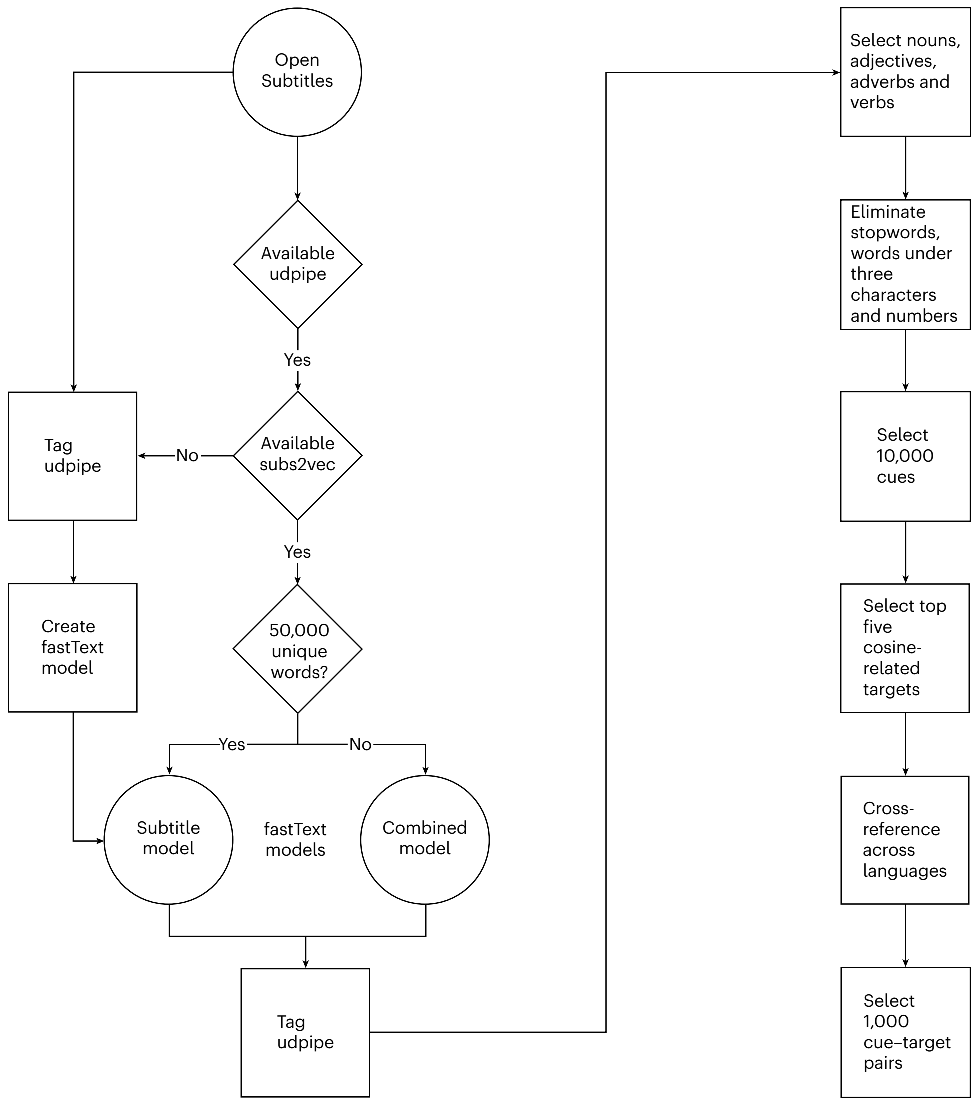

Fig. 2 | Stimulus selection method. Flow chart of the stimulus selection method. Circles represent the data or models used in the decision tree. Diamonds represent a decision criterion for the data selected. Squares represent coding processes or data reduction for the final stimulus set.

and 5%, and the 80% criterion was chosen to only include participants who were engaged in the experiment. Additionally, as noted above, two definitions of accuracy were identified by the lead team, and consequently, both criteria are provided.

The study lasted an average of 26.40 min (s.d. = 303.61). If a participant's computer went to sleep during the study, and they later returned to it (for example, to close the browser), the last time stamp would include the final time the study was open. Therefore, the median completion time of 17.88 min is probably more representative. The participants included in the analysis completed the study in 24.14 min on average (s.d. = 296.83; median, 17.97 min).

Trial-level data. Each language was saved in separate files in the online materials. Supplementary files (https://osf.io/q7e35 and https://osf.io/dmc6u) and code within semanticprimeR (https://osf.io/yd8u4) enable merging trials across concepts and pairings (for example, CAT (English) → KATZE (German) → GATTO (Italian)). If a participant left the study early (for example, their internet disconnected, their computer crashed or they closed the study), the data beyond that point were not recorded. The trial-level data therefore represent all trials displayed during the experiment, and new columns were added to denote different exclusion criteria at the trial level. We expected that participants would provide incorrect answers on some trials, and these trials were marked for exclusion. All timeout trials were marked as missing values in the final dataset. No missing values were imputed.

Trials were also marked for exclusion if they were under the minimum response latency of 160 ms[52]. Furthermore, lab.js automatically codes timeout data with a special marker (that is, data ended on response or timeout as a column), which excludes trials over 3,000 ms as the maximum response latency. However, because of variations in browser/screen refresh rates, some trials were answered with response latencies over 3,000 ms when a participant made a key press at the very end of the trial before timeout. Given the preregistered exclusion rules, these were also marked for exclusion.

The response latencies from each participant's session were then *Z*-scored following Faust et al.[53]. For privacy reasons, we did not collect identifying information to determine whether a person took the experiment multiple times, but as these are considered different sessions, the recommended *Z*-score procedure should control for participant variability at this level. Therefore, the possibility of repeated participation was not detrimental to data collection, especially with the large number of possible stimuli for a participant to receive within each session. Both *Z*-scores and raw response times are included in the provided data files. The Supplementary Information includes the number of trials

https://doi.org/10.1038/s41562-025-02254-x

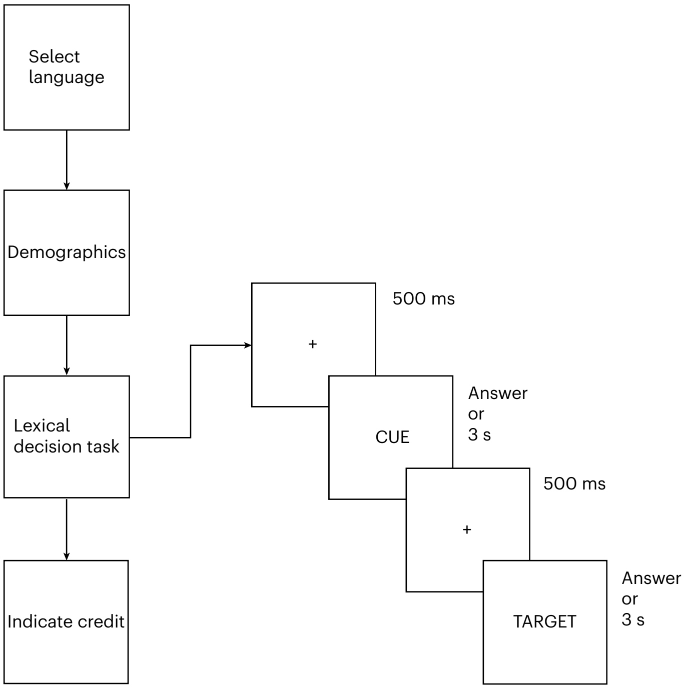

Fig. 3 | Study procedure. Flow chart of the procedure for the study. Within the lexical decision task, the participants were given short breaks after 100 trials. The answer choices for that language were always displayed at the bottom of the screen during the lexical decision task.

and accuracy for each language, for all participants and for analysed participants (https://osf.io/baem5, Supplementary Table 2). The mean Z-scores for all trials, regardless of item or related/unrelated condition, are presented in the summary files online (https://osf.io/baem5). The analyses averaged over item statistics are presented below.

Item-level data. The item-level data files can be matched with lexical information about all stimuli calculated from the OpenSubtitles[20] and subs2vec[46] projects using the semanticprimeR package (https://osf.io/yd8u4)[54]. Please see Fig. 3 for a description of the item selection procedure. The descriptive statistics calculated from the trial-level data are separated by language for each item: mean response latency, average standardized response latency, sample size, standard errors of response latencies and accuracy rate. No data points were excluded for being potential outliers (that is, no response latencies were excluded due to being outliers after the removal of excluded participants and trials mentioned above); however, we used recommended cut-off criteria for absolute-value Z-score outliers at 2.5 and 3.0 (ref. 21), and we calculated these same statistics with those subsets of trials excluded. For all real words, when available, values for age of acquisition, imageability, concreteness, valence, dominance, arousal and familiarity can be merged with the item files. These values do not exist for non-words. Online tables show the item statistics for average item sample size, average Z-scored response time and average s.e. for the Z-scored response latencies separated by item (non-word or word) type and language (https://osf.io/rvt8f, Supplementary Tables 3 and 4). The raw response time averages can be found in Supplementary Table 5. These values exclude both participants and trials from the exclusions listed above, and scores are calculated by creating item means and then averaging all item means.

Priming-level data. In separate files, we prepared information about the priming results in two forms: (1) priming trials that were converted from long data (that is, one trial per row) to wide data (that is, cue–target priming trial combinations paired together on one line) and (2) summary data, which include the list of target words, average response latencies, averaged Z-scored response latencies, sample sizes, standard errors and priming response latency (all files: https://github.com/SemanticPriming/SPAML/tree/v1.0.2; summary: https://osf.io/m8kjv). For each item, priming was defined as the average Z-scored response latency when presented in the unrelated minus the related condition. The timing for DOG–CAT would therefore be subtracted from BUS–CAT to indicate the priming effect for the word CAT. The similarity scores calculated during stimulus selection are provided for merging, as well as other established measures of similarity if they are available in that language. For example, semantic feature overlap norms are also available in Italian[55], German[56], Spanish[23], Dutch[57] and Chinese[58]. The overall priming averages by language are shown in Fig. 4 as part of Hypotheses 1 and 2. Supplementary Fig. 1 demonstrates the same distributions as raw response latencies.

Reliability. Item reliability was calculated by randomly splitting priming trials into two halves, calculating Z-score priming for each half and correlating those scores by item. The results below were calculated on the original accuracy scoring for all trials; the Supplementary Information includes the rescored accuracy versions (https://osf.io/r4fym and https://osf.io/jf28q; summary: https://osf.io/m8kjv). Participant-level reliability was calculated in a similar fashion by splitting participant related–unrelated trials in half, calculating priming as the average unrelated Z-scored response latency minus the related Z-scored response latency and correlating the two priming scores. The Spearman–Brown prophecy formula was applied to the average and median correlations across 100 random runs to estimate overall reliability. The average reliability was 0.56 for items (median, 0.56) and 0.08 for participants (median, 0.08). The Discussion compares these results to previous findings.

The correlation between average item sample size (averaged across both related and unrelated conditions) and item reliability is $r = 0.59$. A linear model of sample size predicting reliability indicates that an average sample size for unrelated and related conditions of $n \approx 557$ participants could potentially achieve a reliability of 0.80. Item reliability is probably impacted by other variables, as languages

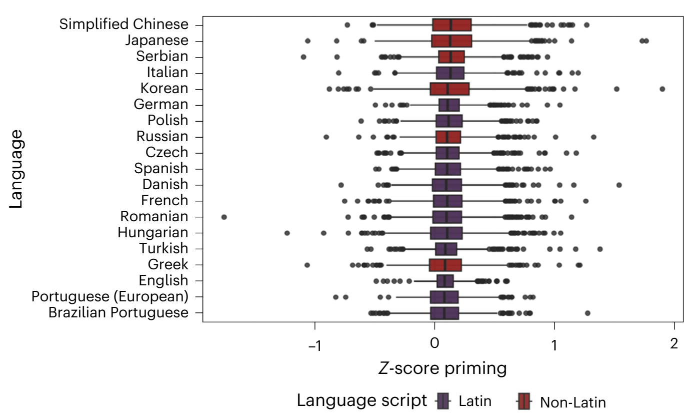

Fig. 4 | Average priming effect distributions. Distribution of average priming effects for languages that met the minimum sample size criteria, shown as box plots. Languages are ordered on the basis of their average priming effect from smallest (bottom) to largest (top). The preregistered language selection for the study included a requirement to ensure at least one non-Latin script within the language choices. The graph colour codes these languages for convenience to highlight the diversity in included languages. These plots represent all item average data without outliers removed (n per language, 1,000; total n, 19,000). The minimum value was $Z = -1.75$, maximum $Z = 1.90$. In each box plot, the median is represented as a solid bar, and the interquartile range as the box. The whiskers extend from the end of the box plot up to 1.5 times the interquartile range. See Supplementary Fig. 1 for the raw response times.

Table 3 | AIC values for intercept-only and random-effects models

	Overall	Z=2.5	Z=3.0
Intercept only	−6,613.93	−14,469.54	−12,977.97
Random effects	−6,711.77	−14,604.55	−13,104.04
Difference	97.84	135.01	126.07

such as Japanese showed higher reliability scores with smaller average item sample sizes (n = 68 versus English n = 356 with nearly identical reliabilities of r = 0.58 and r = 0.56).

Hypothesis 1

Hypothesis 1 predicted finding semantic facilitation wherein the response latencies for related targets would be faster than those for unrelated targets, as shown in Table 1. Hypothesis 1 was tested by fitting an intercept-only regression model using the Z-scored priming response latency as the dependent variable (https://osf.io/rmkag). The priming response latency was calculated by taking the average of the unrelated-pair Z-scored response latency minus the average related-pair response latency within each item by language. Values that are positive and greater than zero (that is, >0.0001) therefore indicate priming because the related pair had a faster response latency than the unrelated pair. The intercept and its 95% CI represent the grand mean of the priming effect across all languages.

The overall Z-scored priming effect was β_0 = 0.12 (s.e. = 0.001; 95% CI, (0.11, 0.12)). This process was repeated for average priming scores calculated without trials that were marked as 2.5 Z-score outliers and 3.0 Z-score outliers separately. These results were consistent with overall priming: $\beta_{0,Z=2.5}$ = 0.10 (s.e. = 0.001; 95% CI, (0.10, 0.11)) and $\beta_{0,Z=3.0}$ = 0.11 (s.e. = 0.001; 95% CI, (0.10, 0.11)). Figure 4 shows the distribution of the average item Z-score effects, ordered by the size of the overall priming effect for each language (see the raw response time effects in Supplementary Fig. 1). The distributions of the priming scores are very similar, with long tails and roughly similar shapes (albeit with more variance in some languages). For comparison to previous publications, the raw response latency priming was β_0 = 30.61 (s.e. = 0.43; 95% CI, (29.78, 31.45)), $\beta_{0,Z=2.5}$ = 27.12 (s.e. = 0.36; 95% CI, (26.51, 27.92)) and $\beta_{0,Z=3.0}$ = 28.08 (s.e. = 0.37; 95% CI, (27.35, 28.81)).

Hypothesis 2

Hypothesis 2 explored the extent to which these semantic priming effects vary across languages. We therefore calculated a random-effects model using the nlme[59] package in R wherein the random intercept of language was added to the overall intercept-only model for Hypothesis 1. Please see Table 3 for the AIC values and their difference scores for comparison. The addition of this parameter improved model fit, supporting significant heterogeneity as the value of the AIC for the random-effects model is two points or more lower than the value of the AIC for the intercept-only model[47]. The standard deviation of the random effect was 0.02 (95% CI, (0.01, 0.03)). The pseudo-R^2 for the model was 0.01 (ref. 60). The random effect was useful in both Z-score 2.5 and 3.0 models, wherein the random-effect sizes were similar to those in the overall model: $Z_{2.5}$ = 0.02 (95% CI, (0.01, 0.02)) and $Z_{3.0}$ = 0.02 (95% CI, (0.01, 0.03)).

Figure 5 shows the forest plot for the average priming effects by language, ordered by the size of the effect without the removal of outliers (see Supplementary Fig. 2 for the raw response time effects). The global priming average is presented on each facet to show how the priming effect changes depending on the removal of outliers. In nearly all languages, the priming effect decreases slightly with the removal of outliers. This figure also shows that the priming effect does vary by language, as supported by the results from Hypothesis 2, but that the effect is probably small, given that pseudo-R^2 was <0.01.

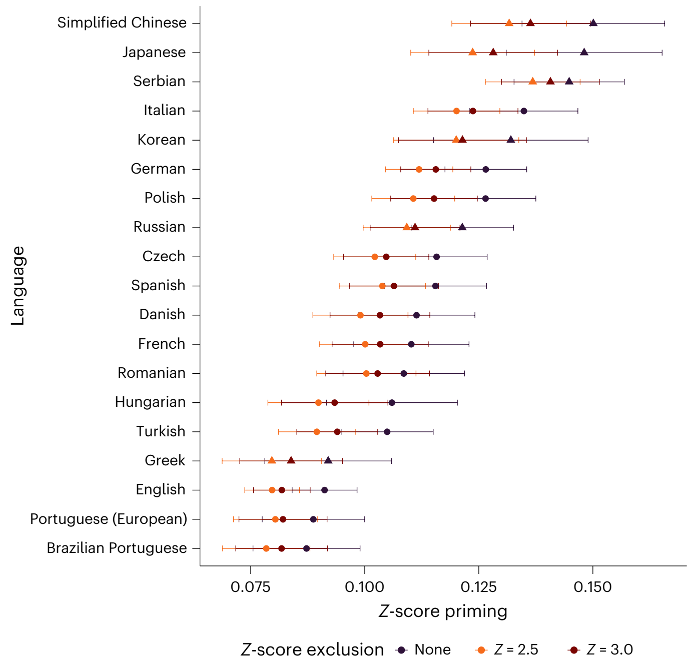

Fig. 5 | Priming effect sizes. Forest plot of average priming effects for each language ordered by priming average when no outliers are removed (least restrictive), Z-scores more than 2.5 are removed (most restrictive) and Z-scores more than 3.0 are removed. Sample sizes are based on item averages with n = 19,000 item averages. The error bars represent the 95% CIs. The plot indicates that all priming averages are positive, and their CIs do not include zero, as the lower end of the graph is approximately Z = 0.07, even with the removal of the outliers shown in Fig. 4. Triangles represent non-Latin languages, and circles represent Latin languages. The languages are ordered on the basis of average priming for the no-Z-score-removal condition from smallest (bottom) to largest (top). See Supplementary Fig. 2 for the raw response times and https://osf.io/m8kjv for the average Z-scores, the average raw response latencies and the standard errors used to create this diagram.

Discussion

This study represents the largest cross-linguistic study on semantic priming to date, with data collection in 30 languages using a set of coordinated stimuli. Using computational models of word embeddings and expanded linguistic corpora, we selected a stimulus set that covered semantic similarity across languages, rather than in a single language to be translated into others. Using a continuous lexical decision task, we collected more than 21 million trials using an adaptive stimulus presentation algorithm that shifted data collection towards uncertainty after a minimum number of trials. Data collection requirements were completed for 19 languages/dialects, with more than 700 participants in each language and coverage of both Latin and non-Latin-based scripts. Given the large proportion of published linguistic research that is still WEIRD[61], we provide a diversity of stimuli, participants and data that can be reused to examine new hypotheses, control stimuli in new studies and create cross-linguistic comparisons for previously found results.

In the 19 analysed languages, we demonstrated consistent non-zero priming effects ranging from Z = 0.09 to 0.15, and these effects are robust to the removal of strong priming pairs with high Z-scores such as ROMEO–JULIET, GOLDEN–SILVER, MENTAL–EMOTIONAL and BLIND–DEAF (that is, the highest positive Z priming scores across all languages, translated into their English counterparts). The Z-score removal also eliminates strong negative pairs, such as RESCUE–SAVE, FASHIONABLE–ELEGANT and POSITION–STATUS. The English dataset provided one of the lowest priming averages, $Z = 0.09$, even with an average cosine relatedness of 0.55 for related pairs (s.d. = 0.11; minimum, 0.22; maximum, 0.90). For comparison, the results of the Semantic Priming Project[21] demonstrated higher priming values when stimulus onset asynchronies were short (200 ms; $Z = 0.21$ for first associates, $Z = 0.14$ for other associates) but comparable values for longer stimulus onset asynchronies (1,200 ms; $Z = 0.16$ for first associates, $Z = 0.10$ for other associates). Given that participants also made lexical decisions on cue words in our study, the results should most closely match the longer stimulus onset asynchrony conditions because there is a longer time before the target is seen; accordingly, our results generally align with the Semantic Priming Project's results for other associates. Our results also demonstrate higher item reliability estimates than some estimates previously shown (0.04 (ref. 40) and 0.17–0.33 (ref. 38)) and are more in line with other estimates (0.66 standardized lexical decision task[39]). The participant reliability estimates are considerably lower than those in previous examinations of the Semantic Priming Project for first associates (0.21–0.27) but somewhat similar to results for other associates (0.07–0.08 (ref. 62)) and other studies (−0.06 to 0.43 (ref. 63)). The large sample sizes in this project probably boosted reliability results for item-level reliability, as the largest samples show some of the strongest reliability coefficients. Researchers interested in predicting semantic priming at the item level are advised to focus on those languages that showed the highest item reliability estimates, most notably Japanese, English and Russian.

Our secondary hypothesis examined the potential heterogeneity of priming effects across languages and revealed small but non-zero differences in levels of priming across languages. Differences between languages may be confounded with differences in data collection sites, participants and other variables. However, one key takeaway from Fig. 4 is the relatively similar distributions found for all languages. While Portuguese and Simplified Chinese show clearly non-overlapping CIs in Fig. 5 in each Z-score calculation, it is somewhat surprising that all means are within the CIs of previous (English) Z-score estimates for priming (that is, stimulus onset asynchrony, 1,200 ms; 95% CI, (0.14, 0.18) for first associates; 95% CI, (0.08, 0.12) for other associates) and how remarkably comparable the results are for each analysed language. Given the potential differences in translation, script, processing, culture and more, this result points to a generalizable cognitive mechanism for semantic priming. With the wealth of data provided in this project, researchers may begin to discern what variables predict differences found in the strength of priming effects at the language level, rather than within individual multilingual populations.

The limitations of this research include the necessity of picking a single design for semantic priming, but it does extend the available data to a new study type (that is, the Semantic Priming Project and others have used a paired (masked) priming task, while this study used a continuous lexical decision task)[2,21]. The study design does provide abundant data for all types of word-processing analyses, but it did not specifically target a single underlying cognitive mechanism for the explanation of priming effects (that is, automatic versus controlled processes). Moreover, only a few self-reported individual demographic variables are present to explore potential reasons for participant variability, and other studies may provide more individual-differences measures, such as reading and vocabulary measures[21]. This limited demographic data collection allowed the study to be conducted easily in many geopolitical regions, as institutional review boards vary widely in their approval of studies that collect identifying measures, especially with overseas data management (that is, they would rather the data be collected and stored locally). Furthermore, this procedure with limited demographic variables represents the normal approach for mega-studies to combat fatigue and different privacy regulations across the globe[64–66]. Finally, not all translated languages completed initial data collection; however, the data are available for use, and ideally, new low-resource languages would be added to new publications of the dataset.

In summary, our results demonstrate semantic priming and its variability across languages and cultural contexts (as multiple languages were collected in different geopolitical regions), using a controlled set of stimuli comprising matching target words. Future research may further explore the sources of variability in semantic priming evident within individuals, items and languages using the provided semanticprimeR package to merge datasets across other psycholinguistic variables. This study demonstrates the effectiveness of large-scale team collaboration in answering cross-linguistic questions, as well as providing resources for future reuse that are more complete (that is, have fewer missing values when combining databases) than individual lab contributions[17]. Although linguistics is largely still WEIRD, big team projects can continue to tackle sampling bias and generalizability problems within the field[43,61,67–69], using grassroots networks such as the PSA[14] and the ManyLanguages community (https://many-languages.com/).

Methods

All deviations from the methods and results can be found in the Supplementary Information (deviation list and https://osf.io/mwuv3). The data, code and other materials can all be found at https://github.com/SemanticPriming/SPAML.

Ethics information

We did not collect any identifiable private or personal data as part of the experiment. This project was approved by Harrisburg University of Science and Technology and conforms to all relevant ethical guidelines and the Declaration of Helsinki, with special care to conform to the General Data Protection Regulation (https://gdpr.eu/). Each research lab obtained local ethical review, relied on the ethical review provided by Harrisburg University or provided evidence of no required ethical review. The institutional review board approvals are available on the OSF: https://osf.io/wrpj4/. Participants were compensated for their participation by course credit or payment depending on individual lab resources. Labs recruited participants via their own local resources. No exclusion criteria for participating in the study were used, except for a minimum age requirement of 18 years (that is, adult participants).

Power analysis

For our power analysis, we first detail the background on how we estimated sample size, explain accuracy in parameter estimation, provide two simulations based on previous research and state the final proposed sample size. We end this section by specifying why this procedure was superior to previous methods and the requirements for publication.

Background. One concern is how to estimate the sample size required for cue–target pairs, as the previous literature indicates variability in their results[40]. Sample sizes of $n = 30$ per study have often been used in an attempt to at least meet some perceived minimum criteria for the central limit theorem. We focused on the lexical decision task for our procedure, wherein participants are simply asked if a concept presented to them is a word (for example, CAT) or a non-word (for example, GAT). The dependent variable in this study was response latency, and we used lexical decision data from the English Lexicon Project[22] and the Semantic Priming Project[21] to estimate the minimum sample size necessary for each item, as previous research has suggested an overall sample size may lead to unreliability in the item-level responses[40]. The English Lexicon Project contains lexical decision task data for over 40,000 words, while the Semantic Priming Project includes 1,661 target words.

Accuracy in parameter estimation. *Description*. In this approach, one selects a minimum sample size, a stopping rule and a maximum sample size. A minimum sample size was defined for each item on the basis of data simulation as described below. For the stopping rule, we focused on finding a CI around a parameter that would be "sufficiently narrow"[50,51,70]. These parameters are often tied to the statistical test or effect size for the study, such as the correlation or contrast between two groups. In this study, we paired accuracy in parameter estimation with a sequential testing procedure to adequately sample each item, rather than estimate an overall effect size. We therefore used the previous lexical decision data to determine our sufficiently narrow confidence by finding a generalized standard error one should expect for well-measured items. After the minimum sample size, each item's standard error was assessed to determine whether the item had met the goals for accuracy in parameter estimation as our stopping rule. If it had, the item was sampled at a lower probability in relation to other items until all items reached the accuracy goals or a maximum sample size determined by our simulations below (https://osf.io/v2y9e).

Estimates from the English Lexicon Project. First, the response latency data for the English Lexicon Project were Z-scored by participant and session as each participant has a somewhat arbitrary average response latency[53]. The data were then filtered for only real-word trials that were correctly answered. The average sample size before removing incorrect answers was 32.69 (s.d. = 0.63) participants with an average retention rate of 84% and 27.41 (s.d. = 6.43) participants after exclusions. The retention rates were skewed due to the large number of infrequent words in the English Lexicon Project, and we used the median retention rate of 91% for later sample size estimations. The median standard error for response latencies in the English Lexicon Project was 0.14, and the mean was 0.16. Because the retention rates were variable across items, we also calculated the average standard error for items that retained at least 30 participants at 0.12. This standard error rate represented the potential stopping rule.

The data were then sampled with replacement to determine the sample size that would provide that standard error value. One hundred words within the data were randomly selected, and samples starting at $n = 5$ to $n = 200$ were selected (increasing in units of five). The standard error for each of these samples was then calculated for the simulation, and the percentage of samples with standard errors at or less than the estimated population value was then tabulated. To achieve 80% of items at or below the proposed standard error, we needed approximately 50 participants per word. This value was used as our minimum sample size for the lexical decision task, and the accuracy standard error level was preliminarily set at 0.12.

Estimates from the Semantic Priming Project. This same procedure was followed with the Semantic Priming Project's lexical decision data on real-word trials. The priming response latencies were expected to be variable, as this priming strength should be predicted by other psycholinguistic variables, such as word relatedness. We therefore aimed to achieve an accurate representation of lexical decision times, from which priming could then be calculated. However, it should be noted that accurately measured response latencies do not necessarily imply reliable priming or difference score data[71], but larger sample sizes should provide more evidence of the picture of item-level reliability. We used these data paired with those from the English Lexicon Project to account for the differences in a lexical decision only versus a priming-focused task. The average standard error in the Semantic Priming Project was less at 0.06, probably because the data in the Semantic Priming Project are generally frequent nouns and only 1,661 concepts, as compared with the 40,000 in the English Lexicon Project. The retention rate for the Semantic Priming Project was less skewed than that for the English Lexicon Project at a median of 97% and mean of 96%. Using the same sampling procedure, we estimated sample sizes of $n = 5$ to $n = 400$ participants increasing by units of five. In this scenario, we found the maximum sample size of 320 participants for 80% of the items to reach the smaller standard error of 0.06. We therefore used 320 as our maximum sample size and the average of the two standard errors found as our stopping rule (that is, 0.09).

Final sample size. Given our minimum, maximum and stopping rule, we then estimated the final sample size per language on the basis of study design characteristics. Participants completed approximately 800 lexical decision trials per session, and each participant completed only 150 of these concepts (75 targets in the related condition and 75 targets in the unrelated condition; cue words were not analysed) that were the target of this sample size analysis (see below for more details on trial composition). The target number of items ($n = 1{,}000$ concepts) was therefore multiplied by the minimum/maximum sample size and the number of conditions (related word pair versus unrelated word pair) and divided by the total number of critical lexical decision trials per participant times the data retention rate (a conservative estimate of 90%). The final estimate for sample size per language was 741 to 4,741 ($(1{,}000 \times 50 \times 2)/(150 \times 0.90)$ to $(1{,}000 \times 320 \times 2)/(150 \times 0.90)$). The complete code and description of this process are detailed in the Supplementary Information (https://osf.io/rxgkf and https://osf.io/v2y9e).

This sample size estimation represents a major improvement from previous database collection studies, as many have used the traditional $n = 30$ to guess at minimum sample size. Because the variability of the sample size was quite large, we employed a stopping procedure to ensure participant time and effort were maximized and data collection was optimized. To summarize, the minimum sample size was 50 participants per word, and the maximum for the adaptive procedure was 320, which resulted in 741 to 4,741 participants per language on the basis of expected usable trials. The total sample size was therefore proposed to be 7,410 to 47,410 participants for ten languages. After 50 participants who answered a real-word item, each concept was examined for standard error, and data collection for that concept was decreased in probability when the standard error reached our average criterion of 0.09. Item probability for selection was also decreased when they reached the maximum proposed sample size ($n = 320$). This process was automated online and checked in a scheduled subroutine.

While 43 languages were identified for possible data collection, we planned to first publish the data when ten languages had reached the appropriate sample size as outlined above on the basis of recruitment of PSA partner labs. We aimed to complete minimum data collection in English, Spanish, Chinese, Portuguese, German, Korean, Russian, Turkish, Czech and Japanese. To date, we have recruited more than 100 researchers in 19 potential languages.

Materials

The following sections detail the important facets of the materials. We first explain the types of word-pair conditions in a semantic priming study (that is, related, unrelated and non-word). Next, we detail how the related word-pair conditions were created using the OpenSubtitles corpora, new computational modelling techniques and the selection procedure.

Word-pair conditions. In a semantic priming study, there are three types of word-pair conditions. In the related-word-pair condition, cue–target pairs are chosen for their similarity or relatedness. Cosine distance is similar to correlation in representing relatedness; however, cosine distance is always positive. A cosine distance of 1 therefore represents the same numeric vectors (perfect similarity), while a cosine distance of 0 represents no similarity between vectors. To create the unrelated condition, cue–target pairs were shuffled so that the cue word was combined with a target word with which it had a negligible cosine distance similarity (that is, <0.15).

Finally, non-word-pair conditions were created by using the Wuggy-like algorithm[72] for non-logographic languages. For logographic languages, we consulted with at least two native speakers to change one stroke or radical such that the characters were a pronounceable word

 https://doi.org/10.1038/s41562-025-02254-x

with no meaning by starting from known non-word lists[73]. Any disagreements between native speakers were resolved by discussion between these speakers. Each cue and target word were first hyphenated using the sylly package and LaTeX style hyphenation[74]. If words were not hyphenated, as they were one syllable or the syllables were not clear, we created bigram character pairs for replacement purposes. The 100,000 most frequent words for each language from the OpenSubtitles data were also hyphenated in this style. From the OpenSubtitles data, we calculated the frequency of each pair of possible hyphenation combinations (for example, NAPKIN → [_, NAP], [NAP, KIN], [KIN, _]) as the transition frequency from Wuggy. For each cue and target, we selected a set of character replacements that kept or closely matched the same number of characters as the original word, minimized transition frequency (that is, the frequency of the replacement was very close to the frequency of the original pair of hyphenated characters) and matched the number of character changes to the number of syllables. At least two native speakers examined each programmatically generated word to ensure they were pronounceable (that is, phonologically valid) and not pseudo-homophones (that is, wherein the pronunciation sounds like a real word, as in KEEP → KEAP)[72]. In cases of disagreement, the native speakers discussed and resolved these inconsistencies. When they marked a non-word for exclusion, a new non-word was generated until speakers agreed it met the rules for non-words. Native speakers also suggested alternatives, which the lead author checked to ensure that they matched the desired non-word characteristics. These files can be found on OSF (https://osf.io/wrpj4/) or GitHub (https://github.com/SemanticPriming/SPAML) under 03_Materials separated by language code.

To control the ability of participants to anticipate or guess the answers, we ensured that half the trials should be answered with a word and half with a non-word. We therefore used 150 related trials (150 words and 0 non-words; 75 pairs), 150 unrelated trials (150 words and 0 non-words; 75 pairs), 200 word–non-word trials (100 words and 100 non-words; this could have been word–non-word or non-word–word combinations to control for answer chaining; 100 pairs) and 300 non-word–non-word trials (0 words and 300 non-words; 150 pairs). These trials were randomly presented to control the transition probability between word and non-word trials (that is, random presentation should ensure trials do not present a word–word–non-word–non-word-style pattern that allows participants to mindlessly guess the answers). Therefore, the yes–no probability was 50% for words–non-words across all trials, and the relatedness proportion for pairs was 18.8%. The exact trial proportions for each language can be found online in our data processing summary, as not all participants completed all trials, which can change the proportions for each language (https://osf.io/zye59).

Similarity calculation. *Corpora*. The OpenSubtitles data include 62 languages or language combinations (for example, Chinese–English mix). We used the 10,000 most frequent nouns, adjectives, adverbs and verbs from each potential language without lemmatization (that is, converting words into their dictionary form, as in RUNS → RUN). The udpipe package[84] is a natural language processing package that contains more than 100 treebanks to assist in part-of-speech tagging (that is, labelling words as noun, verb and so on), parsing (that is, separating blocks of text into words and determining their relationships to other words in a text) and lemmatization. This package was selected for its large coverage of languages with reliable part-of-speech tagging. Cross-referencing the available languages in udpipe with the OpenSubtitles data allowed for the possibility of 43 different languages in this project. See Fig. 2 for the model selection process.

As described in the introduction, the choice of related words based on similarity was key for the study. There are multiple measures of semantic similarity, including the cosine similarity between overlapping features[32], free association probabilities[33,34,75] and local/global coherence values from network models. However, the underlying data for these calculations are inconsistent across languages. One solution is to use the data present in the OpenSubtitles datasets[20] (that is, a large collection of movie subtitles) to calculate word frequency and cosine similarity values. These datasets have been used to calculate word frequencies for the SUBTLEX projects, which have validated their use as strong predictors of cognition-related phenomena[18,76–83]. Cosine similarity was selected over other similarity measures because of the availability of possible languages and models for this project, as described below.

Modelling. The subs2vec project[46] used the OpenSubtitles data to create fastText[85] computational representations for 55 languages. fastText is a distributional vector space model, an extension of word2vec[44,45], wherein each word in a corpus is converted to a vector of numbers that represents the relationship of that word to a number of dimensions. These dimensions can be imagined as a thematic or topic representation of the text. The relationship between these vectors represents the similarity between concepts, as words that have similar or related meanings will appear in similar places and dimensions in a text and will therefore have similar numeric vectors[4,5]. We used the existing models from subs2vec to extract related word concepts for the most frequent concepts identified using the top cosine distance between word vectors. When the model was not present in subs2vec, we recreated the same model using their parameters on the relevant OpenSubtitles data.

Cue selection procedure. The selected token list was then tagged for part of speech using udpipe, selecting tokens that were tagged as nouns, adjectives, adverbs and verbs. From the udpipe output, the lemma for each token was selected to control for high similarity between lemma-token forms (for example, RUN is highly related to RUNS). All stopwords (that is, commonly used words in a language with little semantic meaning such as THE, AN and OF), words with fewer than three characters for non-logographic languages and words with numeric characters were eliminated (that is, 1 would be eliminated but not ONE). The stopword lists can be found in the stopwords package using the Stopwords ISO dataset[86]. This procedure covered all but two languages in our list of 43 possible languages. For the final two languages, we used udpipe to tag the OpenSubtitles directly and calculate word frequency. Additionally, fastText models using the same parameters as those for subs2vec were trained for similarity calculation. The 10,000 most frequent concepts were selected at this point.

The procedure for stimulus selection can be reviewed in the Supplementary Information and is displayed graphically in Fig. 2 (https://osf.io/mz7p4, https://osf.io/s9h3z). If the language was available via subs2vec, the provided subtitle frequency counts were examined. If the language had more than 50,000 unique concepts represented in the subtitle data, we used the subtitle model only. If the subtitles did not provide enough linguistic information (that is, fewer than 50,000 concepts in the corpus), we used the combined Wikipedia and subtitle model[46]. subs2vec contains a model with only the OpenSubtitles data, a model with only Wikipedia for a given language and a combined model of both. The subtitle data have been shown to best represent a language[18,76]; however, not all subtitle projects contain a large enough corpus for the subtitles to cover the breadth of the possible concepts within that language (for example, Afrikaans subtitles only represent approximately 18,000 words).

Target selection procedure. Using the fastText models for each language, we selected the top five cosine distance similarity values for each concept in each language independently, resulting in 50,000 possible cue–target pairs. These were cross-referenced across languages using Google Translate to create a master list of potential cue–target pairings. The related word pairs (n = 1,000) were selected from this list using each cue or target only once, favouring pairs with translations in most languages. The selection procedure was therefore based on the most common cue–target pairs across languages, rather than selecting similar words in one language and then translating. This procedure was programmatic, using Google Translate, which may not produce the most appropriate translation for a word. Native speakers therefore ensured the accurate translation of word pairs using the PSA's translation network for the final selected set in a similar manner as described above. They suggested a more common or appropriate word for items they thought were unusual, and in cases of disagreement, group discussion between the two translators took place. In some instances, translation indicated that a particular language does not have separate concepts for the cue–target pairing. In these instances, we changed the cue word to a related word for that language from the five selected in the original list. Thus, all targets were matched across languages, and as many cues as possible were matched while avoiding repetition within a cue–target pair. Translation information is located at https://osf.io/vdme5 within the 03_Materials folders shared online.

Procedure

We describe the important components of the procedure in this section. First, we detail the implementation of the study, focusing on the timing software and adaptive stimulus section, as not all participants saw all items. We then discuss the study procedure in order, as shown in Fig. 3. The participants first completed a demographic questionnaire, followed by the lexical decision task. We explain how our data complements those from the Semantic Priming Project and finally discuss additional data that researchers can combine with the current dataset.

Implementation

Timing software. While the participants were naïve to the word pairings, the principal investigator knew the pair combinations during data collection and analysis. A small demonstration of the experiment can be found at https://psa007.psysciacc.org/ or recreated from the Supplementary Information (on OSF or GitHub, use the 04_Procedure folder). The study was programmed using lab.js[87], which is an online, open-source study-building software. Precise timing measurement was required for this study. The lab.js team has documented the accuracy of measurement within their framework[88], and previous work has shown no differences between lab- and web-based data collection for response latencies[89]. In addition, SPALEX, a large lexical decision database in Spanish, was collected completely online[23]. We recommended that research labs suggest Chrome as their browser for participants completing the study due to recommendations from the lab.js team. However, meta-information about the browser and operating system were saved when participants completed the experiment to examine for potential implementation differences.

Participants were directed to an online web portal to complete the study, and all data were retained in the online platform with regular backups to the server. The participants were required to complete the study on a computer with a keyboard, rather than on a device with only a touch screen. This requirement allows for tracking of the display of the device, which indicates important aspects about screen size, browser and timing accuracy. To enforce this requirement, the participants were asked to hit the space bar to continue the study.

Adaptive stimulus selection. At the start of data collection, all presented items were randomly selected from the larger item pool by equalizing the probability of inclusion for all words and non-words ($p = 1/1{,}000$ concepts). After the minimum sample size was collected, each word's standard error was checked to determine whether the sample size for that item had reached our accuracy criteria. If so, the probability of sampling that item was decreased by half. Once a concept had reached the maximum required sample size, the probability of sampling was also decreased by half. This procedure allowed for random sampling of the items that still needed participants without eliminating words from the item pool. We therefore ensured that there were always words to randomly select from (that is, to keep the same procedure and number of trials for all participants) and that the randomization was a sampled mix of words that reached accuracy quickly and words that needed more participants (that is, participants did not see only the unusual words at the end of data collection). Once all words reached the stopping criteria or maximum sample size, the probabilities were equalized. We set the minimum, the maximum and a stopping rule for the initial data collection; however, we allowed data collection after these were reached and will post updates to the data using GitHub releases (modelled after the Small World of Words Project[33], which is ongoing). All data were included in our dataset, and the analysis section describes how we indicated exclusion criteria. Data collection was therefore a repeated-measures design in which participants did not see all of the possible stimuli but did see all the possible conditions (related, unrelated and non-word pairs). The participants were blinded to condition, and the explicit link between pairs was not explained to them.

Study procedure. *Demographics*. Participants were given a language-specific link for each research lab. The participants were asked to indicate their gender (that is, male, female, other or prefer not to say), year of birth and education level (that is, none, elementary school, high school, bachelor's, master's, doctorate or their equivalent in the target country of data collection) as demographic variables. They provided their native language in an open text box and selected left or right as their dominant hand for the mapping of word–non-word answer keys (see below). A flow chart of the procedure is provided in Fig. 3.

Lexical decision task. After 100 trials, the participants were shown a short break screen with the option to continue by hitting the space bar after 10 s. This break timed out after 60 s. After eight blocks of 100 trials (800 word–non-word decisions), the experiment ended with a thank-you screen. On this screen, the participants were given instructions on how to indicate that they had completed the study to the appropriate lab. The participants were allowed to take the study multiple times as items were randomly selected for inclusion. An estimate of the time required for the study was approximately 30 min inclusive of practice trials, reading all instructions and breaks. This estimate was based on previous studies of lexical decision times[22], and the final median completion time was approximately 18 min.

Instructions on how to complete a lexical decision task were shown on the next screen, followed by ten practice trials. Each trial started with a fixation cross (+) in the middle of the screen for 500 ms. The stimulus item was then displayed in the middle of the screen in lowercase sans-serif 18-point font (that is, Arial font). On the bottom of the screen, the possible responses were shown as the traditional keys next to the Shift key depending on the most common keyboard layout for that language (that is, Z and / on a QWERTY keyboard or < and - on a QWERTZ keyboard or the numbers 1 and 9 for languages that had many keyboard layouts). Response keys were mapped such that the 'non-word' response option was on the non-dominant-hand side of the keyboard, and the 'word' response option was on the dominant-hand side[90]. Participants made their choice for each concept, and during the practice trials, they received feedback on whether their answer was correct or incorrect. The next stimulus appeared with an intertrial interval of 500 ms (that is, the time between the offset of the first concept response and the onset of the next concept, when the fixation cross was showing). Responses timed out after three seconds and moved on to the next trial. After ten trials, the participants saw the instruction screen again with a reminder that they would now be doing the real task.

Comparison to the Semantic Priming Project. This procedure is a continuous lexical decision task wherein every concept (cue and target) is judged for lexicality (that is, word/non-word). Many priming studies often present cue words for a short time prior to the presentation of target words for lexicality judgement. Evidence from the Semantic Priming Project suggests that the stimulus onset asynchrony (that is, the time between the non-judged cue word and the target word) does not affect overall priming rates (25 versus 23 ms for 200 ms and 1,200 ms). Furthermore, adding the lexicality judgement to each presented concept creates a less obvious link between cue and target to avoid potential conscious expectancy generation effects[91,92]. Even though they appear sequentially in the task, they are not explicitly paired by being a non-judged cue word followed by a judged target word. This procedure therefore differs from that used to collect the data in the Semantic Priming Project, thus extending their work to different conditions. Lucas[15] provides evidence that priming effect sizes are relatively equal across task type (for example, continuous, masked, paired and naming); therefore, we should expect similar results.

Additional data. We then combined available lexical and subject rating data with the priming data. A tutorial is provided in the Supplementary Information on how to download data and combine them with available norms (https://osf.io/yd8u4). Lexical measures such as length, frequency, part of speech and the number of phonemes (that is, sounds in a word) are easily created from the concept or the SUBTLEX projects[76–82]. Subjective measures are concept characteristics that are rated by participants, and we included age of acquisition[93–96] (the approximate age at which a person learned a concept), imageability[97,98] (how easy the concept comes to mind), concreteness[99] (how concrete the concept is), valence (how positive versus negative the concept is), arousal (how excited or calm a concept makes a person), dominance (whether the word denotes something that is weak/subordinate or strong/dominant)[24,26] and familiarity (how well a person knows a concept)[100]. These variables were selected from the list of most published databases for linguistic data[17].

Protocol registration
The preregistration is at https://osf.io/u5bp6 (updated 31 May 2022).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
All raw and processed data are available via GitHub at https://github.com/SemanticPriming/SPAML.

Code availability
All code used for study creation and delivery, data processing, and analyses is available via OSF (https://osf.io/wrpj4/) and GitHub (https://github.com/SemanticPriming/SPAML).

References

1. Meyer, D. E. & Schvaneveldt, R. W. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* **90**, 227–234 (1971).
2. McNamara, T. P. *Semantic Priming* (Psychology Press, 2005); https://doi.org/10.4324/9780203338001
3. Mandera, P., Keuleers, E. & Brysbaert, M. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: a review and empirical validation. *J. Mem. Lang.* **92**, 57–78 (2017).
4. Cree, G. S. & Armstrong, B. C. in *The Cambridge Handbook of Psycholinguistics* (eds Spivey, M. et al.) 259–282 (Cambridge Univ. Press, 2012); https://doi.org/10.1017/CBO9781139029377.014
5. McRae, K. & Jones, M. *Semantic Memory* (Oxford Univ. Press, 2013); https://doi.org/10.1093/oxfordhb/9780195376746.013.0014
6. Rogers, T. T. in *The Cambridge Handbook of Computational Psychology* (ed. Sun, R.) 226–266 (Cambridge Univ. Press, 2001); https://doi.org/10.1017/CBO9780511816772.012
7. Frings, C., Schneider, K. K. & Fox, E. The negative priming paradigm: an update and implications for selective attention. *Psychon. Bull. Rev.* **22**, 1577–1597 (2015).
8. Spruyt, A., De Houwer, J., Everaert, T. & Hermans, D. Unconscious semantic activation depends on feature-specific attention allocation. *Cognition* **122**, 91–95 (2012).
9. McDonough, K. & Trofimovich, P. *Using Priming Methods in Second Language Research* (Routledge, 2011); https://doi.org/10.4324/9780203880944
10. Singh, L. One world, two languages: cross-language semantic priming in bilingual toddlers. *Child Dev.* **85**, 755–766 (2014).
11. Kiefer, M. et al. Neuro-cognitive mechanisms of conscious and unconscious visual perception: from a plethora of phenomena to general principles. *Adv. Cogn. Psychol.* **7**, 55–67 (2011).
12. Steinhauer, K., Royle, P., Drury, J. E. & Fromont, L. A. The priming of priming: evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neurosci. Lett.* **651**, 192–197 (2017).
13. Liu, B., Wu, G., Meng, X. & Dang, J. Correlation between prime duration and semantic priming effect: evidence from N400 effect. *Neuroscience* **238**, 319–326 (2013).
14. Moshontz, H. et al. The Psychological Science Accelerator: advancing psychology through a distributed collaborative network. *Adv. Methods Pract. Psychol. Sci.* **1**, 501–515 (2018).
15. Lucas, M. Semantic priming without association: a meta-analytic review. *Psychon. Bull. Rev.* **7**, 618–630 (2000).
16. Hutchison, K. A. Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychon. Bull. Rev.* **10**, 785–813 (2003).
17. Buchanan, E. M., Valentine, K. D. & Maxwell, N. P. LAB: Linguistic Annotated Bibliography—a searchable portal for normed database information. *Behav. Res.* **51**, 1878–1888 (2019).
18. New, B., Brysbaert, M., Veronis, J. & Pallier, C. The use of film subtitles to estimate word frequencies. *Appl. Psycholing.* **28**, 661–677 (2007).
19. Gimenes, M. & New, B. Worldlex: Twitter and blog word frequencies for 66 languages. *Behav. Res.* **48**, 963–972 (2016).
20. Lison, P. & Tiedemann, J. OpenSubtitles2016: extracting large parallel corpora from movie and TV subtitles. In *Proc. Tenth International Conference on Language Resources and Evaluation* (LREC'16) (eds Calzolari, N. et al.) 923–929 (European Language Resources Association, 2016).
21. Hutchison, K. A. et al. The Semantic Priming Project. *Behav. Res.* **45**, 1099–1114 (2013).
22. Balota, D. A. et al. The English Lexicon Project. *Behav. Res. Methods* **39**, 445–459 (2007).
23. Aguasvivas, J. A. et al. SPALEX: a Spanish lexical decision database from a massive online data collection. *Front. Psychol.* **9**, 2156 (2018).
24. Bradley, M. M. & Lang, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **25**, 49–59 (1994).
25. Warriner, A. B., Kuperman, V. & Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res.* **45**, 1191–1207 (2013).
26. Bradley, M. M. & Lang, P. J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings* (Univ. of Florida, 1999).
27. Brysbaert, M., Warriner, A. B. & Kuperman, V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res.* **46**, 904–911 (2014).

28. Stadthagen-Gonzalez, H. & Davis, C. J. The Bristol norms for age of acquisition, imageability, and familiarity. *Behav. Res. Methods* **38**, 598–605 (2006).
29. Kerr, N. L. HARKing: hypothesizing after the results are known. *Pers. Soc. Psychol. Rev.* **2**, 196–217 (1998).
30. Cree, G. S., McRae, K. & McNorgan, C. An attractor model of lexical conceptual processing: simulating semantic priming. *Cogn. Sci.* **23**, 371–414 (1999).
31. Zannino, G. D., Perri, R., Pasqualetti, P., Caltagirone, C. & Carlesimo, G. A. Analysis of the semantic representations of living and nonliving concepts: a normative study. *Cogn. Neuropsychol.* **23**, 515–540 (2006).
32. Buchanan, E. M., Valentine, K. D. & Maxwell, N. P. English semantic feature production norms: an extended database of 4436 concepts. *Behav. Res.* **51**, 1849–1863 (2019).
33. De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M. & Storms, G. The 'Small World of Words' English word association norms for over 12,000 cue words. *Behav. Res.* **51**, 987–1006 (2019).
34. Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. The University of South Florida free association, rhyme, and word fragment norms. *Behav. Res. Methods Instrum. Comput.* **36**, 402–407 (2004).
35. Landauer, T. K., Foltz, P. W. & Laham, D. An introduction to latent semantic analysis. *Discourse Process.* **25**, 259–284 (1998).
36. Landauer, T. K. & Dumais, S. T. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211–240 (1997).
37. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* **28**, 203–208 (1996).
38. Heyman, T., Hutchison, K. A. & Storms, G. Uncovering underlying processes of semantic priming by correlating item-level effects. *Psychon. Bull. Rev.* **23**, 540–547 (2016).
39. Hutchison, K. A., Balota, D. A., Cortese, M. J. & Watson, J. M. Predicting semantic priming at the item level. *Q. J. Exp. Psychol.* **61**, 1036–1066 (2008).
40. Heyman, T., Bruninx, A., Hutchison, K. A. & Storms, G. The (un) reliability of item-level semantic priming effects. *Behav. Res.* **50**, 2173–2183 (2018).
41. Perea, M., Duñabeitia, J. A. & Carreiras, M. Masked associative/semantic priming effects across languages with highly proficient bilinguals. *J. Mem. Lang.* **58**, 916–930 (2008).
42. Guasch, M., Sánchez-Casas, R., Ferré, P. & García-Albea, J. E. Effects of the degree of meaning similarity on cross-language semantic priming in highly proficient bilinguals. *J. Cogn. Psychol.* **23**, 942–961 (2011).
43. Levisen, C. Biases we live by: Anglocentrism in linguistics and cognitive sciences. *Lang. Sci.* **76**, 101173 (2019).
44. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **26**, 3111–3119 (2013).
45. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. Preprint at https://doi.org/10.48550/arXiv.1301.3781 (2013).
46. van Paridon, J. & Thompson, B. subs2vec: word embeddings from subtitles in 55 languages. *Behav. Res.* **53**, 629–655 (2021).
47. Burnham, K. P. & Anderson, D. R. Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).
48. Tzelgov, J. & Eben-ezra, S. Components of the between-language semantic priming effect. *Eur. J. Cogn. Psychol.* **4**, 253–272 (1992).
49. Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L. & Jain, M. The bilingual lexicon: language-specific units in an integrated network. *J. Verbal Learn. Verbal Behav.* **23**, 519–539 (1984).
50. Kelley, K. Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behav. Res. Methods* **39**, 755–766 (2007).
51. Kelley, K., Darku, F. B. & Chattopadhyay, B. Accuracy in parameter estimation for a general class of effect sizes: a sequential approach. *Psychol. Methods* **23**, 226–243 (2018).
52. Proctor, R. W. & Schneider, D. W. Hick's law for choice reaction time: a review. *Q. J. Exp. Psychol.* **71**, 1281–1299 (2018).
53. Faust, M. E., Balota, D. A., Spieler, D. H. & Ferraro, F. R. Individual differences in information-processing rate and amount: implications for group differences in response latency. *Psychol. Bull.* **125**, 777–799 (1999).
54. Buchanan, E. SemanticPriming/semanticprimeR: semanticprimeR package. *Zenodo* https://doi.org/10.5281/ZENODO.10697999 (2024).
55. Montefinese, M., Ambrosini, E., Fairfield, B. & Mammarella, N. Semantic memory: a feature-based analysis and new norms for Italian. *Behav. Res.* **45**, 440–461 (2013).
56. Kremer, G. & Baroni, M. A set of semantic norms for German and Italian. *Behav. Res.* **43**, 97–109 (2011).
57. Ruts, W. et al. Dutch norm data for 13 semantic categories and 338 exemplars. *Behav. Res. Methods Instrum. Comput.* **36**, 506–515 (2004).
58. Deng, Y. et al. A Chinese Conceptual Semantic Feature Dataset (CCFD). *Behav. Res.* **53**, 1697–1709 (2021).
59. Pinheiro, J., Bates, D., Debroy, S., Sarkar, D. & R Core Team. nlme: Linear and nonlinear mixed effects models. R package v3.1.166 (2017).
60. Bartoń, K. MuMIn: Multi-model inference. R package v1.48.4 (2020).
61. Bochynska, A. et al. Reproducible research practices and transparency across linguistics. *Glossa Psycholinguistics* https://doi.org/10.5070/G6011239 (2023).
62. Yap, M. J., Hutchison, K. A. & Tan, L. C. in *Big Data in Cognitive Science* (ed. Jones, M. N.) 203–226 (Routledge/Taylor & Francis Group, 2017).
63. Stolz, J. A., Besner, D. & Carr, T. H. Implications of measures of reliability for theories of priming: activity in semantic memory is inherently noisy and uncoordinated. *Vis. Cogn.* **12**, 284–336 (2005).
64. Siegelman, N. et al. Rethinking first language–second language similarities and differences in English proficiency: insights from the ENglish Reading Online (ENRO) Project. *Lang. Learn.* **74**, 249–294 (2024).
65. Mandera, P., Keuleers, E. & Brysbaert, M. Recognition times for 62 thousand English words: data from the English Crowdsourcing Project. *Behav. Res.* **52**, 741–760 (2020).
66. Kuperman, V. et al. Text reading in English as a second language: evidence from the Multilingual Eye-Movements Corpus. *Stud. Second Lang. Acquis.* **45**, 3–37 (2023).
67. Andringa, S. & Godfroid, A. Sampling bias and the problem of generalizability in applied linguistics. *Annu. Rev. Appl. Linguist.* **40**, 134–142 (2020).
68. Sulpizio, S. et al. Taboo language across the globe: a multi-lab study. *Behav. Res.* **56**, 3794–3813 (2024).
69. Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D. & Majid, A. Over-reliance on English hinders cognitive science. *Trends Cogn. Sci.* **26**, 1153–1170 (2022).
70. Maxwell, S. E., Kelley, K. & Rausch, J. R. Sample size planning for statistical power and accuracy in parameter estimation. *Annu. Rev. Psychol.* **59**, 537–563 (2008).
71. Overall, J. E. & Woodward, J. A. Unreliability of difference scores: a paradox for measurement of change. *Psychol. Bull.* **82**, 85–86 (1975).

72. Keuleers, E. & Brysbaert, M. Wuggy: a multilingual pseudoword generator. *Behav. Res. Methods* **42**, 627–633 (2010).
73. Tse, C.-S. et al. The Chinese Lexicon Project: a megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behav. Res.* **49**, 1503–1519 (2017).
74. Michalke, M. sylly: Hyphenation and syllable counting for text analysis. R package v0.1.6 (2020).
75. De Deyne, S., Navarro, D. J. & Storms, G. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behav. Res.* **45**, 480–498 (2013).
76. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* **41**, 977–990 (2009).
77. van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. Subtlex-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* **67**, 1176–1190 (2014).
78. Keuleers, E., Brysbaert, M. & New, B. SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav. Res. Methods* **42**, 643–650 (2010).
79. Cai, Q. & Brysbaert, M. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE* **5**, e10729 (2010).
80. Brysbaert, M. et al. The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp. Psychol.* **58**, 412–424 (2011).
81. Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J. & Carreiras, M. Subtitle-based word frequencies as the best estimate of reading behavior: the case of Greek. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2010.00218 (2010).
82. Mandera, P., Keuleers, E., Wodniecka, Z. & Brysbaert, M. Subtlex-pl: subtitle-based word frequency estimates for Polish. *Behav. Res.* **47**, 471–483 (2015).
83. Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A. & Carreiras, M. EsPal: one-stop shopping for Spanish word properties. *Behav. Res.* **45**, 1246–1258 (2013).
84. Wijffels, J. et al. udpipe: Tokenization, parts of speech tagging, lemmatization and dependency parsing with the 'UDPipe' 'NLP' toolkit. R package v0.8.11 (2021).
85. Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching word vectors with subword information. *Trans. Assoc.Comput. Linguist.* **5**, 135–146 (2017).
86. Benoit, K., Muhr, D. & Watanabe, K. stopwords: Multilingual stopword lists. R package v2.3 (2021).
87. Henninger, F., Shevchenko, Y. & Mertens, U.K. et al. lab.js: A free, open, online study builder. *Behav. Res.* **54**, 556–573 (2022).
88. Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. & Hilbig, B. E. Who said browser-based experiments can't have proper timing? Implementing accurate presentation and response timing in browser. *Presentation for the 48th Annual Meeting of The Society for Computers in Psychology* (2018).
89. Hilbig, B. E. Reaction time effects in lab- versus web-based research: experimental evidence. *Behav. Res.* **48**, 1718–1724 (2016).
90. Proctor, R. W. & Cho, Y. S. Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychol. Bull.* **132**, 416–442 (2006).
91. Neely, J. H., Keefe, D. E. & Ross, K. L. Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *J. Exp. Psychol. Learn. Mem. Cogn.* **15**, 1003–1019 (1989).
92. Shelton, J. R. & Martin, R. C. How semantic is automatic semantic priming? *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 1191–1210 (1992).
93. Johnston, R. A. & Barry, C. Age of acquisition and lexical processing. *Vis. Cogn.* **13**, 789–845 (2006).
94. Ghyselinck, M., Lewis, M. B. & Brysbaert, M. Age of acquisition and the cumulative-frequency hypothesis: a review of the literature and a new multi-task investigation. *Acta Psychol.* **115**, 43–67 (2004).
95. Juhasz, B. J. Age-of-acquisition effects in word and picture identification. *Psychol. Bull.* **131**, 684–712 (2005).
96. Brysbaert, M. & Ellis, A. W. Aphasia and age of acquisition: are early-learned words more resilient? *Aphasiology* **30**, 1240–1263 (2016).
97. Richardson, J. T. E. Imageability and concreteness. *Bull. Psychon. Soc.* **7**, 429–431 (1976).
98. Richardson, J. T. E. Concreteness and imageability. *Q. J. Exp. Psychol.* **27**, 235–249 (1975).
99. Paivio, A., Walsh, M. & Bons, T. Concreteness effects on memory: when and why? *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 1196–1204 (1994).
100. Wilson, M. MRC psycholinguistic database: machine-usable dictionary, version 2.00. *Behav. Res. Methods Instrum. Comput.* **20**, 6–10 (1988).

Acknowledgements

A.D.A. was supported by the South-Eastern Norway Regional Health Authority (no. 2020023). J.L.U. was supported by ANID/CONICYT FONDECYT Iniciación 11190673, Programa de Investigación Asociativa en Ciencias Cognitivas (RU-158-2019), Research Center on Cognitive Sciences, Faculty of Psychology, Universidad de Talca, Chile. P.K. was supported by APVV-22-0458. Y.Z. was supported by the National Natural Science Foundation of China (grant nos 72272048, 72343035, 72432003, 71902164, 71972065, 72272049 and 72102060) and the Post-Funding Project of Philosophy and Social Science Research of the Ministry of Education (grant no. 21JHQ088). S.W. was supported by the Deutsche Forschungsgemeinschaft (DFG) Heisenberg Programme (funding ID: 442405852). Y.A.N., A. Stückler, F.T., M.H.C. and S. Pfattheicher supported the data collection in Danish with Interacting Minds Centre seed grant no. 26254. M. Marelli was supported by ERC Consolidator Grant 101087053 (project 'BraveNewWord'). A. Sepehri was supported by ESSEC Business School Research Center (CERESSEC). E.M.B. was supported by funding from the Einstein Foundation Award through the PSA, Harrisburg University of Science and Technology, and the Leibniz Institute for Psychology (ZPID). C.H. was supported by the Fund for Scientific Research Flanders (FWO), grant no. FWO G049821N. P.A. was supported by Fundação para a Ciência e Tecnologia through the Research Center CIS-Iscte (UID/PSI/03125/2020). M. Köster was supported by DFG grant no. 290878970-GRK 2271. M. Cavdan was supported by DFG project no. 502774891-ORA project 'UNTOUCH'. M. A. Vadillo was supported by grant no. PID2020-118583GB-I00 from Agencia Estatal de Investigación (Spain). D. Grigoryev was supported by HSE University Basic Research Program. S.C.R. and P.S. were supported by IDN Being Human Incubator. Y. Yamada was supported by JSPS KAKENHI (JP22K18263). K. Schmidt was supported by the John Templeton Foundation. R.M.R. was supported by the John Templeton Foundation (grant ID: 62631). K.K. was supported by the Austrian Science Fund (FWF) (grant DOI: 10.55776/ESP286). K.B. and E.I. were supported by a grant from the National Science Centre, Poland (2019/35/B/HS6/00528). M.M.E. was supported by Leverhulme Early Career Fellow ECF-2022-761. L.R. was supported by the National Recovery and Resilience Plan (PNRR), Mission 4, Component 2, Investment 1.1, Call for Tender No. 104 published on 2 February 2022 by the Italian Ministry of University and Research, funded by the European Union—NextGenerationEU—project title 'The World in Words: Moving beyond a spatiocentric view of the human mind (acronym: WoWo)', project code 2022TE3XMT, CUP (Rinaldi) F53D23004850006; and by the Italian Ministry of Health (Ricerca Corrente 2024). M. Kowal was supported by the Foundation for Polish Science START scholarship. I.R. was supported by Mediated Society (MEDIS:ON) CZ.02.01.01/00/23_025/0008713, which is co-financed by the European Union and by APVV-23-0421. D.M.-R. was supported by a National Master's Scholarship from the Agencia Nacional de Investigación y Desarrollo of Chile 3537/2023. R.J. was supported by the National Science Center (2020/37/B/HS6/00610). A.M.A. was supported by the OTKA FK 146604 research grant. M. Adamkovič was supported by PRIMUS/24/SSH/017; APVV-22-0458. M. Perea was supported by grant no. CIAICO/2021/172 from the Department of Innovation, Universities, Science and Digital Society of the Valencian Government. J.H.-w.H. was supported by the Research Grant Council of Hong Kong (GRF no. 17608621 to Hsiao). A. Sorokowska was supported by Scientific Excellence Incubator 'Being Human'. W.D. was supported by the FWO, project no. G049821N. K. Wolfe was supported by the Leverhulme Trust (RPG-2020-035). T.V.P. was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, according to the contract on the financial support of the scientific research of teaching staff at accredited higher-education institutions in 2024, contract no. 451-03-65/2024-03/200105. Z.P. was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, as part of the financial support of scientific research at the University of Belgrade—Faculty of Philosophy (contract no. 451-03-66/2024-03/200163). D.A.S.E.-D. was supported by Prince Sultan University through the Applied Linguistics Research Lab (RL-CH-2019/9/1). M. Comesaña was supported by The study corresponding to European Portuguese data was conducted at the Psychology Research Centre (PSI/01662), School of Psychology, University of Minho, and was supported by the Foundation for Science and Technology through the Portuguese State Budget (Ref.: UIDB/PSI/01662/2020). Z.M. and R.Z. were supported by ERDF/ESF project TECHSCALE (no. CZ.02.01.01/00/22_008/0004587). K.F. was supported by a Marie-Curie-Fellowship (882168). F.S. was supported by the Austrian Research Promotion Agency (FFG). S.D.D. was supported by a Social Sciences and Humanities Research Council of Canada grant (grant no. 435-2021-1074). M. Montefinese was supported by the Investment line 1.2 'Funding projects presented by young researchers' (CHILDCONTROL) from the European Union—NextGenerationEU. I.S.P. was supported by a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University). C. Blaison was supported by ANR-18-IDEX-0001 and ANR-20-FRAL-0008. T.N. was supported by the University Excellence Fund of ELTE Eötvös Loránd University, Budapest, Hungary, and the János Bolyai scholarship of the Hungarian Academy of Sciences. D. Muller was supported by Université Grenoble Alpes, Institut Universitaire de France. We thank I. Castillejo for their assistance with data collection. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper. An introductory sentence in the paper was edited for clarity: "Semantic priming is defined as the decrease in response latency (that is, reduced linguistic processing or facilitation) for target words that are semantically related to their cue words in comparison to unrelated cue words". Links to the supplementary materials were added to the Methods for direct access to noted materials.

Author contributions

The authors made the following contributions (https://osf.io/uv27t). E.M.B.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, and writing—review and editing. K.C.: project administration, supervision, and writing—review and editing. T.H.: conceptualization, data curation, methodology, project administration, validation, and writing—review and editing. N.v.B.: methodology, project administration, software, and writing—review and editing. N.A.C.: validation and writing—review and editing. A.I., K.P. and A.E.v.V.: project administration and writing—review and editing. M. Montefinese: conceptualization, investigation, methodology, resources, writing—original draft, and writing—review and editing. N.P.M.: conceptualization, investigation, methodology, and writing—review and editing. J.E.T. and K.D.V.: conceptualization, methodology, writing—original draft, and writing—review and editing. P.A.: funding acquisition, investigation, resources, and writing—review and editing. K.B., L. Boucher, W.M.C. and D.C.V.: investigation, resources, supervision, and writing—review and editing. B. Aczel, A.H.A.-H., E.A., T.B., D.I.B., M.M.B., A.J.B.C., Y.D., D.A.S.E.-D., M.M.E., M.F.-L., P.R.d.S.F., R.M.K.F., C.A.G., H.G., P.A.G., P. Halama, P. Havan, N.C.I., C.I., R.K.I., Y.J., A. Karaaslan, M. Kohút, V.K., J.K., A.I.K., T.J.S.d.L., M.H.C.M., C. Manouilidou, L.A.M., C. Morvinski, A.M., F.E.M., Y.A.N., J.C.O., J.O., M.P.-P., I.P., Z.P., B.P., G.P., E.P., T.B.R., C.R.R., S.R.-F., M. Senderecka, Ç.S., A. Stückler, R.D.S.-C., A.R.T., R.T., U.S.T., J.L.U., M. A. Varga, S.V., T.V.P., G.V., N.W., K.Z., C.C.Y. and S. Patel: investigation, resources, and writing—review and editing. X.M.M.: project administration, resources, and writing—review and editing. J.F.M.: investigation, project administration, and writing—review and editing. F.T.: funding acquisition, investigation, and writing—review and editing. O.J.C., O. A. Acar, M. Adamkovič, G.A., A.M.A., Z.A., S.A., O. A. Ananyeva, M. Andreychik, B. Angele, D.C.A.Q., N.C.A., E. Baskin, L. Batalha, C. Batres, M.S.B., M. Becker, M. Becker, M. Behnke, C. Blaison, E. Brandstätter, N.B., Á.C., Z.G.C., E.C.C., M. Cavdan, L.C., S.E.C., F.X.W.C., C.R.C., E.C., M.H.C., H.C.-P., P.C., M. Comesaña, C.C., S.D.D., O.A.D., W.E.D., B.J.W.D., H.D., R.D., W.D., L.A.E., C.E., T.R.E., G.F., K.F., S.E.F., H.D.F., P.G.-V., D. Gatti, V.G., A.S.G., L.M.G., D. Grigoryev, I.G., H.G., G. Handjaras, C.H., R.M.H., A.H., A.M.H., W.H., J.H.-w.H., G. Huang, E.I., A.D., K.J., D.C.J., R.J., J.J., L. Kaczer, J.A.K., A. Karner, P.K., J.J.K., K.K., M. Kowal, A.O.K., L. Körtvélyessy, F.E.K., M. Köster, M. Kękuś, M.L., C.L., J.L., W.L., G.L., C.A.L., J.G.L., S.E.M., K.M., D.M.-R., N.M., M. Marelli, M. Martínez, M.F.M., A.D.A.M., J.M.-S., D.P.M., Z.M., F.M., M. Misiak, S.M., K.E.M., D. Muller, T.N., M. Naranowicz, I.L.G.N., L.N., C.E.O., F.J.P., A.J.P., M.P.-C., Y.G.P., S. Paydarfard, D.P., M. Peker, M. Perea, S. Pfattheicher, J.P., I.S.P., K.P.-B., Z.Q., G.R., L.R., S.C.R., T.C.R., M.O.R., R.M.R., J.P.R., F.R., E. Sampaolo, A.C.S., F.Ç.S., F.S., K. Schmidt, A. Sepehri, H.O.S., T.S., C.S.Q.S., M.E.S., M. Sirota, A. Sorokowska, P.S., I.D.S., L.M.S., S.L.K.S., D. Steyrl, S. Stieger, A. Studzinska, M. Suarez, A. Szmalec, D. Sznycer, E. Szumowska, S. Söylemez, B.S., J.C.R.T., C.T., P.T., J.T., B.C.T., J.G.T., V.T.-M., A.K.T., B.T., M.T., B.N.T., M. A. Vadillo, C.V., M.E.W.V., M.R.V., L.A.V., F.V., L.M.V., S.W., L.W., G.P.W., D.W., K. Wolfe, A.S.W., Y. Zhou, M.Z., W.Z., Y. Zheng, R.Z., N.M.Z., O.Ç., S.Ç., S.Ö., A.A.Ö., S.M.Ş., D.K., A.A., M.F.C., A.G., C.K., I.R., Y.Z. and X.Z.: investigation and writing—review and editing. A.D.A., B.J.B., A.M.B., J.B., S.-C.C., C.W.C., E.G.D.B., V.E., A.E., C.F., Z.H., K.H., K.I., T.I., X.J., K.L.K., M. Korbmacher, T.K., Y.K., C.M.L., C. Mazzuca, L.C.P.M., D. Moreau, M. Neta, J.J.W.P., A.E.P., T.T.S., K. Sasaki, A. Szala, K.T., C.K.T., K. Wang, Y. Yamada, Y. Yang and C.Z.: resources and writing—review and editing. J.G., M.J.K., U.-D.R., C. Brick and J.W.S.: validation and writing—review and editing. S.C.L.: investigation, project administration, supervision, and writing—review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02254-x.

https://doi.org/10.1038/s41562-025-02254-x

Correspondence and requests for materials should be addressed to Erin M. Buchanan.

Peer review information *Nature Human Behaviour* thanks Marc Brysbaert and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Erin M. Buchanan[1] ✉, Kelly Cuccolo[2], Tom Heyman[3], Niels van Berkel[4], Nicholas A. Coles[5], Aishwarya Iyer[6], Kim Peters[7], A. E. van 't Veer[3], Maria Montefinese[8], Nicholas P. Maxwell[9], Jack E. Taylor[10], Kathrene D. Valentine[11,12], Patrícia Arriaga[13], Krystian Barzykowski[14], Leanne Boucher[15], W. Matthew Collins[15], David C. Vaidis[16,17], Balazs Aczel[18], Ali H. Al-Hoorie[19], Ettore Ambrosini[20], Théo Besson[21], Debora I. Burin[22,23], Muhammad Mussaffa Butt[24], A. J. Benjamin Clarke[25], Yalda Daryani[26], Dina Abdel Salam El-Dakhs[27], Mahmoud M. Elsherif[28,29], Maria Fernández-López[30], Paulo Roberto dos Santos Ferreira[31], Raquel Meister Ko Freitag[32], Carolina A. Gattei[33,34], Hendrik Godbersen[35], Philip A. Grim II[36], Peter Halama[37], Patrik Havan[38], Natalia C. Irrazabal[39], Chris Isloi[40], Rebecca Kvisler Iversen[41], Yoann Julliard[42,43], Aslan Karaaslan[44,45], Michal Kohút[46], Veronika Kohútová[47], Julija Kos[48], Alexandra I. Kosachenko[49], Tiago Jessé Souza de Lima[50], Matthew H. C. Mak[51], Christina Manouilidou[48], Leonardo A. Marciaga[52], Xiaolin Melinna Melinna[53], Jacob Francisco Miranda[54], Coby Morvinski[55], Aishwarya Muppoor[15], F. Elif Müjdeci[56], Yngwie A. Nielsen[57], Juan Carlos Oliveros[58], Jaš Onič[48], Marietta Papadatou-Pastou[59], Ishani Patel[60], Zoran Pavlović[61], Blaž Pažon[48], Gerit Pfuhl[62,63], Ekaterina Pronizius[64], Timo B. Roettger[65], Camilo R. Ronderos[66], Susana Ruiz-Fernandez[67], Magdalena Senderecka[68], Çağlar Solak[69], Anna Stückler[70], Raluca D. Szekely-Copîndean[71,72], Analí R. Taboh[73,74], Rémi Thériault[75], Ulrich S. Tran[64], Fabio Trecca[76], José Luis Ulloa[77], Marton A. Varga[78], Steven Verheyen[79], Tijana Vesić Pavlović[80], Giada Viviani[8], Nan Wang[81], Kristyna Zivna[82], Chen Chu Yun[83], Oliver James Clark[84], Oguz A. Acar[85], Matúš Adamkovič[86,87,88], Giulia Agnoletti[33,89], Atakan M. Akil[18,90], Zainab Alsuhaibani[91], Simona Amenta[92], Olga A. Ananyeva[93], Michael Andreychik[94], Bernhard Angele[95], Danna Catalina Arias Quiñones[96], Nwadiogo Chisom Arinze[97], Adrian Dahl Askelund[98,99], Bradley J. Baker[100], Ernest Baskin[101], Luisa Batalha[102], Carlota Batres[96], Maria Soledad Beato[103], Manuel Becker[104], Maja Becker[16], Maciej Behnke[105], Christophe Blaison[106], Anna M. Borghi[107,108], Eduard Brandstätter[109], Jacek Buczny[110], Nesrin Budak[111], Álvaro Cabana[112], Zhenguang G. Cai[113], Enrique C. Canessa[114,115], Müge Cavdan[116], Luca Cecchetti[117], Sergio E. Chaigneau[118], Feria X. W. Chang[119], Christopher R. Chartier[120], Sau-Chin Chen[121], Elena Cherniaeva[93], Morten H. Christiansen[122,123], Hu Chuan-Peng[83], Patrycja Chwiłkowska[124], Montserrat Comesaña[125], Chin Wen Cong[126], Casey Cowan[127], Stéphane Daniel Dandeneau[128], Oana A. David[129], William E. Davis[130], Elif Gizem Demirag Burak[131], Barnaby James Wyld Dixson[132,133], Hongfei Du[134,135], Rod Duclos[136], Wouter Duyck[137], Liudmila A. Efimova[93], Ciara Egan[127], Vanessa Era[138,139], Thomas R. Evans[140,141], Anna Exner[142], Gilad Feldman[143], Katharina Fellnhofer[144,145], Chiara Fini[146], Sarah E. Fisher[120], Heather D. Flowe[147], Patricia Garrido-Vásquez[148], Daniele Gatti[149], Jason Geller[150], Vaitsa Giannouli[151], Anna Sergeevna Gorokhova[93], Lindsay M. Griener[152], Dmitry Grigoryev[153], Igor Grossmann[154,155], Hesam Ghasemi[156], Giacomo Handjaras[157], Cathy Hauspie[137], Zhiran He[158], Renata M. Heilman[72], Amirmahdi Heydari[159], Alanna M. Hine[127], Karlijn Hoyer[160], Weronika Hryniszak[161], Janet Hui-wen Hsiao[162], Guanxiong Huang[163], Keiko Ihaya[164], Ewa Ilczuk[165], Tatsunori Ishii[166], Andrei Dumbravă[167,168], Katarzyna Jankowiak[169], Xiaoming Jiang[170,171], David C. Johnson[172,173], Rafał Jończyk[169,174], Juhani Järvikivi[175], Laura Kaczer[176], Kevin Leander Kamermans[177], Johannes A. Karl[178,179,180], Alexander Karner[181], Pavol Kačmár[182], Jacob J. Keech[183], M. Justin Kim[184,185], Max Korbmacher[186,187], Kathrin Kostorz[64], Marta Kowal[188], Tomas Kratochvil[189], Yoshihiko Kunisato[190], Anna O. Kuzminska[191], Lívia Körtvélyessy[192], Fatma Ebru Köse[193,194], Massimo Köster[104], Magdalena Kękuś[195], Melanie Labusch[196,197], Claus Lamm[64], Chaak Ming Lau[198], Julieta Laurino[176], Wilbert Law[199,200], Giada Lettieri[201], Carmel A. Levitan[202], Jackson G. Lu[203], Sarah E. MacPherson[204], Klara Malinakova[82], Diego Manriquez-Robles[205], Nicolás Marchant[206], Marco Marelli[92], Martín Martínez[207], Molly F. Matthews[208], Alan D. A. Mattiassi[209], Josefina Mattoli-Sánchez[210], Claudia Mazzuca[107], David P. McGovern[178], Zdenek Meier[82], Filip Melinscak[181], Michal Misiak[188,211], Luis Carlos Pereira Monteiro[212], David Moreau[213,214], Sebastian Moreno[114], Kate E. Mulgrew[215], Dominique Muller[42,43], Tamás Nagy[18], Marcin Naranowicz[216], Izuchukwu L. G. Ndukaihe[217], Maital Neta[218], Lukas Novak[82], Chisom Esther Ogbonnaya[219], Jessica Jee Won Paek[220], Aspasia Eleni Paltoglou[221,222], Francisco J. Parada[223], Adam J. Parker[224], Mariola Paruzel-Czachura[225,226], Yuri G. Pavlov[227], Saeed Paydarfard[228],

Dominik Pegler[64], Mehmet Peker[229], Manuel Perea[197,230], Stefan Pfattheicher[70], John Protzko[231], Irina Sergeevna Prusova[93], Katarzyna Pypno-Blajda[225], Zhuang Qiu[113,232], Ulf-Dietrich Reips[233], Gianni Ribeiro[234,235], Luca Rinaldi[236,237], S. Craig Roberts[188,238], Tanja C. Roembke[239], Marina O. Romanova[93], Robert M. Ross[240], Jan Philipp Röer[241], Filiz Rızaoğlu[242], Toni T. Saari[243], Erika Sampaolo[117], Anabela Caetano Santos[244], F. Çağlar Sarıçiçek[245], Kyoshiro Sasaki[246], Frank Scharnowski[64], Kathleen Schmidt[120], Amir Sepehri[247], Halid O. Serçe[248], A. Timur Sevincer[249], Cynthia S. Q. Siew[119], Matilde Ellen Simonetti[250], Miroslav Sirota[251], Agnieszka Sorokowska[252], Piotr Sorokowski[252], Ian D. Stephen[253], Laura M. Stevens[254], Suzanne L. K. Stewart[255], David Steyrl[64], Stefan Stieger[256], Anna Studzinska[257], Mar Suarez[103], Anna Szala[258,259], Arnaud Szmalec[137,260], Daniel Sznycer[261], Ewa Szumowska[262,263], Sinem Söylemez[69], Bahadır Söylemez[264], Kaito Takashima[265], Christian K. Tamnes[266], Joel C. R. Tan[119], Chengxiang Tang[267], Peter Tavel[82], Julian Tejada[268], Benjamin C. Thompson[269], Jake G. Tiernan[178], Vicente Torres-Muñoz[270], Anna K. Touloumakos[271], Bastien Trémolière[16,272], Monika Tschense[273], Belgüzar Nilay Türkan[274], Miguel A. Vadillo[275], Caterina Vannucci[117], Michael E. W. Varnum[276], Martin R. Vasilev[224], Leigh Ann Vaughn[277], Fanny Verkampt[278], Liliana M. Villar[279,280], Sebastian Wallot[281], Lijun Wang[282], Ke Wang[283], Glenn Patrick Williams[284], David Willinger[256], Kelly Wolfe[285,286], Alexandra S. Wormley[276,287], Yuki Yamada[288], Yunkai Yang[289], Yuwei Zhou[199], Mengfan Zhang[64], Wang Zheng[290], Yueyuan Zheng[291], Chenghao Zhou[292], Radka Zidkova[82], Nina Meret Zumbrunn[178], Ogeday Çoker[293], Sami Çoksan[294,295], Sezin Öner[245], Asil Ali Özdoğru[296,297], Seda Merve Şahin[111], Dauren Kasanov[298], Alexios Arvanitis[299], Cameron Brick[300], Melissa F. Colloff[254], Albina Gallyamova[301], Christopher Koch[302], Ivan Ropovik[303,304], Yucheng Zhang[305], Xingxing Zhou[306], Sneh Patel[15], Jordan W. Suchow[307] & Savannah C. Lewis[120,269]

[1]Analytics, Harrisburg University of Science and Technology, Harrisburg, PA, USA. [2]Independent researcher, Alma, MI, USA. [3]Methodology and Statistics Unit, Institute of Psychology, Leiden University, Leiden, the Netherlands. [4]Department of Computer Science, Aalborg University, Aalborg, Denmark. [5]Center for the Study of Language and Information, Stanford University, Stanford, CA, USA. [6]Department of Psychology, Christ University, Bangalore, India. [7]University of Exeter Business School, University of Exeter, Exeter, UK. [8]Department of Developmental Psychology and Socialisation, University of Padova, Padua, Italy. [9]Psychology, Midwestern State University, Wichita Falls, TX, USA. [10]Department of Psychology, Goethe University Frankfurt, Frankfurt, Germany. [11]General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [12]Medicine, Harvard Medical School, Boston, MA, USA. [13]Department of Psychology, CIS-Iscte, University Institute of Lisbon, Lisbon, Portugal. [14]Institute of Psychology, Faculty of Philosophy, Jagiellonian University, Krakow, Poland. [15]Department of Psychology and Neuroscience, Nova Southeastern University, Fort Lauderdale, FL, USA. [16]CLLE, University of Toulouse, CNRS, Toulouse, France. [17]Institute of Psychology, Social Psychology Laboratory, Paris Cité University, Paris, France. [18]Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary. [19]Royal Commission for Jubail and Yanbu, Jubail Industrial City, Saudi Arabia. [20]Department of Neuroscience, Padova Neuroscience Center, University of Padova, Padua, Italy. [21]Social Psychology Laboratory, Paris Cité University, Boulogne-Billancourt, France. [22]Faculty of Psychology, University of Buenos Aires, CONICET, Buenos Aires, Argentina. [23]Department of Psychology, University of Palermo, Buenos Aires, Argentina. [24]Department of Psychology, Government College University, Lahore, Pakistan. [25]Department of Linguistics, Thammasat University, Pathum Thani, Thailand. [26]Department of Psychology, University of Tehran, Tehran, Iran. [27]College of Humanities and Sciences, Prince Sultan University, Riyadh, Saudi Arabia. [28]Department of Psychology, University of Birmingham, Birmingham, UK. [29]Department of Vision Sciences, University of Leicester, Leicester, UK. [30]Department of Basic Psychology and ERI-Lectura, University of València, València, Spain. [31]Instituto de Psicologia, Federal University of Uberlândia, Uberlândia, Brazil. [32]Departamento de Letras Vernáculas, Federal University of Sergipe, São Cristóvão, Brazil. [33]School of Government, Torcuato Di Tella University, Buenos Aires, Argentina. [34]Institute of Linguistics, Faculty of Philosophy and Letters, University of Buenos Aires, CONICET, Buenos Aires, Argentina. [35]FOM University of Applied Sciences, Essen, Germany. [36]Computer and Information Sciences, Harrisburg University of Science and Technology, Harrisburg, PA, USA. [37]Centre of Social and Psychological Sciences, Slovak Academy of Sciences, Bratislava, Slovakia. [38]Institute of Experimental Psychology, Centre of Social and Psychological Sciences, Slovak Academy of Sciences, Bratislava, Slovak Republic. [39]Department of Psychology, University of Palermo, CONICET, Buenos Aires, Argentina. [40]Unaffiliated researcher, London, UK. [41]Department of Philosophy, Classics, History of Art and Ideas (IFIKK), University of Oslo, Oslo, Norway. [42]LIP/PC2S, University of Grenoble Alpes, Grenoble, France. [43]LIP/PC2S, University of Savoy Mont Blanc, Chambéry, France. [44]Department of Psychology, Izmir Katip Celebi University, Izmir, Turkey. [45]Department of Psychology, Ege University, Izmir, Turkey. [46]Department of Psychology, Faculty of Philosophy and Arts, University of Trnava, Trnava, Slovakia. [47]Faculty of Philosophy and Arts, University of Trnava, Trnava, Slovakia. [48]Department of Comparative and General Linguistics, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia. [49]Department of Psychology, Ural Federal University, Ekaterinburg, Russian Federation. [50]Department of Social and Work Psychology, Social and Affective Neuroscience Institute, University of Brasília, Brasília, Brazil. [51]Department of Psychology, University of Warwick, Coventry, UK. [52]Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL, USA. [53]School of Optometry, Hong Kong Polytechnic University, Hong Kong, China. [54]Psychology Department, California State University—East Bay, Hayward, CA, USA. [55]Department of Management, Ben-Gurion University of the Negev, Be'er Sheva, Israel. [56]Department of Psychology, Bilkent University, Ankara, Turkey. [57]School of Communication and Culture, Interacting Minds Centre, Aarhus University, Aarhus, Denmark. [58]Department of Psychology, Catholic University of Maule, Talca, Chile. [59]Department of Primary Education, National and Kapodistrian University of Athens, Athens, Greece. [60]College of Psychology, Nova Southeastern University, Fort Lauderdale, FL, USA. [61]Department of Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia. [62]Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway. [63]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway. [64]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria. [65]Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway. [66]Center for Languages and Literature, Lund University, Lund, Sweden. [67]Department of Psychology, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus-Senftenberg, Germany. [68]Centre for Cognitive Science, Jagiellonian University, Krakow, Poland. [69]Department of Psychology, Manisa Celal Bayar University, Manisa, Turkey. [70]Department of Psychology and Behavioural Sciences, Aarhus University,

Aarhus, Denmark. [71]Department of Social and Human Research, Romanian Academy, Cluj-Napoca, Romania. [72]Department of Psychology, Babeș-Bolyai University, Cluj-Napoca, Romania. [73]CONICET, Institute of Interdisciplinary and Applied Physics (INFINA), University of Buenos Aires, Buenos Aires, Argentina. [74]School of Business, Neuroscience Lab, Torcuato Di Tella University, Buenos Aires, Argentina. [75]Department of Psychology, University of Quebec in Montreal, Montreal, Quebec, Canada. [76]TrygFonden's Centre for Child Research, School of Communication and Culture, Aarhus University, Aarhus, Denmark. [77]Associative Research Program in Cognitive Sciences, Research Center in Cognitive Sciences, Faculty of Psychology, University of Talca, Talca, Chile. [78]ELTE Eötvös Loránd University, Budapest, Hungary. [79]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands. [80]Faculty of Mechanical Engineering, University of Belgrade, Belgrade, Serbia. [81]Department of Linguistics, The Chinese University of Hong Kong, Hong Kong, China. [82]Olomouc University Social Health Institute, Palacký University Olomouc, Olomouc, Czechia. [83]School of Psychology, Nanjing Normal University, Nanjing, China. [84]Faculty of Health and Education, Manchester Metropolitan University, Manchester, UK. [85]King's Business School, King's College London, London, UK. [86]Institute of Social Sciences CSPS, Slovak Academy of Sciences, Košice, Slovakia. [87]Faculty of Education, Charles University, Prague, Czechia. [88]Faculty of Humanities and Social Sciences, University of Jyväskylä, Jyväskylä, Finland. [89]School of Government, Favaloro University, Buenos Aires, Argentina. [90]Institute of Psychology, Faculty of Humanities and Social Sciences, University of Pécs, Pécs, Hungary. [91]College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. [92]Department of Psychology, University of Milano-Bicocca, Milan, Italy. [93]Laboratory for Psychology of Social Inequality, National Research University Higher School of Economics, Moscow, Russian Federation. [94]Department of Psychological and Brain Sciences, Fairfield University, Fairfield, CT, USA. [95]Nebrija Research Center in Cognition (CINC), Antonio de Nebrija University, Madrid, Spain. [96]Department of Psychology, Franklin and Marshall College, Lancaster, PA, USA. [97]Psychology, Alex Ekwueme Federal University, Ndufu-Alike, Nigeria. [98]Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway. [99]PsychGen Center for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway. [100]Department of Sport, Tourism and Hospitality Management, Temple University, Philadelphia, PA, USA. [101]Department of Food, Pharma, and Healthcare, Saint Joseph's University, Philadelphia, PA, USA. [102]School of Behavioural and Health Sciences, Australian Catholic University, Strathfield, New South Wales, Australia. [103]Faculty of Psychology, University of Salamanca, Salamanca, Spain. [104]Psychological Assessment, University of Marburg, Marburg, Germany. [105]Department of Psychology and Cognitive Science, Cognitive Neuroscience Center, Adam Mickiewicz University, Poznan, Poland. [106]Social Psychology Laboratory, Paris Cité University, Paris, France. [107]Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome, Rome, Italy. [108]Institute of Cognitive Sciences and Technologies, Italian Research Council, Rome, Italy. [109]Department of Economic Psychology, Johannes Kepler University Linz, Linz, Austria. [110]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [111]Department of Psychology, Middle East Technical University, Ankara, Turkey. [112]School of Psychology and CICADA, Interdisciplinary Center for Data Science and Machine Learning, University of the Republic, Montevideo, Uruguay. [113]Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong, China. [114]Faculty of Engineering and Science, University of Adolfo Ibáñez, Viña del Mar, Chile. [115]Deparment of Psychology, Center for Cognition Research, University of Adolfo Ibáñez, Santiago, Chile. [116]Experimental Psychology, Justus Liebig University Giessen, Giessen, Germany. [117]Social and Affective Neuroscience Group, IMT School for Advanced Studies, Lucca, Italy. [118]Center for Social and Cognitive Neuroscience, School of Psychology, University of Adolfo Ibáñez, Santiago, Chile. [119]Department of Psychology, National University of Singapore, Singapore, Singapore. [120]Department of Psychology, Ashland University, Ashland, OH, USA. [121]Department of Human Development and Psychology, Tzu-Chi University, Haulien, Taiwan. [122]Department of Psychology, Cornell University, Ithaca, NY, USA. [123]Interacting Minds Centre, Aarhus University, Aarhus, Denmark. [124]Psychology and Cognitive Science Department, Adam Mickiewicz University, Poznań, Poland. [125]Basic Psychology, Psycholinguistics Research Line, CIPsi, School of Psychology, University of Minho, Braga, Portugal. [126]School of Education, Humanities and Social Sciences, Wawasan Open University, George Town, Malaysia. [127]School of Psychology, University of Galway, Galway, Ireland. [128]Psychology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada. [129]International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babeș-Bolyai University, Cluj-Napoca, Romania. [130]Department of Psychology, Wittenberg University, Springfield, OH, USA. [131]Department of Psychology, University of Oklahoma, Norman, OK, USA. [132]School of Health, University of the Sunshine Coast, Morton Bay, Queensland, Australia. [133]The School of Psychology, The University of Queensland Brisbane, Brisbane, Queensland, Australia. [134]Department of Psychology, Faculty of Arts and Sciences, Beijing Normal University at Zhuhai, Zhuhai, China. [135]Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing, China. [136]Ivey School, Western University, London, Ontario, Canada. [137]Department of Experimental Psychology, Ghent University, Ghent, Belgium. [138]Department of Psychology, Sapienza University of Rome, Rome, Italy. [139]Social Neuroscience Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy. [140]School of Human Sciences, University of Greenwich, London, UK. [141]Institute for Lifecourse Development, University of Greenwich, London, UK. [142]Faculty of Psychology, Ruhr University Bochum, Bochum, Germany. [143]Department of Psychology, University of Hong Kong, Hong Kong, China. [144]Department of Management, Technology, and Economics, ETH Zürich, Zurich, Switzerland. [145]Research and Innovation Management GmbH, Neumarkt an der Ybbs, Austria. [146]Department of Dynamic and Clinical Psychology and Health Studies, Sapienza University of Rome, Rome, Italy. [147]School of Psychology, University of Birmingham, Edgbaston, UK. [148]Department of Psychology, University of Concepción, Concepción, Chile. [149]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy. [150]Department of Psychology and Neuroscience, Boston College, Boston, MA, USA. [151]School of Social Sciences, Hellenic Open University, Patras, Greece. [152]Centre for Comparative Psycholinguistics, University of Alberta, Edmonton, Alberta, Canada. [153]Center for Sociocultural Research, National Research University Higher School of Economics, Moscow, Russian Federation. [154]Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada. [155]African Centre for Epistemology and Philosophy of Science, University of Johannesburg, Johannesburg, South Africa. [156]Marketing, ESSEC Business School, Cergy, France. [157]IMT School for Advanced Studies, Lucca, Italy. [158]Education Department, Henan University, Kaifeng, China. [159]Psychology, University of Tehran, Tehran, Iran. [160]Developmental Psychology, University of Amsterdam, Amsterdam, the Netherlands. [161]Faculty of Philosophy, Applied Memory Research Laboratory, Institute of Psychology, Doctoral School in the Social Sciences, Jagiellonian University, Krakow, Poland. [162]Division of Social Science, Hong Kong University of Science and Technology, Kowloon, China. [163]Department of Media and Communication, City University of Hong Kong, Kowloon, China. [164]Center for Liberal Arts, Fukuoka Institute of Technology, Fukuoka, Japan. [165]Faculty of Philosophy, Jagiellonian University, Krakow, Poland. [166]Department of Psychology, Japan Women's University, Tokyo, Japan. [167]George I.M. Georgescu Institute of Cardiovascular Diseases, Iași, Romania. [168]Alexandru Ioan Cuza University, Iași, Romania. [169]Faculty of English, Adam Mickiewicz University, Poznań, Poland. [170]Institute of Language Sciences, Shanghai International Studies University, Shanghai, China. [171]Key Laboratory of Language Science and Multilingual Artificial Intelligence, Shanghai International Studies University, Shanghai, China. [172]Department of Behavioral Sciences, York College, City University of New York, New York, NY, USA. [173]Department of Psychology, City University of New York, New York, NY, USA. [174]Cognitive Neuroscience Center, Adam Mickiewicz University, Poznań, Poland. [175]Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada. [176]Department of Physiology, Molecular, and Cellular Biology, University of Buenos Aires, Buenos Aires, Argentina. [177]Independent researcher, Amsterdam, the Netherlands. [178]School of Psychology, Dublin City University, Dublin, Ireland. [179]University of Zürich, Zurich, Switzerland. [180]School of Psychology Te Herenga Waka, Victoria University of Wellington, Wellington, New Zealand. [181]Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Vienna, Austria. [182]Department of Psychology, Faculty of Arts, Pavol Jozef Šafárik University, Košice, Slovakia. [183]School of Applied Psychology, Griffith University, Mt Gravatt, Queensland, Australia. [184]Department of Psychology, Sungkyunkwan University, Seoul, South Korea. [185]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea. [186]Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway. [187]Mohn Medical Imaging and Visualization Centre, Bergen, Norway. [188]IDN Being Human, Institute of Psychology, University of Wrocław, Wrocław, Poland. [189]International Research Team on Internet and Society, Faculty of Social Studies, Masaryk University, Brno, Czech Republic. [190]Department of Psychology, Senshu University, Kawasaki, Japan. [191]Faculty of Management, University of Warsaw, Warsaw, Poland. [192]Department of British and American Studies, Pavol Jozef Šafárik University, Košice, Slovakia. [193]Psychology Department, Aydın Adnan Menderes University, Aydın, Turkey. [194]Psychology Department, Durham University, Durham, UK. [195]Faculty of Psychology, SWPS University, Krakow, Poland. [196]ERI-Lectura and Department of Methodology, University of València, Valencia, Spain. [197]Nebrija Research Center in Cognition (CINC), Nebrija University, Madrid, Spain. [198]Centre for Research on Linguistics and Language Studies, Education University of Hong Kong, Tai Po, China. [199]Department of Psychology, Education University of Hong Kong, Hong Kong, China. [200]Centre for Psychosocial Health, The Education University of Hong Kong, Hong Kong, China. [201]Affective Physiology and Interoception Group, MoMiLab, IMT School for Advanced Studies, Lucca, Italy. [202]Cognitive Science, Occidental College, Los Angeles, CA, USA. [203]MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [204]Human Cognitive Neuroscience, University of Edinburgh, Edinburgh, UK. [205]Department of Psychology, Catholic University of Temuco, Temuco, Chile. [206]Center for Social and Cognitive Neuroscience, University of Adolfo Ibáñez, Valparaíso, Chile. [207]Department of Psychology, University of Navarra, Pamplona, Spain. [208]Psychology, University of Waterloo, Waterloo, Ontario, Canada. [209]Department of Education, Literatures, Intercultural Studies, Languages and Psychology, University of Florence, Florence, Italy. [210]Faculty of Psychology, Center for Studies in Human Neuroscience and Neuropsychology, Diego Portales University, Santiago, Chile. [211]School of Anthropology & Museum Ethnography, University of Oxford, Oxford, UK. [212]Institute of Biological Sciences, Federal University of Pará, Belém, Brazil. [213]School of Psychology, University of Auckland, Auckland, New Zealand. [214]Center for Brain Research, University of Auckland, Auckland, New Zealand. [215]School of Health, University of the Sunshine Coast, Petrie, Queensland, Australia. [216]Faculty of English, Adam Mickiewicz University, Poznan, Poland. [217]Department of Psychology, Alex Ekwueme Federal University, Ndufu-Alike, Nigeria. [218]Department of Psychology, Center for Brain, Biology, and Behavior, University of Nebraska–Lincoln, Lincoln, NE, USA. [219]Psychology Department, Alex Ekwueme Federal University, Ndufu-Alike, Nigeria. [220]Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, Bloomington, IN, USA. [221]School of Psychology, Manchester Metropolitan University, Manchester, UK. [222]School of Psychology, Social Work and Public Health (PSP), Oxford Brookes University, Oxford, UK. [223]Department of Psychology, Diego Portales University, Santiago, Chile. [224]Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, London, UK. [225]Institute of Psychology, University of Silesia in Katowice, Katowice, Poland. [226]Penn Center for Neuroaesthetics, University of Pennsylvania, Philadelphia, PA, USA. [227]Institute of Medical Psychology and Behavioral Neurobiology, University of Tuebingen, Tuebingen, Germany. [228]Psychology and Educational Science, Shahid Beheshti University, Tehran, Iran. [229]Faculty of Letters, Department of Psychology, Ege University, Izmir, Turkey. [230]Department of Methodology, University of València, Valencia, Spain. [231]Psychological Science, Central Connecticut State University, New Britain, CT, USA. [232]Faculty of Humanities and Social Sciences, City University of Macau, Macau, China. [233]Department of Psychology, University of Konstanz, Konstanz, Germany. [234]School of Psychology, University of Queensland, Brisbane, Queensland, Australia. [235]School of Law and Justice, University of Southern Queensland, Ipswich, Queensland, Australia. [236]Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy. [237]Cognitive Psychology Unit, IRCCS Mondino Foundation, Pavia, Italy. [238]Department of Psychology, University of Stirling, Stirling, UK. [239]Chair of Cognitive and Experimental Psychology, Institute of Psychology, RWTH Aachen University, Aachen, Germany. [240]Department of Philosophy, Macquarie University, Sydney, New South Wales, Australia. [241]Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany. [242]Foreign Language Education Department, Pamukkale University, Denizli, Turkey. [243]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [244]CIS-Iscte, University Institute of Lisbon, Lisbon, Portugal. [245]Department of Psychology, Kadir Has University, İstanbul, Turkey. [246]Faculty of Informatics, Kansai University, Takatsuki, Japan. [247]Department of Marketing, ESSEC Business School, Cergy, France. [248]Psychology Department, Bahçeşehir University, İstanbul, Turkey. [249]Institute of Substainability Psychology, Leuphana University Lüneburg, Lunenburg, Germany. [250]Institute of Psychology, RWTH Aachen University, Aachen, Germany. [251]Department of Psychology, University of Essex, Colchester, UK. [252]Institute of Psychology, University of Wrocław, Wrocław, Poland. [253]Department of Psychology, Bournemouth University, Bournemouth, UK. [254]School of Psychology, University of Birmingham, Birmingham, UK. [255]Division of Psychology, University of Chester, Chester, UK. [256]Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria. [257]Humanities Department, Icam School of Engineering, Toulouse, France. [258]Department of Anthropology, Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wrocław, Poland. [259]Center for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland. [260]Psychological Sciences Research Institute, Catholic University of Louvain, Louvain-la-Neuve, Belgium. [261]Department of Psychology, Oklahoma State University, Stillwater, OK, USA. [262]Institute of Psychology, Jagiellonian University, Krakow, Poland. [263]Department of Psychology, University of Maryland, College Park, MD, USA. [264]Philosophy, Pamukkale University, Denizli, Turkey. [265]Graduate School of Human-Environment Studies, Kyushu University, Fukuoka, Japan. [266]Department of Psychology, University of Oslo, Oslo, Norway. [267]School of Economics and Management, Beijing Jiaotong University, Beijing, China. [268]Department of Psychology, Federal University of Sergipe, Aracaju, Brazil. [269]Department of Psychology, University of Alabama—Tuscaloosa, Tuscaloosa, AL, USA. [270]Research Center in Cognitive Sciences, University of Talca, Talca, Chile. [271]Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece. [272]Psychology, University of Quebec at Trois-Rivières, Trois-Rivières, Quebec, Canada. [273]Institute for Sustainability Psychology, School of Sustainability, Leuphana University Lüneburg, Lunenburg, Germany. [274]Department of Cognitive Psychology, Trier University, Trier, Germany. [275]Department of Basic Psychology, Autonomous University of Madrid, Madrid, Spain. [276]Department of Psychology, Arizona State University, Tempe, AZ, USA. [277]Department of Psychology, Ithaca College, Ithaca, NY, USA. [278]LAPCOS, Côte d'Azur University, Nice, France. [279]Department of Psychological Sciences, Central Connecticut State University, New Britain, CT, USA. [280]Department of Psychology, Southern New Hampshire University, Manchester, NH, USA. [281]Institute of Sustainability Psychology, Leuphana University Lüneburg, Lunenburg, Germany. [282]School of Management, Wuhan University of Technology, Wuhan, China. [283]Darden School of Business, University of Virginia, Charlottesville, VA, USA. [284]Department of Psychology, Faculty of Health and Life Sciences, Northumbria University, Newcastle-upon-Tyne, UK. [285]Psychology Department, School of Philosophy & Language Sciences, University of Edinburgh, Edinburgh, UK. [286]Psychology Department, Center for Applied Behavioural Sciences, School of Social Sciences, Heriot-Watt University, Edinburgh, UK. [287]Department of Psychology, University of Michigan, Ann Arbor, MI, USA. [288]Faculty of Arts and Science, Kyushu University, Fukuoka, Japan. [289]Institute of Psychology and Behavior, Henan University, Kaifeng, China. [290]School of Psychology and Cognitive Science, East China Normal University, Shanghai, China. [291]Division of Social Science, Hong Kong University of Science and Technology, Hong Kong, China. [292]Department of Psychology, New York University, New York, NY, USA. [293]Department of

 https://doi.org/10.1038/s41562-025-02254-x

Psychology, Pamukkale University, Denizli, Turkey. [294]Network for Economic and Social Trends, Western University, London, Ontario, Canada. [295]Department of Psychology, Erzurum Technical University, Erzurum, Turkey. [296]Department of Psychology, Marmara University, İstanbul, Turkey. [297]Department of Psychology, Üsküdar University, İstanbul, Turkey. [298]Laboratory of Neurotechnology, Ural Federal University, Ekaterinburg, Russian Federation. [299]Department of Psychology, University of Crete, Rethymno, Greece. [300]Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands. [301]Centre for Sociocultural Research, National Research University Higher School of Economics, Moscow, Russian Federation. [302]Department of Psychology, George Fox University, Newberg, OR, USA. [303]Institute for Research and Development of Education, Faculty of Education, Charles University in Prague, Prague, Czech Republic. [304]Institute of Psychology, Czech Academy of Sciences, Brno, Czechia. [305]Southampton Business School, University of Southampton, Southampton, UK. [306]Institute of Facility Agriculture, Guangdong Academy of Agricultural Science, Guangzhou, China. [307]School of Business, Stevens Institute of Technology, Hoboken, NJ, USA. ✉e-mail: ebuchanan@harrisburgu.edu

nature portfolio

Corresponding author(s): Erin Buchanan

Last updated by author(s): May 5, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All code used for study creation and delivery, data processing, and analyses are available on OSF (https://osf.io/wrpj4/) and GitHub (https://github.com/SemanticPriming/SPAML). We used lab.js to collect the data (v22: https://lab.js.org/).
Data analysis	All code used for study creation and delivery, data processing, and analyses are available on OSF (https://osf.io/wrpj4/) and GitHub (https://github.com/SemanticPriming/SPAML). The semanticprimer R package (v0.0.2: https://doi.org/10.5281/zenodo.10698000) includes custom code created for this project. Other R packages used for analyses, project tracking dashboards, and stimuli creation included broom (v1.0.7), cluster (v2.1.6), codebook (v0.9.5), countrycode (v1.6.0), cowplot (v1.1.3), dendextend (v1.19.0), dplyr (v1.1.4), DT (v0.33), factoextra (v1.0.7), FactoMineR (v2.11), faux (v1.2.1), flextable (v0.9.7), forestplot (v3.1.6), ggplot2 (v3.5.2), ggrepel (v0.9.6), ggridges (v0.5.6), googlesheets4 (v1.1.1), here (v1.0.1), janitor (v2.2.0), jsonlite (v2.0.0), knitr (v1.50), labelled (v2.13.0), LexOPS (v0.3.1), lsa (v0.73.3), maps (v3.4.2.1), metafor (v4.6.0), moments (v0.14.1), MuMIn (v1.48.4), NCmisc (v1.2.0), nlme (v3.1.166), papaja (v0.1.3), parameters (v0.24.0), performance (v0.12.4), plyr (v1.8.9), psych (v2.5.3), purrr (v1.0.4), pwr (v1.3.0), quanteda (v4.1.0), rcanvas (v0.0.0.9001), readr (v2.1.5), readxl (v1.4.3), reshape (v0.8.9), rio (v1.2.3), RSQLite (v2.3.8), rvest (v1.0.4), semanticprimeR (v0.1.0), shiny (v1.10.0), shinydashboard (v0.7.3), stopwords (v2.3), stringdist (v0.9.12), stringi (v1.8.7), stringr (v1.5.1), sylly (v0.1.6), tenzing (v0.3.0), tibble (v3.2.1), tidyr (v1.3.1), tidytext (v0.4.2), tidyverse (v2.0.0), tm (v0.7.15), uaparserjs (v0.3.5), udpipe (v0.8.11), and widyr (v0.1.5).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Full SPAML Dataset (v1.0.2): This includes trial-level, participant-level, item-level, and priming-level data for 30 languages.
• GitHub: https://github.com/SemanticPriming/SPAML/releases/tag/v1.0.2

Semantic Priming Project: Used as a benchmark and comparison for this study.
• Available at: https://www.montana.edu/attmemlab/spp.html
• Original publication: Hutchison et al., 2013

English Lexicon Project: Used to inform sample size and estimate standard errors in lexical decision tasks.
• Available at: https://osf.io/n63s2/
• Original publication: Balota et al., 2007

OpenSubtitles Dataset: Used to compute word frequencies and generate word embeddings across languages.
• Now available at: https://opus.nlpl.eu/OpenSubtitles/corpus/version/OpenSubtitles
• Original publication: Lison & Tiedemann, 2016

subs2vec Embeddings: Pretrained fastText-style embeddings derived from OpenSubtitles (used for similarity calculations).
• Avaliable at: https://github.com/nicolasdupre/subs2vec
• Original publication: van Paridon & Thompson, 2021

Lexical Norms & Psycholinguistic Data (when available in the target language):
• Includes norms such as age of acquisition, valence, concreteness, and imageability.
• Integrated where possible via the semanticprimeR package: https://github.com/SemanticPriming/semanticprimeR

Supplementary Metadata & IRB Approvals:
• Labs, participant demographics, stimulus translation details, and ethical review information
• See supplemental materials with article for list of links

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender

The sample of participants self-identified as female (55.49%), male (37.39%), with the rest either missing data, not wanting to indicate their gender, or other. We use female, male, other, and prefer not to say because these were the English labels on the survey. We asked participants to indicate their gender. Current norms suggest we should have used woman and man instead. We report the labels that were on the survey.

No gender analyses were conducted, as these variables were not expected to impact the results.

Reporting on race, ethnicity, or other socially relevant groupings

Looking at the entire sample, participants indicated they had completed high school (42.77%), some college (7.63%), college (30.47%), a master's degree (9.30%), and other options (less than High School, Doctorate, or missing). Participants included in the analysis also followed this pattern: high school (46.02%), some college (8.34%), college (31.97%), and a master's degree (9.61%). College was used to indicate university-type experience (community college or otherwise). "Some college" indicated that they had not completed a degree but had completed some credits. Please note we use the terms here that were listed on the survey, but the terminology for education was localized to the data collection area.

The average self-reported age for all participants was M = 31.4 years (SD = 15.0), ranging from 18 to 104 years (Mdn = 24, IQR = 20 – 39).

Population characteristics

See above.

Recruitment

Labs recruited participants using their own local resources, which varied by site and country. Recruitment methods included distribution through social media, course participation for academic credit, and paid platforms such as MTurk, Prolific, and Respondi. A detailed table of recruitment strategies by site—including platform used, compensation method, testing modality (in-person or online), and sampling context—is available in the supplementary materials (see https://osf.io/ty4hp).

Because recruitment strategies were decentralized, there is potential for self-selection bias, especially in online, unpaid, or convenience samples where participants with higher interest in language or psychology research may be overrepresented. In paid samples, participant pools may skew toward individuals with greater digital literacy or those motivated by financial incentives. Additionally, demographic data (e.g., education level, native language, age) varied widely across labs and geopolitical regions, potentially contributing to non-random sampling variance across languages.

These sources of bias are unlikely to impact within-language priming effects, which rely on within-subject comparisons.

However, they may introduce noise or systematic differences in between-language comparisons, especially if cultural, educational, or platform-related factors correlate with response latencies or task engagement. The large sample size, preregistered exclusion criteria, and consistency in task design across sites help mitigate these concerns.

Ethics oversight	Harrisburg University of Science and Technology IRB File No. 20211110 https://osf.io/sg5ac

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Therefore, data collection was a repeated-measures design in which participants did not see all of the possible stimuli, but did see all the possible conditions (related, unrelated, and nonword pairs). Participants were blinded to condition, and the explicit link between pairs was not explained to participants. Data are quantitative.

Research sample

The following statistics are calculated by session, which generally represents one participant; however, participants could have taken the study multiple times. We will describe these sessions as participants for ease of reading. We present the full sample information and the analyzed sample information to demonstrate that the data analyzed are similar to the full dataset. The sample of participants self-identified as female (55.49%), male (37.39%), with the rest either missing data, not wanting to indicate their gender, or other. We use female, male, other, and prefer not to say because these were the English labels on the survey. We asked participants to indicate their gender. Current norms suggest we should have used woman and man instead. We report the labels that were on the survey. If the data were filtered to select only participants that were included in the analysis, the participants self-identified as predominantly female (60.95%) or male (37.44%). Looking at the entire sample, participants indicated they had completed high school (42.77%), some college (7.63%), college (30.47%), a master's degree (9.30%), and other options (less than High School, Doctorate, or missing). Participants included in the analysis also followed this pattern: high school (46.02%), some college (8.34%), college (31.97%), and a master's degree (9.61%). College was used to indicate university-type experience (community college or otherwise). "Some college" indicated that they had not completed a degree but had completed some credits. Please note we use the terms here that were listed on the survey, but the terminology for education was localized to the data collection area. Please see https://osf.io/vdgkr for the full participant information.
Full language percent tables can be found in the supplementary materials (https://osf.io/ta6wf, https://osf.io/652h8, Table S1). The data indicates that the pattern of native languages was similar in the full data and data used for analysis. The average self-reported age for all participants was M = 31.4 years (SD = 15.0), ranging from 18 to 104 years (Mdn = 24, IQR = 20 – 39). In the demographic questions, we asked the participants to enter their year of birth, and the high maximum values likely belonged to participants who entered the minimum possible year allowable in the data collection form. The data of the participants included in the analysis showed the same age pattern: M = 30.4 (SD = 14.2) ranging from 18 to 104 (Mdn = 24, IQR = 20 – 37).

Sampling strategy

Sampling strategy was generally convenience sampling within each lab.

For our power analysis, we first detail the background on how we estimated sample size, explain accuracy in parameter estimation, provide two simulations based on previous research, and the final proposed sample size. We end this section by specifying why this procedure was superior to previous methods and the requirements for publication.
Background
One concern is how to estimate the sample size required for cue-target pairs, as the previous literature indicates variability in their results40. Sample sizes of N = 30 per study have often been used in an attempt to at least meet some perceived minimum criteria for the central limit theorem. We focused on the lexical decision task for our procedure, wherein participants are simply asked if a concept presented to them is a word (e.g., CAT) or nonword (e.g., GAT). The dependent variable in this study was response latency, and we used lexical decision data from the English Lexicon Project22 and the Semantic Priming Project21 to estimate the minimum sample size necessary for each item, as previous research has suggested an overall sample size may lead to unreliability in the item-level responses40. The English Lexicon Project contains lexical decision task data for over 40,000 words, while the Semantic Priming Project includes 1,661 target words.
Accuracy in parameter estimation (AIPE)
AIPE description. In this approach, one selects a minimum sample size, a stopping rule, and a maximum sample size. A minimum sample size was defined for all items based on data simulation below. For the stopping rule, we focused on finding a confidence interval around a parameter that would be "sufficiently narrow"50,51,71. These parameters are often tied to the statistical test or effect size for the study, such as correlation or contrast between two groups. In this study, we paired accuracy in parameter estimation with a sequential testing procedure to adequately sample each item, rather than estimate an overall effect size. Therefore, we used the previous lexical decision data to determine our sufficiently narrow confidence by finding a generalized standard error one should expect for well measured items. After the minimum sample size, each item's standard error was assessed to determine if the item had met the goals for accuracy in parameter estimation as our stopping rule. If so, the item was sampled at a lower probability in relation to other items until all items reach the accuracy goals or a maximum sample size determined by our simulations below (https://osf.io/v2y9e).
Estimates from the English Lexicon Project. First, the response latency data for the English Lexicon Project were z-scored by participant and session as each participant has a somewhat arbitrary average response latency53. The data were then subset for only real word trials that were correctly answered. The average sample size before removing incorrect answers was 32.69 (SD = 0.63)

participants with an average retention rate of 84% and 27.41 (SD = 6.43) participants after exclusions. The retention rates were skewed due to the large number of infrequent words in the English Lexicon Project, and we used the median retention rate of 91% for later sample size estimations. The median standard error for response latencies in the English Lexicon Project was 0.14, and the mean was 0.16. Because the retention rates were variable across items, we also calculated the average standard error for items that retained at least 30 participants at 0.12. This standard error rate represented the potential stopping rule.
The data were then sampled with replacement to determine the sample size that would provide that standard error value. One hundred words within the data were randomly selected, and samples starting at n = 5 to n = 200 were selected (increasing in units of five). The standard error for each of these samples was then calculated for the simulation, and the percent of samples with standard errors at or less than the estimated population value was then tabulated. In order to achieve 80% of items at or below the proposed standard error, we needed approximately 50 participants per word. This value was used as our minimum sample size for a lexical decision task, and the accuracy standard error level was preliminarily set at 0.12.
Estimates from the Semantic Priming Project. This same procedure was examined with the Semantic Priming Project's lexical decision data on real word trials. The priming response latencies were expected to be variable, as this priming strength should be predicted by other psycholinguistic variables, such as word relatedness. Therefore, we aimed to achieve an accurate representation of lexical decision times, from which priming could then be calculated. However, it should be noted that accurately measured response latencies do not necessarily imply "reliable" priming or difference score data72, but larger sample sizes should provide more evidence of the picture of item-level reliability. We used these data paired with the English Lexicon Project to account for the differences in a lexical decision only versus priming focused task. The average standard error in the Semantic Priming Project was less at 0.06, likely for two reasons: the data in the Semantic Priming Project are generally frequent nouns and only 1,661 concepts, as compared to the 40,000 in the English Lexicon Project. The retention rate for the Semantic Priming Project was less skewed than the English Lexicon Project at a median of 97% and mean of 96%. Using the same sampling procedure, we estimated sample sizes of n = 5 to n = 400 participants increasing by units of 5. In this scenario, we found the maximum sample size of 320 participants for 80% of the items to reach the smaller standard error of 0.06. Therefore, we used 320 as our maximum sample size, and the average of the two standard errors found as our stopping rule, i.e., 0.09.
Final sample size. Given our minimum, maximum, and stopping rule, we then estimated the final sample size per language based on study design characteristics. Participants completed approximately 800 lexical decision trials per session, and each participant only completed 150 of these concepts (75 targets in the related condition, 75 targets in the unrelated condition; cue words were not analyzed) that were the target of this sample size analysis (see below for more details on trial composition). Therefore, the target number of items (n = 1000 concepts) was multiplied by the minimum/maximum sample size, and conditions (related word pair versus unrelated word pair) and divided by the total number of critical lexical decision trials per participant times the data retention rate (a conservative estimate of 90%). The final estimate for sample size per language was 741 to 4741 [(1000*50*2) / (150*.90); (1000*320*2) / 150*.90]. The complete code and description of this process are detailed in our supplemental documents (https://osf.io/rxgkf, https://osf.io/v2y9e).
This sample size estimation represents a major improvement from previous database collection studies, as many have used the traditional N = 30 to guess at minimum sample size. Because the variability of the sample size was quite large, we employed a stopping procedure to ensure participant time and effort were maximized, and data collection was optimized. To summarize, the minimum sample size was 50 participants per word and the maximum for the adaptive procedure was 320, which results in 741 to 4741 participants per language based on expected usable trials. Therefore, the total sample size was proposed to be 7410 to 47410 participants for ten languages. After 50 participants who answered a real word item, each concept was examined for standard error, and data collection for that concept was decreased in probability when the standard error reached our average criterion of 0.09. Item probability for selection was also decreased when they reached the maximum proposed sample size (n = 320). This process was automated online and checked in a scheduled subroutine.

Data collection

Data collection was online via computer with a keyboard (as the spacebar was required for the study). The researcher was sometimes in the room with participants (for labs that collected data in person) but generally participants could complete the study on their own. The corresponding author knew what conditions word pairs were in, but other researchers and participants were blinded to the conditions.

Timing

Start: 2022-08-03 03:32:13 UTC
Stop: 2024-02-14 15:27:48 UTC

Data exclusions

Preregistered criterion:

Exclusion summary
Data were excluded for the following reasons in this order (per the pre-registered plan):
1) Participant-level data: the entire participant's data were removed from the analyses if:
a. A participant did not indicate at least 18 years of age.
b. A participant did not complete at least 100 trials.
c. A participant did not achieve 80% correct.
2) Trial level data: individual trials were removed from the analyses in the following instances:
a. Timeout trials (i.e., no response given in 3 s window). This value was chosen to ensure that the experiment was completed in under 30 minutes on average, while giving an appropriate amount of time in a lexical decision study to answer (using the Semantic Priming Project as rubric for general trial length).
b. Incorrectly answered trials.
c. Response latencies shorter than 160 ms52.
3) Trial level exclusions dependent on test: Participant sessions were Z-scored as described below, and trials were marked for exclusion in the dataset. Each analysis was tested with the full data and then without these values:
a. Response latencies over the absolute value of Z = 2.5.
b. Response latencies over the absolute value of Z = 3.0.

35,904 participants opened the study link, with 31,645 participants proceeding to complete at least one study trial (i.e., past the practice trials). Of these participants, 26,971 were retained for analysis because they met our three participant-level inclusion criteria. The analyses reported in the study examine only those languages that met the minimum data criteria, which includes 32,897 total participants, 29,155 of whom completed at least one trial, 25,163 met the strict inclusion criteria.

Non-participation	We were unable to record the number of participants who declined participation outright, as individuals could close the study before any data were collected. However, among those who opened the study link, 4,259 participants (11.9%) did not complete at least one real (non-practice) trial and were therefore excluded from all analyses. These cases are considered incomplete responses and likely reflect dropout or disengagement prior to beginning the main task.
Randomization	At the start of data collection, all presented items were randomly selected from the larger item pool by equalizing the probability of inclusion for all words and nonwords (p = 1/1000 concepts). After the minimum sample size was collected, each word's standard error was checked to determine if the sample size for that item had reached our accuracy criteria. If so, the probability of sampling that item was decreased by half. Once a concept has reached the maximum required sample size, the probability of sampling was also be decreased by half. This procedure allowed for random sampling of the items that still need participants without eliminating words from the item pool. Therefore, we ensured that there were always words to randomly select from (i.e., to keep the same procedure and number of trials for all participants) and that the randomization was a sampled mix of words that reach accuracy quickly and words that need more participants (i.e., participants do not only see the unusual words at the end of data collection). Once all words reached the stopping criteria or maximum sample size, the probabilities were equalized. We set minimum, maximum, and a stopping rule for the initial data collection; however, we allowed data collection after these were reached and will post updates to the data using GitHub releases (modeled after the Small World of Words Project33, which is ongoing). All data were included in our dataset, and the analysis section describes how we indicated exclusion criteria.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*