Measuring the Semantic Priming Effect Across Many Languages

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. M. Collins[15], David C. Vaidis[16,17], Balazs Aczel[18], Ali H. Al-Hoorie[19], Ettore Ambrosini[8], Théo Besson[17], Debora I. Burin[20,21], Muhammad M. Butt[22], A. J. Benjamin Clarke[23], Yalda Daryani[24], Dina A. S. El-Dakhs [25], Mahmoud M. Elsherif[26,27], Maria Fernández-López[28], Paulo R. S. Ferreira[29], Raquel M. K. Freitag[30], Carolina A. Gattei[20,31], Hendrik Godbersen[32], Philip A. Grim II[1], Peter Halama[33], Patrik Havan[33], Natalia C. Irrazabal[21], Chris Isloi[34], Rebecca K. Iversen[35], Yoann Julliard[36,37], Aslan Department of Psychology, IIzmir Katip Celebi University, Izmir, Türkiye[38,39], Michal Kohút[40], Veronika Kohútová[40], Julija Kos[41], Alexandra I. Kosachenko[42], Tiago J. S. d. Lima[43], Matthew HC Mak[44], Christina Manouilidou[41], Leonardo A. Marciaga[45], Xiaolin M. Melinna[46], Jacob F. Miranda[47], Coby Morvinski[48], Aishwarya Muppoor[15], F. Elif Müjdeci[49], Yngwie A. Nielsen[50], Juan C. Oliveros[51], Jaš Onič[41], Marietta Papadatou-Pastou[52], Ishani Patel[15], Zoran Pavlović[53], Blaž Pažon[41], Gerit Pfuhl[54,55], Ekaterina Pronizius[56], Timo B. Roettger[35], Camilo R. Ronderos[35], Susana Ruiz-Fernandez[57], Magdalena Senderecka[14], Çağlar Solak[58], Anna Stückler[50], Raluca D. Szekely-Copîndean[59,60], Analí R. Taboh[20,31], Rémi Thériault[61], Ulrich S. Tran[56], Fabio Trecca[50], José Luis Ulloa[62], Marton A. Varga[18], Steven Verheyen[63], Tijana Vesić Pavlović[53], Giada Viviani[64], Nan Wang[65], Kristyna Zivna[66], Chen C. Yun[67], Oliver J. Clark[68], Oguz A. Acar[69], Matúš Adamkovič[33,70,71], Giulia Agnoletti[31,72], Atakan M. Akil[18,73], Zainab Alsuhaibani[74], Simona Amenta[75], Olga A. Ananyeva[76], Michael Andreychik[77], Bernhard Angele[78,79], Danna C. Arias Quiñones[80], Nwadiogo C. Arinze[81], Adrian D. Askelund[82,83], Bradley J. Baker[84], Ernest Baskin[85], Luisa Batalha[86], Carlota Batres[80], Maria S. Beato[87], Manuel Becker[88], Maja Becker[16], Maciej Behnke[89], Christophe Blaison[17], Anna M. Borghi[90,91], Eduard Brandstätter[92], Jacek Buczny[93], Nesrin Budak[94], Álvaro Cabana[95], Zhenguang G. Cai[65], Enrique C. Canessa[96], Ignacio Castillejo[97], Müge

Cavdan[98], Luca Cecchetti[99], Sergio E. Chaigneau[96], Feria X. W. Chang[100], Christopher R. Chartier[101], Sau-Chin Chen[102], Elena Cherniaeva[76], Morten H. Christiansen[50,103], Hu Chuan-Peng[67], Patrycja Chwiłkowska[89], Montserrat Comesaña[104], Chin Wen Cong[105], Casey Cowan[106], Stéphane D. Dandeneau[107], Oana A. David[60], William E. Davis[108], Elif G. Demirag Burak[109], Barnaby J. W. Dixson[110,111], Hongfei Du[112,113], Rod Duclos[114], Wouter Duyck[115], Liudmila A. Efimova[76], Ciara Egan[106], Vanessa Era[90,116], Thomas R. Evans[117], Anna Exner[118], Gilad Feldman[119], Katharina Fellnhofer[120,121], Chiara Fini[90], Sarah E. Fisher[101], Heather D. Flowe[26], Patricia Garrido-Vásquez[122], Daniele Gatti[123], Jason Geller [124], Vaitsa Giannouli[125], Anna S. Gorokhova[76], Lindsay M. Griener[126], Dmitry Grigoryev[76], Igor Grossmann[127], Mohammadhesam Hajighasemi[128], Giacomo Handjaras[99], Cathy Hauspie[115], Zhiran He[129], Renata M. Heilman[60], Amirmahdi Heydari[24], Alanna M. Hine[106], Karlijn Hoyer[130], Weronika Hryniszak[14], Janet H.-w. Hsiao[131], Guanxiong Huang[132], Keiko Ihaya[133], Ewa Ilczuk[14], Tatsunori Ishii[134], Andrei Dumbravă[135,136], Katarzyna Jankowiak[89], Xiaoming Jiang[137], David C. Johnson[138], Rafał Jończyk[89], Juhani Järvikivi[126], Laura Kaczer[20], Kevin L. Kamermans[2], Johannes A. Karl[139], Alexander Karner[56], Pavol Kačmár[140], Jacob J. Keech[141], M. Justin Kim[142,143], Max Korbmacher[144,145], Kathrin Kostorz[56], Marta Kowal[146], Tomas Kratochvil[147], Yoshihiko Kunisato[148], Anna O. Kuzminska[149], Lívia Körtvélyessy[140], Fatma E. Köse[150,151], Massimo Köster[88], Magdalena Kękuś[152], Melanie Labusch[28,153], Claus Lamm[56], Chaak Ming Lau[154], Julieta Laurino[20], Wilbert Law[154], Giada Lettieri[99], Carmel A. Levitan[155], Jackson G. Lu[156], Sarah E. MacPherson[157], Klara Malinakova[66], Diego Manriquez-Robles[158], Nicolás Marchant[96], Marco Marelli[75], Martín Martínez[159], Molly F. Matthews[127], Alan D. A. Mattiassi[160], Josefina Mattoli-Sánchez[161], Claudia Mazzuca[90], David P. McGovern[139], Zdenek Meier[66], Filip Melinscak[56], Michal Misiak[146,162], Luis C. P. Monteiro[163], David Moreau[164], Sebastian Moreno[96], Kate E. Mulgrew[110], Dominique Muller[36,37,165], Tamás Nagy[18], Marcin Naranowicz[89], Izuchukwu L. G. Ndukaihe[81], Maital Neta[166], Lukas Novak[66], Chisom E. Ogbonnaya[81], Jessica Jee Won Paek[167], Aspasia E. Paltoglou[68,168], Francisco J. Parada[161], Adam J. Parker[169], Mariola Paruzel-

Czachura[170,171], Yuri G. Pavlov[172], Saeed Paydarfard[173], Dominik Pegler[56], Mehmet Peker[39], Manuel Perea[28,153], Stefan Pfattheicher[50], John Protzko[174], Irina S. Prusova[76], Katarzyna Pypno-Blajda[170], Zhuang Qiu[65,175], Ulf-Dietrich Reips[176], Gianni Ribeiro[177,178], Luca Rinaldi[123,179], S. Craig Roberts[146,180], Tanja C. Roembke[181], Marina O. Romanova[76], Robert M. Ross[182], Jan P. Röer[183], Filiz Rızaoğlu[184], Toni T. Saari[185], Erika Sampaolo[99], Anabela Caetano Santos[13], F. Çağlar Sarıçiçek[186], Kyoshiro Sasaki[187], Frank Scharnowski[56], Kathleen Schmidt[101], Amir Sepehri[128], Halid O. Serçe[188], A. T. Sevincer[189], Cynthia S. Q. Siew[100], Matilde E. Simonetti[181], Miroslav Sirota[190], Agnieszka Sorokowska[146], Piotr Sorokowski[146], Ian D. Stephen[78], Laura M. Stevens[26], Suzanne L. K. Stewart[191], David Steyrl[56], Stefan Stieger[192], Anna Studzinska[193], Mar Suarez[87], Anna Szala[194,195], Arnaud Szmalec[115,196], Daniel Sznycer[197], Ewa Szumowska[14,198], Sinem Söylemez[58], Bahadır Söylemez[184], Kaito Takashima[199], Christian K. Tamnes[35], Joel C. R. Tan[100], Chengxiang Tang[200], Peter Tavel[66], Julian Tejada[30], Benjamin C. Thompson[201], Jake G. Tiernan[139], Vicente Torres-Muñoz[62], Anna K. Touloumakos[202], Bastien Trémolière[16,203], Monika Tschense[189], Belgüzar N. Türkan[204], Miguel A. Vadillo[97], Caterina Vannucci[99], Michael E. W. Varnum[205], Martin R. Vasilev[169], Leigh Ann Vaughn[206], Fanny Verkampt[207], Liliana M. Villar[174,208], Sebastian Wallot[189], Lijun Wang[209], Ke Wang[210], Glenn P. Williams[211], David Willinger[192], Kelly Wolfe[157,212], Alexandra S. Wormley[205,213], Yuki Yamada[199], Yunkai Yang[129], YUWEI ZHOU[154], Mengfan Zhang[56], Wang Zheng[214], Yueyuan Zheng[131], Chenghao Zhou[215], Radka Zidkova[66], Nina M. Zumbrunn[139], Ogeday Çoker[184], Sami Çoksan[114,216], Sezin Öner[186], Asil A. Özdoğru[217,218], Seda M. Şahin[94], Dauren Kasanov[42], Alexios Arvanitis[219], Cameron Brick[130], Melissa F. Colloff[26], Albina Gallyamova[76], Christopher Koch[220], Ivan Ropovik[221,222], Yucheng C. Zhang[223], Xingxing Zhou[224], Sneh Patel[15], Jordan W. Suchow[225], & Savannah C. Lewis[101,201]

[1] Harrisburg University of Science and Technology

[2] Independent Researcher

[3] Leiden University

[4] Aalborg University

[5] Stanford University

[6] Christ University

[7] University of Exeter

[8] University of Padova

[9] Midwestern State University

[10] Goethe University Frankfurt

[11] Massachusetts General Hospital

[12] Harvard Medical School

[13] University Institute of Lisbon

[14] Jagiellonian University

[15] Nova Southeastern University

[16] University of Toulouse

[17] Paris Cité University

[18] ELTE Eötvös Loránd University

[19] Royal Commission for Jubail and Yanbu

[20] University of Buenos Aires

[21] University of Palermo

[22] Government College University

[23] Thammasat University

[24] University of Tehran

[25] Prince Sultan University

[26] University of Birmingham

[27] University of Leicester

[28] University of València

[29] Federal University of Uberlândia

[30] Federal University of Sergipe

[31] Torcuato Di Tella University

[32] FOM University of Applied Sciences

[33] Slovak Academy of Sciences

[34] Unaffiliated Researcher

[35] University of Oslo

[36] University of Grenoble Alpes

[37] University Savoie Mont Blanc

[38] Izmir Katip Celebi University

[39] Ege University

[40] University of Trnava

[41] University of Ljubljana

[42] Ural Federal University

[43] University of Brasília

[44] University of Warwick

[45] Illinois Institute of Technology

[46] The Hong Kong Polytechnic University

[47] California State University - East Bay

[48] Ben-Gurion University of the Negev

[49] Bilkent University

[50] Aarhus University

[51] Catholic University of Maule

[52] National and Kapodistrian University of Athens

[53] University of Belgrade

[54] UiT The Arctic University of Norway

[55] Norwegian University of Science and Technology

[56] University of Vienna

[57] Brandenburg University of Technology Cottbus-Senftenberg

[58] Manisa Celal Bayar University

[59] Romanian Academy

[60] Babeș-Bolyai University

[61] University of Quebec in Montreal

[62] University of Talca

[63] Erasmus University Rotterdam

[64] University of Padua

[65] The Chinese University of Hong Kong

[66] Palacký University Olomouc

[67] Nanjing Normal University

[68] Manchester Metropolitan University

[69] King's College London

[70] Charles University

[71] University of Jyväskylä

[72] Universidad Favaloro

[73] University of Pécs

[74] Imam Mohammad Ibn Saud Islamic University

[75] University of Milano-Bicocca

[76] National Research University Higher School of Economics

[77] Fairfield University

[78] Bournemouth University

[79] Antonio de Nebrija University

[80] Franklin and Marshall College

[81] Alex Ekwueme Federal University

[82] Lovisenberg Diaconal Hospital

[83] Norwegian Institute of Public Health

[84] Temple University

[85] Saint Joseph's University

[86] Australian Catholic University

[87] University of Salamanca

[88] University of Marburg

[89] Adam Mickiewicz University

[90] Sapienza University of Rome

[91] Italian Research Council

[92] Johannes Kepler University Linz

[93] Vrije Universiteit Amsterdam

[94] Middle East Technical University

[95] University of the Republic

[96] University of Adolfo Ibáñez

[97] Autonomous University of Madrid

[98] Justus Liebig University Giessen

[99] IMT School for Advanced Studies

[100] National University of Singapore

[101] Ashland University

[102] Tzu-Chi University

[103] Cornell University

[104] University of Minho

[105] Wawasan Open University

[106] University of Galway

[107] Memorial University of Newfoundland

[108] Wittenberg University

[109] University of Oklahoma

[110] University of the Sunshine Coast

[111] The University of Queensland Brisbane

[112] Beijing Normal University at Zhuhai

[113] Beijing Normal University

[114] Western University

[115] Ghent University

[116] IRCCS Santa Lucia Foundation

[117] University of Greenwich

[118] Ruhr University Bochum

[119] University of Hong Kong

[120] ETH Zürich

[121] Research and Innovation Management GmbH

[122] University of Concepción

[123] University of Pavia

[124] Boston College

[125] Hellenic Open University

[126] University of Alberta

[127] University of Waterloo

[128] ESSEC Business School

[129] Henan University

[130] University of Amsterdam

[131] Hong Kong University of Science and Technology

[132] City University of Hong Kong

[133] Fukuoka Institute of Technology

[134] Japan Women's University

[135] George I.M. Georgescu Institute of Cardiovascular Diseases

[136] Alexandru Ioan Cuza University

[137] Shanghai International Studies University

[138] City University of New York

[139] Dublin City University

[140] Pavol Jozef Šafárik University

[141] Griffith University

[142] Sungkyunkwan University

[143] Institute for Basic Science

[144] Western Norway University of Applied Sciences

[145] Mohn Medical Imaging and Visualization Centre

[146] University of Wrocław

[147] Masaryk University

[148] Senshu University

[149] University of Warsaw

[150] Aydın Adnan Menderes University

[151] Durham University

[152] SWPS University

[153] Nebrija University

[154] The Education University of Hong Kong

[155] Occidental College

[156] Massachusetts Institute of Technology

[157] University of Edinburgh

[158] Catholic University of Temuco

[159] University of Navarra

[160] University of Florence

[161] Diego Portales University

[162] University of Oxford

[163] Federal University of Pará

[164] University of Auckland

[165] University Institute of France

[166] University of Nebraska-Lincoln

[167] Indiana University

[168] Oxford Brookes University

[169] University College London

[170] University of Silesia

[171] University of Pennsylvania

[172] University of Tuebingen

[173] Shahid Beheshti University

[174] Central Connecticut State University

[175] City University of Macau

[176] University of Konstanz

[177] The University of Queensland

[178] University of Southern Queensland

[179] IRCCS Mondino Foundation

[180] University of Stirling

[181] RWTH Aachen University

[182] Macquarie University

[183] Witten/Herdecke University

[184] Pamukkale University

[185] University of Helsinki

[186] Kadir Has University

[187] Kansai University

[188] Bahçeşehir University

[189] Leuphana University Lüneburg

[190] University of Essex

[191] University of Chester

[192] Karl Landsteiner University of Health Sciences

[193] Icam School of Engineering

[194] Polish Academy of Sciences

[195] Nicolaus Copernicus University

[196] Catholic University of Louvain

[197] Oklahoma State University

[198] University of Maryland

[199] Kyushu University

[200] Beijing Jiaotong University

[201] The University of Alabama - Tuscaloosa

[202] Panteion University of Social and Political Science

[203] University of Quebec at Trois-Rivières

[204] Trier University

[205] Arizona State University

[206] Ithaca College

[207] Côte d’Azur University

[208] Southern New Hampshire University

[209] Wuhan University of Technology

[210] University of Virginia

[211] Northumbria University

[212] Heriot-Watt University

[213] University of Michigan

[214] East China Normal University

[215] New York University

[216] Erzurum Technical University

[217] Marmara University

[218] Üsküdar University

[219] University of Crete

[220] George Fox University

[221] Charles University in Prague

[222] Czech Academy of Sciences

[223] University of Southampton

[224] Guangdong Academy of Agricultural Science

[225] Stevens Institute of Technology

Corresponding author: Erin M. Buchanan, ebuchanan@harrisburgu.edu

Abstract

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 analyzed) by creating the largest available database of semantic priming values based on 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 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 July 15, 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, 5/31/2022). Since OSF has updated their system and old files are no longer viewable with the proper time stamps (see https://osf.io/en8ur), we point to the GitHub file that is time stamped appropriately: https://github.com/SemanticPriming/SPAML/blob/736f846b973cea8c994e6aa958b0df9b5d636c3d/07_Manuscript/SPAML_RR_NHB_V4.docx or the pdf format at https://github.com/SemanticPriming/SPAML/blob/79fbad2ef9ac357a55ac1113722f32b82330540b/07_Manuscript/SPAML_RR_NHB_V4.pdf after acceptance (all time stamped before data collection began).

Measuring the Semantic Priming Effect Across Many Languages

Semantic priming is a well-studied cognitive phenomenon whereby participants are shown a cue word (e.g., DOG) followed by either a semantically related (e.g., CAT) or unrelated (e.g., BUS) target word[1]. Semantic priming is defined as the decrease in response latency (i.e., 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 of attention[7,8], studies of bi/multilingual people[9,10], on neurodivergent individuals such as those affected by Parkinson's disease, aphasia, or schizophrenia, and 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 (e.g., frequency, familiarity, 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 significant 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 (e.g., 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. 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 (e.g., DOG-CAT) and unrelated (e.g., 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 (i.e., number of shared features between words)[30–32], association strength (i.e., the probability of a first word eliciting a second word when simply shown the first word)[33,34], or text co-occurrence (i.e., 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 nonwords, 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 (i.e., lexical decision or naming speed) for the related word-pair conditions compared to unrelated word-pair conditions. The proposed 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 Hutchison et al.'s[39] study, 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 only contains 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 challenging, as large-scale data in this area are sparse.

Although it is challenging, using newer computational techniques[44,45] and recently published corpora[20,46], a broader coverage dataset in up to 43 languages is possible. Therefore, this study aims to provide data that complement and extend the published data, which would 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 will address the proposed limitations by increasing sample size to hopefully improve reliability and expanding beyond the English language within the same target stimuli. From these openly shared data, two research questions will be assessed as detailed in Table 1:

1) Is semantic priming a non-zero effect? To assess this research question, we will examine the confidence interval of the semantic priming effect to determine if the lower limit of the confidence interval is greater than zero using an intercept-only regression model estimating across all languages. Therefore, we predict semantic facilitation with reduced response latencies for related word-pair conditions in comparison to unrelated word-pair conditions.
2) Does the semantic priming effect vary across languages when examining the same target stimuli? We will add a random intercept of language to the model estimated in Hypothesis 1 to estimate the variability of priming across languages. We will conclude there is variability between priming effects for languages when the AIC for the random-intercept model is two or more points less than the AIC for the model in Hypothesis 1[47]. To contextualize these results, we will provide a forest plot of the priming effects for languages to demonstrate the pattern of variability. For Hypothesis 2, we do not specify predicted directions for the effects but do expect potential variability in priming effects across languages. It is logical to expect differences in language due to culture, orthography, alphabet, etc., 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 will use 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 will be 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, stimuli 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. Thus, differences in priming can 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, along with identification of the languages that reached the pre-registered minimum sample size. Next, the research labs and ethics involved in the project are discussed. We then detail the exclusion criteria from the pre-registered 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 pre-registered descriptives in the supplementary materials.

Languages

Forty-three languages were originally identified for possible data collection based on 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 pre-registered 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 3 provides a summary of the data collection for each language with respect to the number of included participants (based on the pre-registered data inclusion rules), the number of participants excluded, the proportion of correct answers for participants included (i.e., participant accuracy scores were calculated, and then the average of participant accuracy scores for each language were calculated), and the median completion time for included participants in minutes (https://osf.io/bqpk2). A complete breakdown of gender, education, age, and stimuli completion can be found in the supplementary materials (https://osf.io/y3dk7). The following 19 languages met the minimum data collection requirements and will be analyzed in this manuscript: 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%; 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 pre-registered 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).

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 materials provide links to the IRB 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 Figure 3 for a visualization of the entire sample during data collection.

[Figure 3]

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 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 analyzed dataset. 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 pre-registered plan calculated accuracy as $\frac{N_{Correct}}{N_{Trials\ Seen}}$ in the planned scripts; however, an administrative team discussion revealed that the pre-registered report's definition of accuracy could alternatively be interpreted as $\frac{N_{Correct}}{N_{Answered}}$. 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 $\frac{N_{Correct}}{N_{Trials\ Seen}}$ for analysis, while an analysis using the rescored accuracy $\frac{N_{Correct}}{N_{Answered}}$ can be found in the supplementary materials. The analyses reported below 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, 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 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).

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%) based on browser meta-data. 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 includes 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 (8.18%), 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 online data (https://osf.io/93kep, https://osf.io/3yab7, Table S1). 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 compared to 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% to 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 minutes (*SD* = 303.61). If a participant's computer went to sleep during the study, and they later returned to it (e.g., to close the browser), the last timestamp would include the final time the study was open. Therefore, the median completion time is likely more representative, *Mdn* = 17.88 minutes. The participants included in the analysis completed the study in 24.14 minutes on average (*SD* = 296.83, *Mdn* = 17.97 minutes).

Trial-level data

Each language was saved in separate files in the online materials. Supplementary files (https://osf.io/q7e35, https://osf.io/dmc6u) and code within *semanticprimeR* (https://osf.io/yd8u4) enable merging trials across concepts and pairings (e.g., CAT [English] → KATZE [German] → GATTO [Italian]). If a participant left the study early (e.g., Internet disconnected, computer crashed, closed the study), the data beyond that point were not recorded. Therefore, the trial-level data represents 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 an incorrect answer on some trials, and these trials were marked for exclusion. All timeout trials were marked as missing values in the final data set. No missing values were imputed.

Trials were also marked for exclusion if they were under the minimum response latency of 160 ms[52]. Further, *lab.js* automatically codes timeout data with a special marker (i.e., data ended on response or timeout as a column), which excludes trials over 3000 ms as the maximum response latency. However, because of variations in browser/screen refresh rates, some trials were answered with response latencies over 3000 ms when a participant made a key press at the very end of the trial before timeout. Given the pre-registered 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 if 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*-score and raw response times are included in the provided data files. The supplemental material includes the number of trials and accuracy for each language, for all participants, and for analyzed participants (https://osf.io/baem5, Table S2). 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]. The descriptive statistics calculated from the trial-level data is 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 a potential outlier (i.e., no response latencies were excluded due to being an "outlier" after removal of excluded participants and trials mentioned above); however, we used a recommended cut-off criterion for absolute value *Z*-score outliers at 2.5 and 3.0[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 values can be merged with the item files. These values do not exist for nonwords. Online tables show the item statistics for average item sample size, average *Z*-scored response time, average *SE* for the Z-scored response latencies separated by item (nonword, word) type and language (https://osf.io/rvt8f, Table S3, S4). The raw response time averages can be found in Table S5. 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 (i.e., one trial per row) to wide data (i.e., cue-target priming trial combinations paired together on one line), and 2) summary data, which includes 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. Therefore, the timing for DOG-CAT would be subtracted from BUS-CAT to indicate the priming effect for the word CAT. The similarity scores calculated during stimuli 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 Figure 1 as part of Hypotheses 1 and 2. Figure S1 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, and the supplementary materials include the rescored accuracy versions (https://osf.io/r4fym, 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 and 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 correlation across 100 random runs to estimate overall reliability. The average reliability was .56 for items (Mdn = .56), and.08 for participants (Mdn = .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 = .59. A linear model of sample size predicting reliability indicates that an average sample size for unrelated and related conditions of $n \sim 557$ participants could potentially achieve a reliability of .80. Item reliability is likely impacted by other variables, as languages 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 = .58 and r = .56).

Hypothesis 1

Hypothesis 1 predicted finding semantic facilitation wherein the response latencies for related targets would be faster than 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. Therefore, values that are positive and greater than zero (i.e., > 0.0001) indicate priming because the related pair had a faster response latency than the unrelated pair. The intercept and its 95% confidence interval represent the grand mean of the priming effect across all languages.

The overall *Z*-scored priming effect was $b_0 = 0.12$, $SE = 0.001$, *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: $b_{0Z2.5} = 0.10$, $SE = 0.001$, *95%CI* [0.10, 0.11], and $b_{0Z3.0} = 0.11$, $SE = 0.001$, *95%CI* [0.10, 0.11]. Figure 1 denotes the distribution of the average item *Z*-score effects, ordered by the size of the overall priming effect for each language (see raw response time effects in Figure S1). 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 $b_0 = 30.61$, $SE = 0.43$, *95%CI* [29.78, 31.45], $b_{0Z2.5} = 27.12$, $SE = 0.36$, *95%CI* [26.51, 27.92], and $b_{0Z3.0} = 28.08$, $SE = 0.37$, *95%CI* [27.35, 28.81].

[Figure 1]

Hypothesis 2

Hypothesis 2 explored the extent to which these semantic priming effects vary across languages. Therefore, we 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 2 for AIC values and their difference scores for comparison. The addition of this parameter improved model fit supporting significant heterogeneity as the value of AIC for the random effects model is two points or more lower than the value of 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 .01[60]. The random effect was useful in both Z-score 2.5 and 3.0 models wherein the random effect sizes were similar to the overall model: $Z_{2.5}$ = 0.02, *95% CI* [0.01, 0.02], $Z_{3.0}$ = 0.02, *95% CI* [0.01, 0.03].

Figure 2 portrays the forest plot for the average priming effects by language, ordered by the size of the effect without the removal of outliers (see Figure S2 for raw response time effects). The global priming average is presented on each facet to show how the priming effect changes based 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 likely small, given pseudo-R^2 was < .01.

[Figure 2]

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, more than 21 million trials were collected using an adaptive stimulus presentation algorithm that shifted data collection toward 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 analyzed languages, we demonstrated consistent non-zero priming effects ranging from *Z* = 0.09 to 0.15, and this effect is robust to the removal of strong priming pairs with high *Z*-scores such as ROMEO-JULIET, GOLDEN-SILVER, MENTAL-EMOTIONAL, and BLIND-DEAF (i.e., 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 (*SD* = 0.11, min = 0.22, max = 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 (1200 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 SOA 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 (.04[40], .17-.33[38]) and are more in line with other estimates (.66 standardized LDT[39]). The participant reliability estimates are considerably lower than previous examinations of the Semantic Priming Project for first associates (.21-.27) but somewhat similar to results for other associates (.07-.08[62]) and other studies (-.06-.43[63]). The large sample sizes in this project likely 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 Figure 1 is the relatively similar distributions found for all languages. While Portuguese and Simplified Chinese show clearly non-overlapping confidence intervals in Figure 2 in each *Z*-score calculation, it is somewhat surprising that all means are within the confidence intervals of previous (English) *Z*-score estimates for priming (i.e., stimulus onset asynchrony 1200 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 analyzed 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 (i.e., the Semantic 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 (i.e., 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 (i.e., they would rather the data be collected and stored locally). Further, 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" (i.e., 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 like the Psychological Science Accelerator[14] and the ManyLanguages community[70].

Method

All deviations to method and results can be found in the supplemental 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 will not collect any identifiable private or personal data as part of the experiment. This project was approved by Harrisburg University of Science and Technology conforming to all relevant ethical guidelines and the Declaration of Helsinki, with special care to conform to the General Data Protection Regulation (GDPR; eugdpr.org). Each research lab will obtain local ethical review, rely on the ethical review provided by Harrisburg University, or provide evidence of no required ethical review. The IRB approvals are available on the Open Science Framework (OSF): https://osf.io/wrpj4/. Participants may be compensated for their participation by course credit or payment depending on individual lab resources. Labs will recruit participants via their own local resources. No exclusion criteria for participating in the study will be used, except for a minimum age requirement of 18 years (i.e., 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 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 (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 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 (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 latency[53]. 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 data[72], 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.

While 43 languages were identified for possible data collection, we planned to first publish the data when ten languages have reached the appropriate sample size as outlined above based on 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 details the important facets of the materials. We first explain the types of word-pair conditions in a semantic priming study (i.e., related, unrelated, and nonword). Next, we detail how the related word-pair conditions were created using the OpenSubtitles corpora, new computational modeling 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. Therefore, a cosine distance of 1 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 (i.e., < .15).

Finally, nonword pair conditions were created by using the Wuggy-like algorithm[73] for non-logographic languages. For logographic languages, we consulted with at least two native speakers to change one stroke or radical such that the character(s) were a pronounceable word with no meaning by starting from known nonword lists[74]. 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[75]. 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 (e.g., 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 matched closely to the same number of characters as the original word, minimized transition frequency (i.e., 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 (i.e., phonologically valid) and not pseudo-homophones (i.e., wherein the pronunciation sounds like a real word, KEEP → KEAP)[73]. In cases of disagreement, the native speakers discussed and resolved these inconsistencies. When they marked a nonword for exclusion, a new nonword was generated until speakers agreed it met the rules for nonwords. Native speakers also suggested alternatives, which the lead author checked to ensure that they matched the desired nonword 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 nonword. Therefore, we used 150 related trials (150 word / 0 nonword; 75 pairs), 150 unrelated trials (150 word / 0 nonword; 75 pairs), 200 word-nonword trials (100 word / 100 nonword, this could have been word-nonword or nonword-word combinations to control for answer chaining; 100 pairs), and 300 nonword-nonword trials (0 word / 300 nonword; 150 pairs). These trials were randomly presented to control the transition probability between word and nonword trials (i.e., random presentation should ensure trials do not present a word-word-nonword-nonword style pattern that allows participants to mindlessly guess the answers). Therefore, the yes-no probability was 50% for words-nonwords 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 proportions for each language (https://osf.io/zye59).

Similarity calculation

Corpora. 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[33,34,76], and local/global coherence values from network models. However, the underlying data for these calculations are inconsistent across languages. Therefore, one solution is to use the data present in the OpenSubtitles datasets[20] (i.e., 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 cognitive related phenomena[18,77–84]. Cosine similarity was selected over other similarity measures because of the availability of possible languages and models for this project, as described below.

The OpenSubtitles data includes 62 languages or language combinations (e.g., Chinese-English mix). We used the 10,000 most frequent nouns, adjectives, adverbs, and verbs from each potential language without lemmatization (i.e., converting words into their dictionary form RUNS → RUN). The *udpipe* package[85] is a natural language processing package that contains more than 100 treebanks to assist in part of speech tagging (i.e., labeling words as noun, verb, etc.), parsing (i.e., separating blocks of text into words and their relationship 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 Figure 4 for the model selection process.

[Figure 4]

Modeling. The subs2vec project[46] used the OpenSubtitles data to create *fastText*[86] computational representation 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 procedure for stimuli selection can be reviewed in our supplementary materials and is displayed graphically in Figure 4 (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 has more than 50,000 unique concepts represented in the subtitle data, we used the subtitle model only. If the subtitles do not provide enough linguistic information (i.e., fewer than 50,000 concepts in the corpus), we used the combined Wikipedia and subtitle model[46]. subs2vec contains models with only the OpenSubtitles data, only Wikipedia for a given language, and a combined model of both. The subtitle data has shown to best represent a language[18,77]; however, not all subtitle projects contain a large enough corpus for the subtitles to cover the breadth of the possible concepts within that language (e.g., Afrikaans subtitles only represent approximately 18,000 words).

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 (e.g., RUN is highly related to RUNS). All stopwords (i.e., commonly used words in a language with little semantic meaning such as THE, AN, OF), words with fewer than three characters for non-logographic languages, and words with numeric characters were eliminated (i.e., 1 would be eliminated but not ONE). The stopword lists can be found in the *stopwords* package using the Stopwords ISO dataset[87]. 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 subs2vec were trained for similarity calculation. The 10,000 most frequent concepts were selected at this point.

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* = 1000) were selected from this list using each cue or target only once, favoring pairs with translations in most languages. Therefore, the selection procedure was 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. Therefore, native speakers 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 may have indicated that a particular language does not have separate concepts for the cue-target pairing. In this instance, 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 while avoiding repetition within a cue-target pair as best possible. Translation information is located at https://osf.io/vdme5 within the 03_Materials folders shared online.

Procedure

We describe the important components to the procedure in this section. First, we detail the implementation of the study, focusing on the timing software and adaptive stimuli section, as not all participants see all items. We then discuss the study procedure in order, as shown in Figure 5. First, participants completed a demographic questionnaire, followed by the lexical decision task. We explain how our data compliments the Semantic Priming Project and finally, discuss additional data that researchers can combine with the current dataset.

[Figure 5]

Implementation

Timing software

While 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 our supplemental materials (on OSF or GitHub use the 04_Procedure folder). The study was programmed using *lab.js*[88], which is an online, open-source, study-building software. Precise timing measurement was required for this study, and the *lab.js* team has documented the accuracy of measurement within their framework[89], and previous work has shown no differences between lab and web-based data collection for response latencies[90]. 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 took 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. 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. In order to enforce this requirement, participants were asked to hit the spacebar to continue the study.

Adaptive stimuli 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 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 Project[33], which is ongoing). All data were included in our dataset, and the analysis section describes how we indicated exclusion criteria. 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.

Study Procedure

Demographics. Participants were given a language specific link for each research lab. Participants were asked to indicate their gender (i.e., male, female, other, prefer not to say), year of birth, and education level (i.e., none, elementary school, high school, bachelors, masters, 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-nonword answer keys (see below). A flow chart of the procedure is provided in Figure 5.

Lexical decision task. Instructions on how to complete a lexical decision task were shown on the next screen, followed by 10 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 (i.e., Arial font, dog). 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 (i.e., Z and / on a QWERTY keyboard or < and - on a QWERTZ keyboard or numbers 1 and 9 for languages that had many keyboard layouts). Response keys were mapped such that the "nonword" response option was on the non-dominant hand side of the keyboard, and the "word" response option was on the dominant hand side[91]. Participants made their choice for each concept, and during the practice trials, they received feedback if their answer was correct or incorrect. The next stimulus appeared with an intertrial interval of 500 ms (i.e., the time between the offset of the first concept response and 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 10 trials, participants saw the instruction screen again with a reminder that they would now be doing the real task.

After 100 trials, the participants were shown a short break screen with the option to continue by hitting the spacebar after 10 seconds. This break timed out after 60 seconds. After eight blocks of 100 trials (800 word-nonword decisions), the experiment ended with a thank you screen. On this screen, participants were given instructions on how to indicate that they had completed the study to the appropriate lab. Participants were allowed to take the study multiple times as items were randomly selected for inclusion. An estimate for the time required for the study was approximately 30 minutes 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 minutes.

Comparison to the Semantic Priming Project. This procedure is a continuous lexical decision task wherein every concept (cue and target) is judged for lexicality (i.e., word/nonword). Many priming studies often present cue words for a short period of time prior to the presentation of target words for lexicality judgment. Evidence from the Semantic Priming Project suggests that the stimulus onset asynchrony (i.e., time between non-judged cue word and target word) does not affect overall priming rates (25 versus 23 ms for 200 ms and 1200 ms). Further, adding the lexicality judgment to each presented concept creates a less obvious link between cue and target to avoid potential conscious expectancy generation effects[92,93]. 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. Therefore, this procedure varies from the data collected 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 (e.g., continuous, masked, paired, and naming), and therefore, we should expect similar results.

Additional data. We then combined available lexical and subject rating data with the priming data, and a tutorial is provided in the supplementary documentation on how to download data and combine with available norms (https://osf.io/yd8u4). Lexical measures, such as length, frequency, part of speech, and the number of phonemes (i.e., sounds in a word) are easily created from the concept or the SUBTLEX projects[77–83]. Subjective measures are concept characteristics that are rated by participants, and we included age of acquisition[94–97] (approximate age you learned a concept), imageability[98,99] (how easy the concept comes to mind), concreteness[100] (how concrete is the concept), valence (how positive versus negative is the concept), arousal (how excited or calm a concept makes a person), dominance (the word denotes something that is weak/subordinate or strong/dominant)[24,26], and familiarity (how well a person knows a concept)[101]. These variables were selected from the list of most published databases for linguistic data[17].

Protocol Registration

The pre-registration is at https://osf.io/u5bp6 (updated 5/31/2022).

Data Availability

All raw and processed data will be available for download from GitHub: https://github.com/SemanticPriming/SPAML.

Code Availability

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).

Acknowledgements

- A.D.A. was supported by the South-Eastern Norway Regional Health Authority (#2020023).
- J.L.U. was supported by ANID/CONICYT FONDECYT Iniciación 11190673, Programa de Investigación Asociativa (PIA) en Ciencias Cognitivas (RU-158-2019), Research Center on Cognitive Sciences (CICC), Faculty of Psychology, Universidad de Talca, Chile
- P.K. was supported by APVV-22-0458.
- Y.C.Z. was supported by the National Natural Science Foundation of China (Grant No. 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 DFG Heisenberg Programme (funding ID: 442405852).
- Y.A.N., A. Stückler, F.T., M.H.C., and S. Pfattheicher were supported by Data collection in Danish was supported by the Interacting Minds Centre through the 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 Psychological Science Accelerator, Harrisburg University of Science and Technology, and the Leibniz Institute for Psychology (ZPID).
- C.H. was supported by Fund for Scientific Research Flanders (FWO), grant number FWO G049821N.
- P.A. was supported by Fundação para a Ciência e Tecnologia (FCT), through the Research Center CIS_Iscte (UID/PSI/03125/2020).
- I.C. was supported by Funded by Comunidad de Madrid PhD grants: PIPF-2022/COM-24573.
- M. Köster was supported by Funded by Deutsche Forschungsgemeinschaft (DFG) - grant number 290878970-GRK 2271.

- M. Cavdan was supported by German Research Foundation DFG-project no. 502774891-ORA project “UNTOUCH”.
- M.A. Vadillo was supported by Grant 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 John Templeton Foundation.
- R.M.R. was supported by 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 2.2.2022 by the Italian Ministry of University and Research (MUR), 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 (FNP) 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 National Master’s Scholarship by the Agencia Nacional de Investigación y Desarrollo of Chile 3537/2023.
- R.n.a.J. was supported by the National Science Center (2020/37/B/HS6/00610).
- A.M.A. was supported by OTKA FK 146604 research grant.
- M. Adamkovič was supported by PRIMUS/24/SSH/017; APVV-22-0458.
- M. Perea was supported by Reference of Grant: CIAICO/2021/172, Funder: Department of Innovation, Universities, Science and Digital Society of the Valencian Government.
- J.H.-w.H. was supported by Research Grant Council of Hong Kong (GRF #17608621 to Hsiao).
- A. Sorokowska was supported by Scientific Excellence Incubator "Being Human".
- W.D. was supported by The Fund for Scientific Research Flanders (FWO), project number 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 number: 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 the scientific research at the University of Belgrade – Faculty of Philosophy (contract number 451-03-66/2024-03/200163).
- D.A.S.E.-D. was supported by The researcher would like to thank Prince Sultan University for funding this project 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 (FCT) 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 (SSHRC) [grant number 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, 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.

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

An introductory sentence in the manuscript was edited for clarity: “Semantic priming is defined as the decrease in response latency (i.e., reduced linguistic processing or facilitation) for target words that are semantically related to their cue words in comparison to unrelated cue words”. Links to supplemental materials were added to the method section for direct access to noted materials.

Author contributions

The authors made the following contributions (https://osf.io/uv27t):

- Erin M. Buchanan: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing, NA
- Kelly Cuccolo: Project administration, Supervision, Writing - review & editing
- Tom Heyman: Conceptualization, Data curation, Methodology, Project administration, Validation, Writing - review & editing
- Niels van Berkel: Methodology, Project administration, Software, Writing - review & editing
- Nicholas A. Coles: Validation, Writing - review & editing
- Aishwarya Iyer: Project administration, Writing - review & editing
- Kim Peters: Project administration, Writing - review & editing
- E. van ’t Veer: Project administration, Writing - review & editing
- Maria Montefinese: Conceptualization, Investigation, Methodology, Resources, Writing - original draft, Writing - review & editing
- Nicholas P. Maxwell: Conceptualization, Investigation, Methodology, Writing - review & editing
- Jack E. Taylor: Conceptualization, Methodology, Writing - original draft, Writing - review & editing

- Kathrene D. Valentine: Conceptualization, Methodology, Writing - original draft, Writing - review & editing
- Patrícia Arriaga: Funding acquisition, Investigation, Resources, Writing - review & editing
- Krystian Barzykowski: Investigation, Resources, Supervision, Writing - review & editing
- Leanne Boucher: Investigation, Resources, Supervision, Writing - review & editing
- W. M. Collins: Investigation, Resources, Supervision, Writing - review & editing
- David C. Vaidis: Investigation, Resources, Supervision, Writing - review & editing
- Balazs Aczel: Investigation, Resources, Writing - review & editing
- Ali H. Al-Hoorie: Investigation, Resources, Writing - review & editing
- Ettore Ambrosini: Investigation, Resources, Writing - review & editing
- Théo Besson: Investigation, Resources, Writing - review & editing
- Debora I. Burin: Investigation, Resources, Writing - review & editing
- Muhammad M. Butt: Investigation, Resources, Writing - review & editing
- J. Benjamin Clarke: Investigation, Resources, Writing - review & editing
- Yalda Daryani: Investigation, Resources, Writing - review & editing
- Dina A. S. El-Dakhs : Investigation, Resources, Writing - review & editing
- Mahmoud M. Elsherif: Investigation, Resources, Writing - review & editing
- Maria Fernández-López: Investigation, Resources, Writing - review & editing
- Paulo R. S. Ferreira: Investigation, Resources, Writing - review & editing
- Raquel M. K. Freitag: Investigation, Resources, Writing - review & editing
- Carolina A. Gattei: Investigation, Resources, Writing - review & editing
- Hendrik Godbersen: Investigation, Resources, Writing - review & editing
- Philip A. Grim II: Investigation, Resources, Writing - review & editing
- Peter Halama: Investigation, Resources, Writing - review & editing
- Patrik Havan: Investigation, Resources, Writing - review & editing
- Natalia C. Irrazabal: Investigation, Resources, Writing - review & editing
- Chris Isloi: Investigation, Resources, Writing - review & editing
- Rebecca K. Iversen: Investigation, Resources, Writing - review & editing
- Yoann Julliard: Investigation, Resources, Writing - review & editing
- Aslan Department of Psychology, IIzmir Katip Celebi University, Izmir, Türkiye: Investigation, Resources, Writing - review & editing
- Michal Kohút: Investigation, Resources, Writing - review & editing
- Veronika Kohútová: Investigation, Resources, Writing - review & editing
- Julija Kos: Investigation, Resources, Writing - review & editing
- Alexandra I. Kosachenko: Investigation, Resources, Writing - review & editing
- Tiago J. S. d. Lima: Investigation, Resources, Writing - review & editing
- Matthew HC Mak: Investigation, Resources, Writing - review & editing
- Christina Manouilidou: Investigation, Resources, Writing - review & editing
- Leonardo A. Marciaga: Validation, Visualization, Writing - review & editing
- Xiaolin M. Melinna: Project administration, Resources, Writing - review & editing
- Jacob F. Miranda: Investigation, Project administration, Writing - review & editing
- Coby Morvinski: Investigation, Resources, Writing - review & editing

- Aishwarya Muppoor: Investigation, Resources, Writing - review & editing
- F. Elif Müjdeci: Resources, Validation, Writing - review & editing
- Yngwie A. Nielsen: Investigation, Resources, Writing - review & editing
- Juan C. Oliveros: Investigation, Resources, Writing - review & editing
- Jaš Onič: Investigation, Resources, Writing - review & editing
- Marietta Papadatou-Pastou: Investigation, Resources, Writing - review & editing
- Ishani Patel: Investigation, Resources, Writing - review & editing
- Zoran Pavlović: Investigation, Resources, Writing - review & editing
- Blaž Pažon: Investigation, Resources, Writing - review & editing
- Gerit Pfuhl: Investigation, Resources, Writing - review & editing
- Ekaterina Pronizius: Investigation, Resources, Writing - review & editing
- Timo B. Roettger: Investigation, Resources, Writing - review & editing
- Camilo R. Ronderos: Investigation, Resources, Writing - review & editing
- Susana Ruiz-Fernandez: Investigation, Resources, Writing - review & editing
- Magdalena Senderecka: Investigation, Resources, Writing - review & editing
- Çağlar Solak: Investigation, Resources, Writing - review & editing
- Anna Stückler: Investigation, Resources, Writing - review & editing
- Raluca D. Szekely-Copîndean: Investigation, Resources, Writing - review & editing
- Analí R. Taboh: Investigation, Resources, Writing - review & editing
- Rémi Thériault: Investigation, Resources, Writing - review & editing
- Ulrich S. Tran: Investigation, Resources, Writing - review & editing
- Fabio Trecca: Funding acquisition, Investigation, Writing - review & editing
- José Luis Ulloa: Investigation, Resources, Writing - review & editing
- Marton A. Varga: Investigation, Resources, Writing - review & editing
- Steven Verheyen: Investigation, Resources, Writing - review & editing
- Tijana Vesić Pavlović: Investigation, Resources, Writing - review & editing
- Giada Viviani: Investigation, Resources, Writing - review & editing
- Nan Wang: Investigation, Resources, Writing - review & editing
- Kristyna Zivna: Investigation, Resources, Writing - review & editing
- Chen C. Yun: Investigation, Resources, Writing - review & editing
- Oliver J. Clark: Investigation, Writing - review & editing
- Oguz A. Acar: Investigation, Writing - review & editing
- Matúš Adamkovič: Investigation, Writing - review & editing
- Giulia Agnoletti: Investigation, Writing - review & editing
- Atakan M. Akil: Investigation, Writing - review & editing
- Zainab Alsuhaibani: Investigation, Writing - review & editing
- Simona Amenta: Investigation, Writing - review & editing
- Olga A. Ananyeva: Investigation, Writing - review & editing
- Michael Andreychik: Investigation, Writing - review & editing
- Bernhard Angele: Investigation, Writing - review & editing
- Danna C. Arias Quiñones: Investigation, Writing - review & editing
- Nwadiogo C. Arinze: Investigation, Writing - review & editing

- Adrian D. Askelund: Resources, Writing - review & editing
- Bradley J. Baker: Resources, Writing - review & editing
- Ernest Baskin: Investigation, Writing - review & editing
- Luisa Batalha: Investigation, Writing - review & editing
- Carlota Batres: Investigation, Writing - review & editing
- Maria S. Beato: Investigation, Writing - review & editing
- Manuel Becker: Investigation, Writing - review & editing
- Maja Becker: Investigation, Writing - review & editing
- Maciej Behnke: Investigation, Writing - review & editing
- Christophe Blaison: Investigation, Writing - review & editing
- Anna M. Borghi: Resources, Writing - review & editing
- Eduard Brandstätter: Investigation, Writing - review & editing
- Jacek Buczny: Resources, Writing - review & editing
- Nesrin Budak: Investigation, Writing - review & editing
- Álvaro Cabana: Investigation, Writing - review & editing
- Zhenguang G. Cai: Investigation, Writing - review & editing
- Enrique C. Canessa: Investigation, Writing - review & editing
- Ignacio Castillejo: Investigation, Writing - review & editing
- Müge Cavdan: Investigation, Writing - review & editing
- Luca Cecchetti: Investigation, Writing - review & editing
- Sergio E. Chaigneau: Investigation, Writing - review & editing
- Feria X. W. Chang: Investigation, Writing - review & editing
- Christopher R. Chartier: Investigation, Writing - review & editing
- Sau-Chin Chen: Resources, Writing - review & editing
- Elena Cherniaeva: Investigation, Writing - review & editing
- Morten H. Christiansen: Investigation, Writing - review & editing
- Hu Chuan-Peng: Investigation, Writing - review & editing
- Patrycja Chwiłkowska: Investigation, Writing - review & editing
- Montserrat Comesaña: Investigation, Writing - review & editing
- Chin Wen Cong: Resources, Writing - review & editing
- Casey Cowan: Investigation, Writing - review & editing
- Stéphane D. Dandeneau: Investigation, Writing - review & editing
- Oana A. David: Investigation, Writing - review & editing
- William E. Davis: Investigation, Writing - review & editing
- Elif G. Demirag Burak: Resources, Writing - review & editing
- Barnaby J. W. Dixson: Investigation, Writing - review & editing
- Hongfei Du: Investigation, Writing - review & editing
- Rod Duclos: Investigation, Writing - review & editing
- Wouter Duyck: Investigation, Writing - review & editing
- Liudmila A. Efimova: Investigation, Writing - review & editing
- Ciara Egan: Investigation, Writing - review & editing
- Vanessa Era: Resources, Writing - review & editing

- Thomas R. Evans: Investigation, Writing - review & editing
- Anna Exner: Resources, Writing - review & editing
- Gilad Feldman: Investigation, Writing - review & editing
- Katharina Fellnhofer: Investigation, Writing - review & editing
- Chiara Fini: Resources, Writing - review & editing
- Sarah E. Fisher: Investigation, Writing - review & editing
- Heather D. Flowe: Investigation, Writing - review & editing
- Patricia Garrido-Vásquez: Investigation, Writing - review & editing
- Daniele Gatti: Investigation, Writing - review & editing
- Jason Geller: Validation, Writing - review & editing
- Vaitsa Giannouli: Investigation, Writing - review & editing
- Anna S. Gorokhova: Investigation, Writing - review & editing
- Lindsay M. Griener: Investigation, Writing - review & editing
- Dmitry Grigoryev: Investigation, Writing - review & editing
- Igor Grossmann: Investigation, Writing - review & editing
- Mohammadhesam Hajighasemi: Investigation, Writing - review & editing
- Giacomo Handjaras: Investigation, Writing - review & editing
- Cathy Hauspie: Investigation, Writing - review & editing
- Zhiran He: Resources, Writing - review & editing
- Renata M. Heilman: Investigation, Writing - review & editing
- Amirmahdi Heydari: Investigation, Writing - review & editing
- Alanna M. Hine: Investigation, Writing - review & editing
- Karlijn Hoyer: Resources, Writing - review & editing
- Weronika Hryniszak: Investigation, Writing - review & editing
- Janet H.-w. Hsiao: Investigation, Writing - review & editing
- Guanxiong Huang: Investigation, Writing - review & editing
- Keiko Ihaya: Resources, Writing - review & editing
- Ewa Ilczuk: Investigation, Writing - review & editing
- Tatsunori Ishii: Resources, Writing - review & editing
- Andrei Dumbravă: Investigation, Writing - review & editing
- Katarzyna Jankowiak: Investigation, Writing - review & editing
- Xiaoming Jiang: Resources, Writing - review & editing
- David C. Johnson: Investigation, Writing - review & editing
- Rafał Jończyk: Investigation, Writing - review & editing
- Juhani Järvikivi: Investigation, Writing - review & editing
- Laura Kaczer: Investigation, Writing - review & editing
- Kevin L. Kamermans: Resources, Writing - review & editing
- Johannes A. Karl: Investigation, Writing - review & editing
- Alexander Karner: Investigation, Writing - review & editing
- Pavol Kačmár: Investigation, Writing - review & editing
- Jacob J. Keech: Investigation, Writing - review & editing
- M. Justin Kim: Validation, Writing - review & editing

- Max Korbmacher: Resources, Writing - review & editing
- Kathrin Kostorz: Investigation, Writing - review & editing
- Marta Kowal: Investigation, Writing - review & editing
- Tomas Kratochvil: Resources, Writing - review & editing
- Yoshihiko Kunisato: Resources, Writing - review & editing
- Anna O. Kuzminska: Investigation, Writing - review & editing
- Lívia Körtvélyessy: Investigation, Writing - review & editing
- Fatma E. Köse: Investigation, Writing - review & editing
- Massimo Köster: Investigation, Writing - review & editing
- Magdalena Kękuś: Investigation, Writing - review & editing
- Melanie Labusch: Investigation, Writing - review & editing
- Claus Lamm: Investigation, Writing - review & editing
- Chaak Ming Lau: Resources, Writing - review & editing
- Julieta Laurino: Investigation, Writing - review & editing
- Wilbert Law: Investigation, Writing - review & editing
- Giada Lettieri: Investigation, Writing - review & editing
- Carmel A. Levitan: Investigation, Writing - review & editing
- Jackson G. Lu: Investigation, Writing - review & editing
- Sarah E. MacPherson: Investigation, Writing - review & editing
- Klara Malinakova: Investigation, Writing - review & editing
- Diego Manriquez-Robles: Investigation, Writing - review & editing
- Nicolás Marchant: Investigation, Writing - review & editing
- Marco Marelli: Investigation, Writing - review & editing
- Martín Martínez: Investigation, Writing - review & editing
- Molly F. Matthews: Investigation, Writing - review & editing
- Alan D. A. Mattiassi: Investigation, Writing - review & editing
- Josefina Mattoli-Sánchez: Investigation, Writing - review & editing
- Claudia Mazzuca: Resources, Writing - review & editing
- David P. McGovern: Investigation, Writing - review & editing
- Zdenek Meier: Investigation, Writing - review & editing
- Filip Melinscak: Investigation, Writing - review & editing
- Michal Misiak: Investigation, Writing - review & editing
- Luis C. P. Monteiro: Resources, Writing - review & editing
- David Moreau: Resources, Writing - review & editing
- Sebastian Moreno: Investigation, Writing - review & editing
- Kate E. Mulgrew: Investigation, Writing - review & editing
- Dominique Muller: Investigation, Writing - review & editing
- Tamás Nagy: Investigation, Writing - review & editing
- Marcin Naranowicz: Investigation, Writing - review & editing
- Izuchukwu L. G. Ndukaihe: Investigation, Writing - review & editing
- Maital Neta: Resources, Writing - review & editing
- Lukas Novak: Investigation, Writing - review & editing

- Chisom E. Ogbonnaya: Investigation, Writing - review & editing
- Jessica Jee Won Paek: Resources, Writing - review & editing
- Aspasia E. Paltoglou: Resources, Writing - review & editing
- Francisco J. Parada: Investigation, Writing - review & editing
- Adam J. Parker: Investigation, Writing - review & editing
- Mariola Paruzel-Czachura: Investigation, Writing - review & editing
- Yuri G. Pavlov: Investigation, Writing - review & editing
- Saeed Paydarfard: Investigation, Writing - review & editing
- Dominik Pegler: Investigation, Writing - review & editing
- Mehmet Peker: Investigation, Writing - review & editing
- Manuel Perea: Investigation, Writing - review & editing
- Stefan Pfattheicher: Investigation, Writing - review & editing
- John Protzko: Investigation, Writing - review & editing
- Irina S. Prusova: Investigation, Writing - review & editing
- Katarzyna Pypno-Blajda: Investigation, Writing - review & editing
- Zhuang Qiu: Investigation, Writing - review & editing
- Ulf-Dietrich Reips: Validation, Writing - review & editing
- Gianni Ribeiro: Investigation, Writing - review & editing
- Luca Rinaldi: Investigation, Writing - review & editing
- S. Craig Roberts: Investigation, Writing - review & editing
- Tanja C. Roembke: Investigation, Writing - review & editing
- Marina O. Romanova: Investigation, Writing - review & editing
- Robert M. Ross: Investigation, Writing - review & editing
- Jan P. Röer: Investigation, Writing - review & editing
- Filiz Rızaoğlu: Investigation, Writing - review & editing
- Toni T. Saari: Resources, Writing - review & editing
- Erika Sampaolo: Investigation, Writing - review & editing
- Anabela Caetano Santos: Investigation, Writing - review & editing
- F. Çağlar Sarıçiçek: Investigation, Writing - review & editing
- Kyoshiro Sasaki: Resources, Writing - review & editing
- Frank Scharnowski: Investigation, Writing - review & editing
- Kathleen Schmidt: Investigation, Writing - review & editing
- Amir Sepehri: Investigation, Writing - review & editing
- Halid O. Serçe: Investigation, Writing - review & editing
- T. Sevincer: Investigation, Writing - review & editing
- Cynthia S. Q. Siew: Investigation, Writing - review & editing
- Matilde E. Simonetti: Investigation, Writing - review & editing
- Miroslav Sirota: Investigation, Writing - review & editing
- Agnieszka Sorokowska: Investigation, Writing - review & editing
- Piotr Sorokowski: Investigation, Writing - review & editing
- Ian D. Stephen: Investigation, Writing - review & editing
- Laura M. Stevens: Investigation, Writing - review & editing

- Suzanne L. K. Stewart: Investigation, Writing - review & editing
- David Steyrl: Investigation, Writing - review & editing
- Stefan Stieger: Investigation, Writing - review & editing
- Anna Studzinska: Investigation, Writing - review & editing
- Mar Suarez: Investigation, Writing - review & editing
- Anna Szala: Resources, Writing - review & editing
- Arnaud Szmalec: Investigation, Writing - review & editing
- Daniel Sznycer: Investigation, Writing - review & editing
- Ewa Szumowska: Investigation, Writing - review & editing
- Sinem Söylemez: Investigation, Writing - review & editing
- Bahadır Söylemez: Investigation, Writing - review & editing
- Kaito Takashima: Resources, Writing - review & editing
- Christian K. Tamnes: Resources, Writing - review & editing
- Joel C. R. Tan: Investigation, Writing - review & editing
- Chengxiang Tang: Investigation, Writing - review & editing
- Peter Tavel: Investigation, Writing - review & editing
- Julian Tejada: Investigation, Writing - review & editing
- Benjamin C. Thompson: Investigation, Writing - review & editing
- Jake G. Tiernan: Investigation, Writing - review & editing
- Vicente Torres-Muñoz: Investigation, Writing - review & editing
- Anna K. Touloumakos: Investigation, Writing - review & editing
- Bastien Trémolière: Investigation, Writing - review & editing
- Monika Tschense: Investigation, Writing - review & editing
- Belgüzar N. Türkan: Investigation, Writing - review & editing
- Miguel A. Vadillo: Investigation, Writing - review & editing
- Caterina Vannucci: Investigation, Writing - review & editing
- Michael E. W. Varnum: Investigation, Writing - review & editing
- Martin R. Vasilev: Investigation, Writing - review & editing
- Leigh Ann Vaughn: Investigation, Writing - review & editing
- Fanny Verkampt: Investigation, Writing - review & editing
- Liliana M. Villar: Investigation, Writing - review & editing
- Sebastian Wallot: Investigation, Writing - review & editing
- Lijun Wang: Investigation, Writing - review & editing
- Ke Wang: Resources, Writing - review & editing
- Glenn P. Williams: Investigation, Writing - review & editing
- David Willinger: Investigation, Writing - review & editing
- Kelly Wolfe: Investigation, Writing - review & editing
- Alexandra S. Wormley: Investigation, Writing - review & editing
- Yuki Yamada: Resources, Writing - review & editing
- Yunkai Yang: Resources, Writing - review & editing
- Yuwei Zhou: Investigation, Writing - review & editing
- Mengfan Zhang: Investigation, Writing - review & editing

- Wang Zheng: Investigation, Writing - review & editing
- Yueyuan Zheng: Investigation, Writing - review & editing
- Chenghao Zhou: Resources, Writing - review & editing
- Radka Zidkova: Investigation, Writing - review & editing
- Nina M. Zumbrunn: Investigation, Writing - review & editing
- Ogeday Çoker: Investigation, Writing - review & editing
- Sami Çoksan: Investigation, Writing - review & editing
- Sezin Öner: Investigation, Writing - review & editing
- Asil A. Özdoğru: Investigation, Writing - review & editing
- Seda M. Şahin: Investigation, Writing - review & editing
- Dauren Kasanov: Investigation, Writing - review & editing
- Alexios Arvanitis: Investigation, Writing - review & editing
- Cameron Brick: Validation, Writing - review & editing
- Melissa F. Colloff: Investigation, Writing - review & editing
- Albina Gallyamova: Investigation, Writing - review & editing
- Christopher Koch: Investigation, Writing - review & editing
- Ivan Ropovik: Investigation, Writing - review & editing
- Yucheng C. Zhang: Investigation, Writing - review & editing
- Xingxing Zhou: Investigation, Writing - review & editing
- Sneh Patel: Investigation, Resources, Writing - review & editing
- Jordan W. Suchow: Validation, Writing - review & editing
- Savannah C. Lewis: Investigation, Project administration, Supervision, Writing - review & editing.

Competing interests

The authors declare no competing interests.

Tables

Table 1. Pre-registered Design Table

Question	Hypothesis	Sampling plan (e.g., power analysis)	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 in comparison to unrelated word pairs. H_0: Response latencies for related word-pairs will be slower or equal to those for unrelated word-pairs.	We will sample participants on items until they reach a desired accuracy in parameter estimation confidence interval width (*SE* = 0.09).	We will calculate the mean and 95% confidence interval 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 will be repeated for the data with 2.5 *Z*-score outlier trials excluded and 3.0 *Z*-score outlier trials excluded.	The results will support H_A when the lower limit of the confidence interval is positive and non-zero > 0.0001 The results will be inconclusive when the lower limit of the confidence interval is negative or zero ≤ 0.0001.

Does the semantic priming effect vary across languages?	H_A: Priming response latencies will be variable between languages (i.e., heterogeneous). H_0: Priming response latencies will not be variable between languages (i.e., homogenous).	We will sample participants on items until they reach a desired accuracy in parameter estimation confidence interval width (*SE* = 0.09).	We will add a random-intercept of language to the previous intercept-only model to assess overall heterogeneity. These calculations will be repeated for the data with 2.5 *Z*-score outlier trials excluded and 3.0 *Z*-score outlier trials excluded.	The results will support H_A when the ΔAIC (intercept-only minus random-intercept) is ≥ 2 points. The results will be inconclusive when the ΔAIC (intercept-only minus random-intercept) is < 2 points.

Table 2. AIC Values for Intercept-Only and Random-Effects Model

	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

Table 3. Language Data Collection Sample Sizes, Accuracy, and Median Study Completion Time in Minutes

Language	*N* Include	*N* Exclude	Proportion Correct	*Mdn* Time (Minutes)
Arabic	133	102	0.92	18.67
Czech	1074	362	0.94	19.76
Danish	829	167	0.93	18.70
Dutch	184	25	0.93	17.60
English	5122	1607	0.92	17.64
Farsi	192	110	0.95	17.71
French	869	142	0.95	17.68
German	2628	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	1085	142	0.95	18.10
Japanese	1165	680	0.94	18.69
Korean	975	601	0.91	17.59
Norwegian	85	17	0.93	20.08
Polish	1188	318	0.94	19.15
Portuguese (Combined)	1178	332	0.93	18.25
Romanian	741	174	0.94	19.65
Russian	1806	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	1468	284	0.94	18.04
Thai	65	20	0.95	18.34
Traditional Chinese	174	67	0.92	18.05
Turkish	2218	790	0.93	17.83
Urdu	315	381	0.88	22.15

Note. *Mdn* = median.

Figure Captions

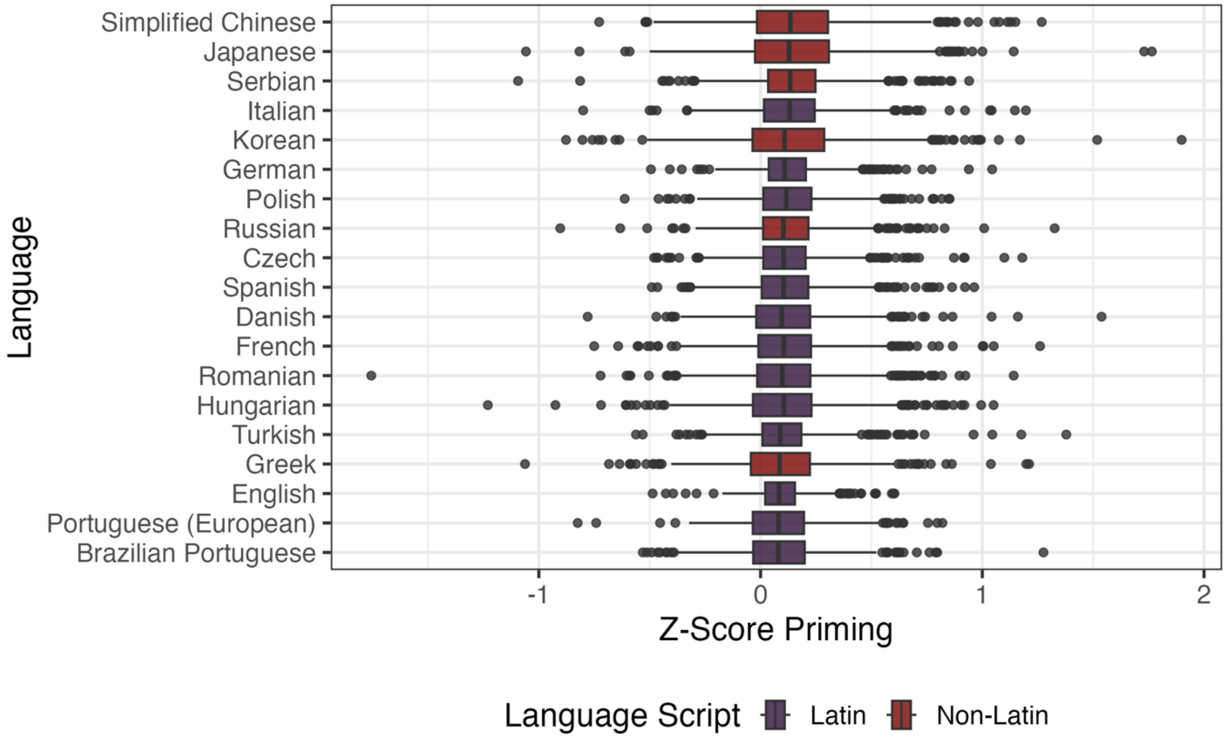

Figure 1 Average priming effect distributions. Distribution of average priming effects for languages that met the minimum sample size criteria using boxplots. Order of languages is based on their average priming effect from smallest (bottom) to largest (top). The pre-registered language selection for the study included a requirement to ensure at least one non-Latin script within the language choices. The graph color codes these languages for convenience to highlight the diversity in included languages. This plot represents all item average data without outliers removed (*n* per language = 1000, total *n* = 19000). The minimum value was *Z* = -1.75, maximum *Z* = 1.90, with the median represented as a solid bar and the interquartile range as the box for the boxplot. The whiskers extend from the end of the boxplot up to 1.5 times the interquartile range. See Figure S1 for raw response times.

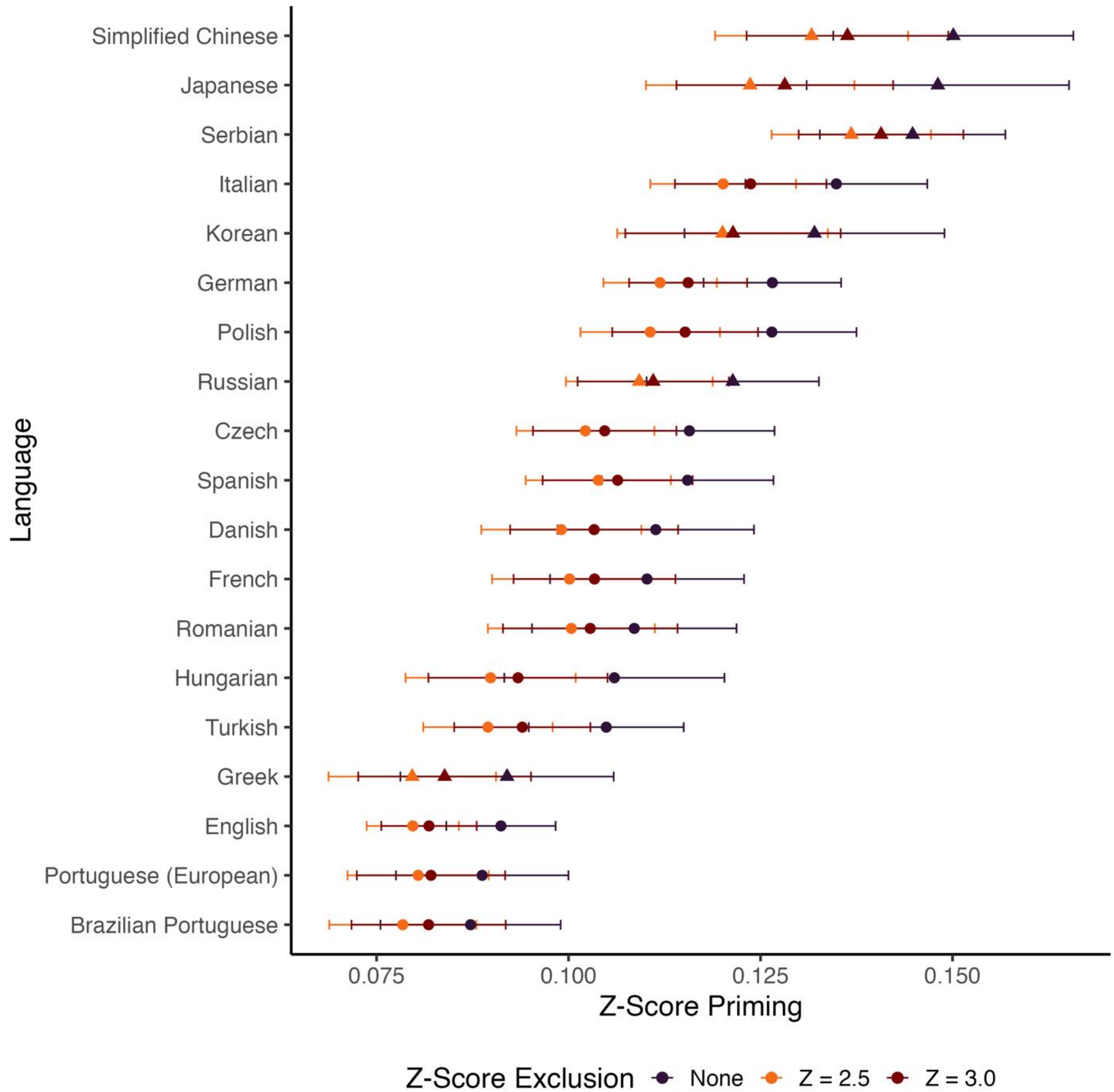

Figure 2 Priming effect sizes. Forest plot of average priming effects for each language ordered by priming average when no outliers are removed (least restrictive), *Z*-scores more than 2.5 are removed (most restrictive), and *Z*-scores more than 3.0 are removed. Sample sizes are based on item averages with *n* = 19000 item averages. Error bars represent a 95% confidence interval. The plot indicates that all priming averages are positive, and their confidence intervals 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 Figure 1. Triangles represent non-Latin languages for convenience, and languages are ordered based on average priming for the no Z-score removal condition from smallest (bottom) to largest (top). See Figure S2 for raw response times, and https://osf.io/m8kjv for the average Z-scores, average raw response latencies, and the standard errors used to create this diagram.

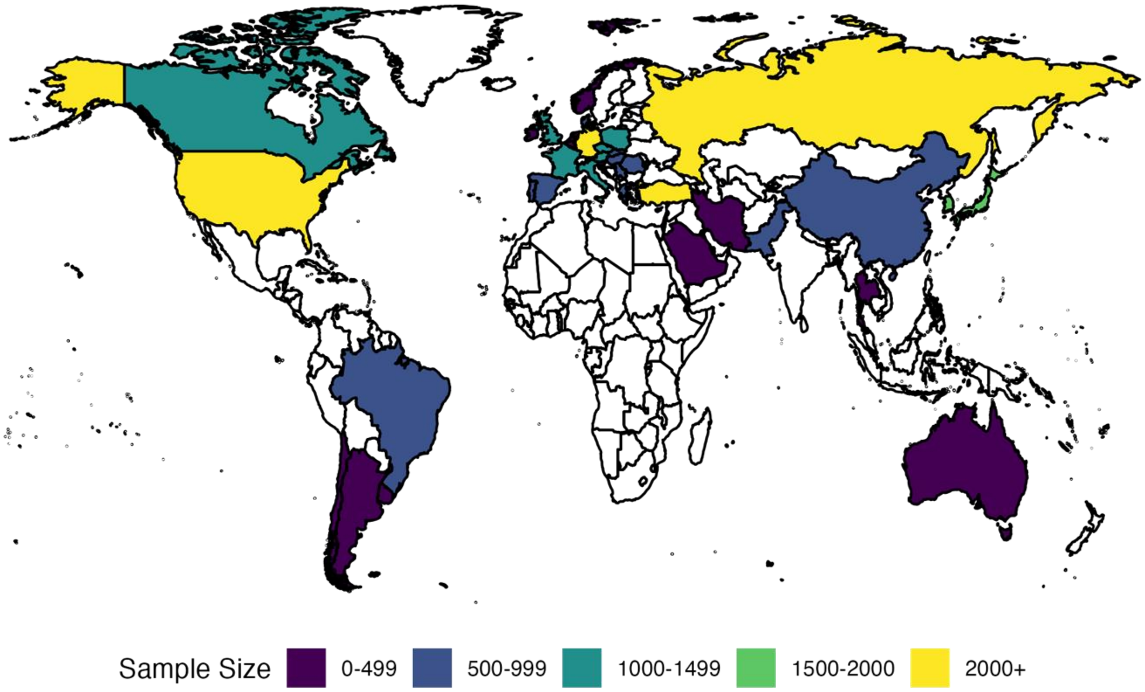

Figure 3 Sample sizes for region and language. Binned sample sizes based on research lab geopolitical region and data collection language demonstrating the full data available for reuse from the project.

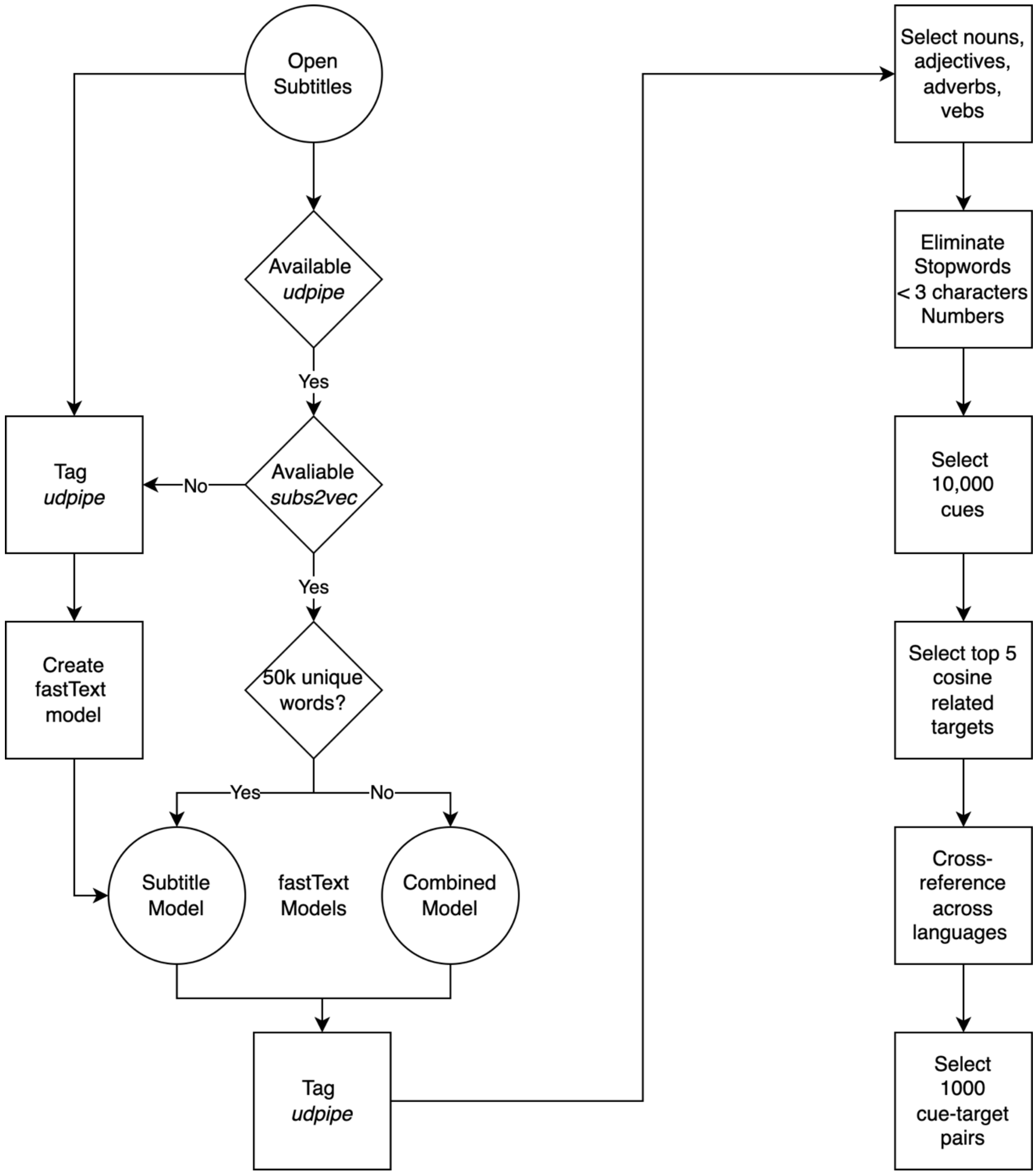

Figure 4 Stimuli selection method. Flow chart of the stimuli selection method. Circles represent the data or models used in the decision tree. Diamonds represent a decision criterion for the data selected. Squares represent coding processes or data reduction for the final stimuli set.

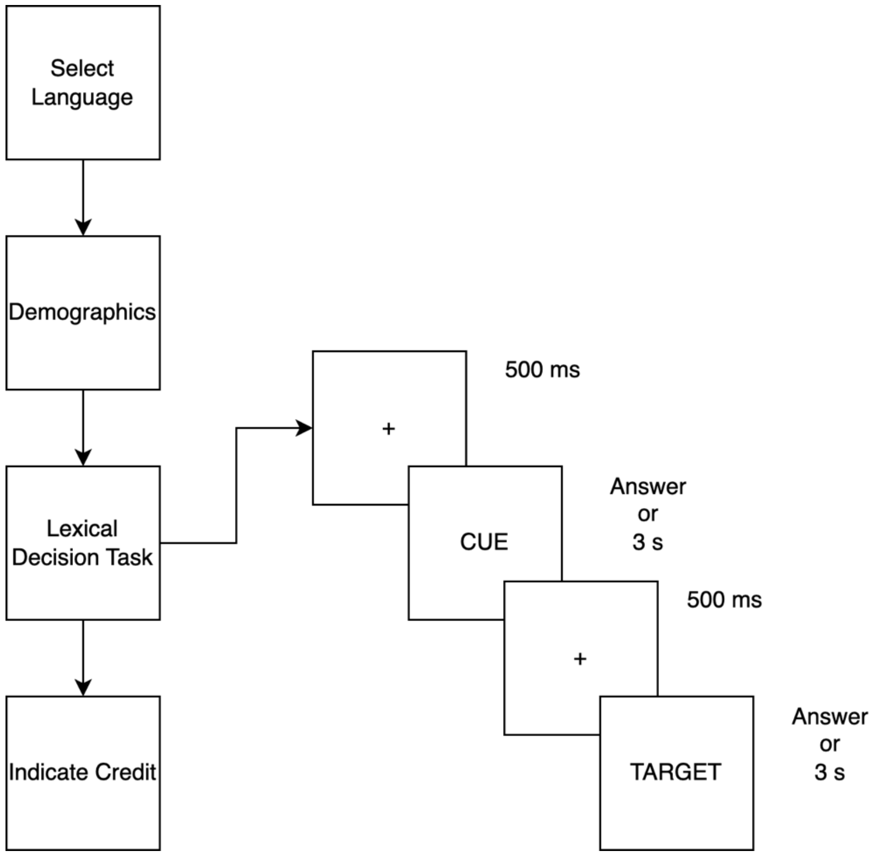

Figure 5 Study procedure. Flow chart of the procedure for the study. Within the lexical decision task, participants were given short breaks after 100 trials. The answer choices for that language were always displayed at the bottom of the screen during the lexical decision task.

References

1. Meyer, D. E. & Schvaneveldt, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* **90**, 227–234 (1971).
2. McNamara, T. P. *Semantic Priming*. (Psychology Press, 2005). doi: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. *Journal of Memory and Language* **92**, 57–78 (2017).
4. Cree, G. S. & Armstrong, B. C. Computational Models of Semantic Memory. in *The Cambridge Handbook of Psycholinguistics* (eds. Spivey, M., McRae, K. & Joanisse, M.) 259–282 (Cambridge University Press, 2012). doi:10.1017/CBO9781139029377.014.
5. McRae, K. & Jones, M. *Semantic Memory*. (Oxford University Press, 2013). doi:10.1093/oxfordhb/9780195376746.013.0014.
6. Rogers, T. T. Computational Models of Semantic Memory. in *The Cambridge Handbook of Computational Psychology* (ed. Sun, R.) 226–266 (Cambridge University Press, 2001). doi: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). doi: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. *Advances in Cognitive Psychology* **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. *Neuroscience Letters* **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. *Advances in Methods and Practices in Psychological Science* **1**, 501–515 (2018).

15. Lucas, M. Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review* **7**, 618–630 (2000).

16. Hutchison, K. A. Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review* **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. *Applied Psycholinguistics* **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. (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. *Behavior Research 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. *Journal of Behavior Therapy and Experimental 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*. (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. *Behavior Research 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. *Cognitive Science* **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. *Cognitive Neuropsychology* **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. *Behavior Research Methods, Instruments, & Computers* **36**, 402–407 (2004).

35. Landauer, T. K., Foltz, P. W. & Laham, D. An introduction to latent semantic analysis. *Discourse Processes* **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. *Psychological Review* **104**, 211–240 (1997).

37. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers* **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. *Quarterly Journal of Experimental Psychology* **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. *Journal of Memory and Language* **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. *Journal of Cognitive Psychology* **23**, 942–961 (2011).

43. Levisen, C. Biases we live by: Anglocentrism in linguistics and cognitive sciences. *Language Sciences* **76**, 101173 (2019).

44. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. in *Advances in neural information processing systems* 3111–3119 (2013).

45. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]* (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. *Sociological Methods & Research* **33**, 261–304 (2004).

48. Tzelgov, J. & Eben-ezra, S. Components of the between-language semantic priming effect. *European Journal of Cognitive Psychology* **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. *Journal of Verbal Learning and Verbal Behavior* **23**, 519–539 (1984).

50. Kelley, K. Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research 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. *Psychological Methods* **23**, 226–243 (2018).

52. Proctor, R. W. & Schneider, D. W. Hick's law for choice reaction time: A review. *Quarterly Journal of Experimental Psychology* **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. *Psychological Bulletin* **125**, 777–799 (1999).

54. Buchanan, E. SemanticPriming/semanticprimeR: semanticprimeR package. [object Object] 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. *Behavior Research Methods, Instruments, & Computers* **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. & Team, R. C. nlme: Linear and nonlinear mixed effects models. (2017).
60. Bartoń, K. MuMIn: Multi-Model Inference. (2020).
61. Bochynska, A. *et al.* Reproducible research practices and transparency across linguistics. Preprint at https://doi.org/10.31222/osf.io/rcews (2022).
62. Yap, M. J., Hutchison, K. A. & Tan, L. C. Individual differences in semantic priming performance: Insights from the semantic priming project. in *Big data in cognitive science* 203–226 (Routledge/Taylor & Francis Group, New York, NY, US, 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. *Visual Cognition* **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. *Language Learning* **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. *Annual Review of Applied Linguistics* **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 in Cognitive Sciences* **26**, 1153–1170 (2022).

70. ManyLanguages. https://many-languages.com/.

71. 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).

72. Overall, J. E. & Woodward, J. A. Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin* **82**, 85–86 (1975).

73. Keuleers, E. & Brysbaert, M. Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* **42**, 627–633 (2010).

74. 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).

75. Michalke, M. sylly: Hyphenation and Syllable Counting for Text Analysis. (2020).

76. 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).

77. 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. *Behavior Research Methods* **41**, 977–990 (2009).

78. van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology* **67**, 1176–1190 (2014).

79. Keuleers, E., Brysbaert, M. & New, B. SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods* **42**, 643–650 (2010).

80. Cai, Q. & Brysbaert, M. SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS ONE* **5**, e10729 (2010).

81. Brysbaert, M. *et al.* The Word Frequency Effect: A Review of Recent Developments and Implications for the Choice of Frequency Estimates in German. *Experimental Psychology* **58**, 412–424 (2011).

82. 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. Psychology* **1**, (2010).

83. Mandera, P., Keuleers, E., Wodniecka, Z. & Brysbaert, M. Subtlex-pl: subtitle-based word frequency estimates for Polish. *Behav Res* **47**, 471–483 (2015).

84. 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).

85. Wijffels, J. *et al.* udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the ‘UDPipe’ ‘NLP’ Toolkit. (2021).

86. Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching Word Vectors with Subword Information. *arXiv preprint arXiv:1607.04606* (2016).

87. Benoit, K., Muhr, D. & Watanabe, K. stopwords: Multilingual Stopword Lists. (2021).

88. Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J. & Hilbig, B. E. *Lab.Js: A Free, Open, Online Study Builder*. https://osf.io/fqr49 (2019) doi:10.31234/osf.io/fqr49.

89. 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. (2018).

90. Hilbig, B. E. Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behav Res* **48**, 1718–1724 (2016).

91. Proctor, R. W. & Cho, Y. S. Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin* **132**, 416–442 (2006).

92. 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 1003–1019 (1989).

93. Shelton, J. R. & Martin, R. C. How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18**, 1191–1210 (1992).

94. Johnston, R. A. & Barry, C. Age of acquisition and lexical processing. *Visual Cognition* **13**, 789–845 (2006).

95. 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 Psychologica* **115**, 43–67 (2004).

96. Juhasz, B. J. Age-of-Acquisition Effects in Word and Picture Identification. *Psychological Bulletin* **131**, 684–712 (2005).

97. Brysbaert, M. & Ellis, A. W. Aphasia and age of acquisition: are early-learned words more resilient? *Aphasiology* **30**, 1240–1263 (2016).

98. Richardson, J. T. E. Imageability and concreteness. *Bull. Psychon. Soc.* **7**, 429–431 (1976).

99. Richardson, J. T. E. Concreteness and Imageability. *Quarterly Journal of Experimental Psychology* **27**, 235–249 (1975).

100. Paivio, A., Walsh, M. & Bons, T. Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**, 1196–1204 (1994).

101. Wilson, M. MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers* **20**, 6–10 (1988).

Links to Supplementary Materials

Please note: all files are synced to OSF through GitHub. We have also included the folder you can find files in if the GitHub add-on is not working on OSF. Since you cannot link directly to a folder on OSF storage, we also indicated where on OSF to find the folder.

Complete Files

- Open Science Framework: https://osf.io/wrpj4/
- GitHub: https://github.com/SemanticPriming/SPAML

Ethics

- Ethics Component OSF Link: https://osf.io/ycn7z/
- Ethics/Lab Table Summary: https://osf.io/ty4hp
  - GitHub: 06_Analysis > supplemental

Power Analysis

- Power analysis code: https://osf.io/v2y9e
  - Github: 02_Power

Method

- Materials separated by language:
  - OSF: 03_Materials
  - Github: 03_Materials
  - The readme explains the stimuli selection and creation procedure: https://osf.io/mz7p4
- *lab.js* Scripts to recreate the experiment:
  - OSF: 04_Procedure
  - Github: 04_Procedure
- Language Table Information: https://osf.io/y3dk7
  - GitHub: 06_Analysis > supplemental
- Deviation Guide: https://osf.io/mwuv3
  - GitHub: 06_Analysis > supplemental
- Translation Information: https://osf.io/vdme5
  - Github: 03_Materials readme

Data

- Data Release: https://github.com/SemanticPriming/SPAML/tree/v1.0.2
- Data Processing Scripts:
  - OSF: 05_Data > data_processing
  - Github: 05_Data > data_processing
- Data Processing Checks/Summary: https://osf.io/zye59
  - Github: 05_Data
- Codebooks:
  - OSF: 05_Data > codebooks
  - Github: 05_Data > codebooks
  - Codebook full data: https://osf.io/xz6nk

    - Codebook item data: https://osf.io/5u9t6
    - Codebook participant data: https://osf.io/9a368
    - Codebook priming trial level data: https://osf.io/49nzq
    - Codebook priming summarized level data: https://osf.io/sx26p
        - Summary table of the sample size calculations: https://osf.io/kv6am
    - Codebook trial data: https://osf.io/s2kqd
- *semanticprimeR* tutorial: https://osf.io/yd8u4

Analyses

- Scripts:
    - OSF: 06_Analysis
    - Github: 06_Analysis
    - Method: https://osf.io/bqpk2
    - Descriptive Statistics
        - Participants: https://osf.io/vdgkr
        - Trials: https://osf.io/baem5
        - Items: https://osf.io/rvt8f
        - Priming: https://osf.io/m8kjv
    - Hypothesis testing: https://osf.io/rmkag
    - Supplemental Meta-Analysis: https://osf.io/rke82
        - Github: 06_Analysis > supplemental
- Supplemental Tables/Summaries:
    - Note: A summary of labs and languages is also in this folder, but linked above
    - Github: 06_Analysis > supplemental
    - Native Language:
        - Overall Native Language Frequency: https://osf.io/ta6wf
        - Analysis Participants Native Language Frequency: https://osf.io/652h8
        - Rescored Analysis Participants Native Language Frequency: https://osf.io/b3y6r
    - Browser Language:
        - Overall Browser Language Frequency: https://osf.io/93kep
        - Analysis Participants Browser Language Frequency: https://osf.io/3yab7
        - Rescored Analysis Participants Browser Language Frequency: https://osf.io/adhbe
    - Lab Reports:
        - Native Language by Lab: https://osf.io/hnrgk
        - Operating System by Lab: https://osf.io/gud6v
        - Web Browser by Lab: https://osf.io/egk9w
        - Language Locale by Lab: https://osf.io/wt3xn
    - Language Reports:
        - Native Language by Language: https://osf.io/5b72x
        - Operating System by Language: https://osf.io/9dwqb
        - Web Browser by Language: https://osf.io/bn7uv
        - Language Locale by Language: https://osf.io/dyh4e
    - Reliability data files:
        - Item Reliability: https://osf.io/r4fym
        - Participant Reliability: https://osf.io/jf28q

Manuscript

- Pre-registration: https://osf.io/u5bp6
- Registered Report: https://osf.io/preprints/osf/q4fjy
- Tenzing chart: https://osf.io/uv27t
    - Github: 08_Credit

Deviation List

Unrelated-pair cosine value deviations

For English, cosine similarity for unrelated pairs were shuffled until all but one pair was less than .15. The pair (ONE-TORTURE) that did not achieve this criterion had a cosine similarity of .20, as the word ONE is a high-frequency word with high cosine similarity values to all targets. For Korean, we increased the unrelated cosine criterion to .20 to find the lowest possible cosine values, as below .15 was not possible for approximately 100 pairs due to the smaller word set size. For Czech, the maximum cosine for unrelated pairs was ~ .16. For Japanese, nearly all pairs were related at very high levels (i.e., M = .80 for cosine). The Japanese model (*fastText*) was created in the same way as described in the subs2vec paper (as it was not available in the subs2vec dataset), but these cosine values are improbable. We shuffled the pairs for the unrelated trials and picked the lowest possible combination for running the study. For Serbian, Simplified Chinese, and Traditional Chinese, the same problem occurred in that all word pairs were very highly correlated. We followed the same procedure as described for Japanese.

Nonword deviations

Translators suggested new nonword options from the computationally generated list. Given that the translators were native speakers, we relied upon their expertise for this component. These suggestions were implemented before data collection. After implementation of trials into the online experiment, a few words were found to be incorrectly marked as nonwords or were misspelled in the dataset. These trials were corrected during data collection or post-data collection in the data processing scripts. These deviations and issues are noted in the data processing files found online.

Word selection deviations

We planned to filter OpenSubtitles for words with at least three characters (excluding logographic languages). This process was completed, and all cue words were at least three characters in length; however, when we matched cues to high-cosine targets, several two-letter words were included. Additionally, due to translation suggestions and cross-referencing, some other two-letter words were also included. For example, in English, MAKE-GO, DOWN-UP, and ENTER-GO were included as potential related cue-target pairs for target selection.

Adaptive implementation deviations

One potential issue with some data collection options labs wanted to use, such as MTurk and Prolific, was the speed of data collection. For example, a researcher can collect data from thousands of participants in an hour via these services. Our study was designed to collect data more slowly across time and to implement the stimuli randomization and selection algorithm. If hundreds of participants came to the study at the same time, we would unevenly collect data on the current stimuli because there is no time to update the stimuli counts. To control for the speed of collection using these sites and any other simultaneous participant runs (i.e., classroom testing), multiple versions of the study were programmed, and participants were assigned to a random version via Qualtrics randomizer. They were then redirected back to their paid provider. Each language continued to use the adaptive randomization and selection algorithm. A summary of data collection procedures by lab is available in the supplementary materials

For large paid samples funded by ZPID and Harrisburg University (https://leibniz-psychology.org/: Japanese, Russian, Turkish, Czech, and Korean), we created 14 different randomizations that evenly distributed the pairs across the study with a small overlap because the important trial combinations (word–word) do not evenly distribute. These were static during the data-collection process to ensure that we obtained 50+ participants in the paid samples for each word–word trial combination. After initial large-scale data collection, the algorithm was turned back on for PSA labs collecting data in those languages.

Additionally, to allow randomization to be more frequent during early stages of data collection, we ran the algorithm randomization process every five minutes once the data collection for a language started. As data size increased, we increased the time interval, to account for the time it took for the algorithm code to run, so that each randomization could finish before the next one was scheduled to start. This process also ensured that the .json files of randomized stimuli were not overwritten or corrupted if two processes were running at once.

An error in the stimulus-writing process led to partial data collection from some participants who appeared to have completed the experiment. The error involved a failure to write new stimuli to the folder used to run the experiment (and therefore, participants were given incorrect practical trials for the first six real blocks followed by two correctly formatted trial blocks before we recognized the error). These tests and inappropriate trials were excluded (please see the data check files for languages and the number of trials affected, summary: https://osf.io/zye59, 05_Data includes all processing files). Other coding-related issues included a typo that showed one trial pair twice at the beginning of the study (affected languages were Czech, English, Japanese, Korean, Russian, and Turkish), instances of garbled items in non-Latin language scripts (e.g., where symbols were shown instead of the Cyrillic characters in Russian), and typos in word spellings. These issues were fixed as soon as they were discovered.

Last, when examining data-collection progress, we noticed that Korean did not have all matched related-unrelated pairs. This error happened during the shuffle to get low cosine values, resulting in too many unrelated trial combinations. Thirty-three new trial combinations were added to ensure each related target had a corresponding unrelated target. In Arabic, the research labs requested that we exclude specific word pairs due to their taboo nature; this request was honored, and thus, the total number of possible stimuli is lower in that language.

Priming calculation deviations

In some cases, a target word was repeated due to language translation. This repetition occurred when translators indicated that there were not separate words for targets within their language, resulting in repeated targets. We created pairs of translations (i.e., cue-target-related1, cue-target-unrelated1, cue-target-related2, cue-target-unrelated2) to ensure each pair only gets subtracted once. For example, if SPOON-CHEESE and TREE-CHEESE (unrelated) needed to be paired with MOUSE-CHEESE and CHEDDAR-CHEESE (related), we ensured each version was only combined once: SPOON-CHEESE minus MOUSE-CHEESE and TREE-CHEESE minus CHEDDAR-CHEESE. For Korean, the extra unrelated pairs accidentally implemented (see above) were excluded in the priming calculation. When the unrelated target was repeated multiple times with no matching related target (i.e., one related target, three unrelated targets), we selected the lowest cosine unrelated target pair to be the comparison condition and discarded the rest of the unrelated pairs. This procedure also allowed us to control the slightly higher cosine values found for unrelated pairs in Korean.

Supplemental Tables

Table S1. Native and Browser Languages for the Overall and Analyzed Participants

	Native Language		Browser Language	
Language	Overall %	Analyzed %	Overall %	Analyzed %
English	15.83	17.19	27.35	27.65
Turkish	8.41	8.63	8.60	8.30
German	7.80	9.39	8.53	9.72
Missing	7.76	1.65	2.85	2.61
Russian	7.61	6.99	8.10	6.99
Spanish	5.39	6.13	4.85	5.35
Japanese	5.03	4.51	5.54	4.57
Polish	4.36	4.65	4.35	4.35
Korean	4.23	3.81	4.58	3.72
Portuguese (Combined)	4.06	4.37	3.98	4.15
Czech	3.88	4.07	4.15	4.04
Italian	3.74	4.38	3.54	4.09
French	2.80	3.31	2.83	3.25
Danish	2.79	3.20	2.61	2.90
Hungarian	2.72	2.96	2.36	2.45
Mandarin	2.58	2.68	NA	NA
Greek	2.35	2.73	1.60	1.73
Serbian	2.27	2.66	0.45	0.50
Romanian	1.99	2.23	0.96	1.08
Chinese	0.62	0.57	2.43	2.24

Note. Native language was coded as Cantonese or Mandarin when the participant used those terms for more specificity. Participants also used a more generic term “Chinese”, and the more specific terminology and generic terms are both included in the table. Browser language meta-data only included “Chinese”, and therefore, is the terminology used here. Values are sorted in descending order by overall native language.

Table S2. Total of Lexical Decision Task (LDT) Trials and Accuracy Proportion by Word-Nonword Trial

	All Participants		Analyzed Participants		All Participants		Analyzed Participants	
Language	Total Nonword Trials	Total Word Trials	Total Nonword Trials	Total Word Trials	Accuracy Nonword	Accuracy Word	Accuracy Nonword	Accuracy Word
Czech	446,465	447,172	396,459	397,150	0.91	0.95	0.94	0.97
Danish	344,582	345,061	311,920	312,264	0.89	0.94	0.92	0.95
English	2,245,604	2,252,266	1,961,546	1,968,289	0.87	0.94	0.91	0.95
French	349,804	350,247	331,078	331,316	0.93	0.96	0.94	0.96
German	1,090,365	1,090,615	1,022,547	1,022,866	0.92	0.95	0.93	0.96
Greek	280,819	281,564	264,274	264,915	0.93	0.94	0.95	0.95
Hungarian	310,186	309,954	279,322	279,126	0.91	0.93	0.94	0.94
Italian	442,736	443,774	420,132	420,889	0.94	0.96	0.95	0.96
Japanese	445,883	444,659	379,645	378,968	0.90	0.92	0.94	0.96
Korean	388,661	390,327	321,070	322,260	0.87	0.92	0.91	0.94
Polish	492,714	492,552	448,989	448,941	0.92	0.95	0.94	0.96
Portuguese (Combined)	495,485	495,373	456,065	456,166	0.89	0.95	0.91	0.96
Romanian	304,296	304,271	278,125	278,246	0.92	0.96	0.93	0.97
Russian	795,078	793,816	652,446	652,149	0.91	0.93	0.95	0.96
Serbian	285,389	285,498	262,660	262,664	0.92	0.95	0.93	0.96
Simplified Chinese	327,479	327,869	274,613	274,870	0.88	0.93	0.92	0.95
Spanish	586,901	586,488	556,113	555,740	0.92	0.95	0.93	0.96
Turkish	898,853	897,783	788,613	788,008	0.91	0.94	0.94	0.95
Overall	10,531,300	10,539,289	9,405,617	9,414,827	0.90	0.94	0.93	0.96

Table S3. Total Number of Unique Trials and Average Trials Per Item

			All Trials		*Z* < 2.5		*Z* < 3.0	
Language	*N* Unique Nonword	*N* Unique Word	*M* Trials Nonword	*M* Trials Word	*M* Trials Nonword	*M* Trials Word	*M* Trials Nonword	*M* Trials Word
Brazilian Portuguese	1,946	1,956	180.75	208.70	172.05	205.65	175.09	206.71
Czech	1,981	1,969	185.05	193.07	176.56	190.18	179.43	191.16
Danish	1,957	1,954	145.73	151.12	138.84	148.48	141.14	149.35
English	1,978	2,000	889.16	932.03	851.22	915.36	863.12	920.45
French	1,976	1,936	156.07	163.90	149.51	161.36	151.66	162.17
German	1,957	1,946	484.48	499.54	463.33	491.11	470.60	493.85
Greek	1,949	1,924	120.51	130.60	115.71	127.85	117.35	128.73
Hungarian	1,936	1,924	134.59	135.65	129.57	132.80	131.25	133.73
Italian	1,992	1,991	197.80	201.52	189.60	198.37	192.38	199.40
Japanese	1,989	1,953	177.24	183.63	170.69	179.39	172.89	180.63
Korean	1,857	1,938	154.96	154.65	149.13	151.40	150.93	152.33
Polish	1,985	1,949	211.16	219.87	202.23	216.29	205.28	217.44
Portuguese (European)	1,965	1,956	183.61	209.07	174.44	206.09	177.64	207.10
Romanian	1,966	1,952	130.63	136.68	124.39	134.80	126.59	135.45
Russian	1,996	1,998	306.39	309.55	294.25	303.59	298.45	305.57
Serbian	1,960	1,957	123.51	128.09	117.67	126.54	120.04	127.15
Simplified Chinese	1,993	1,842	126.09	140.62	120.99	137.76	122.60	138.63
Spanish	1,989	1,941	259.36	273.35	247.93	269.43	251.68	270.71
Turkish	1,866	1,929	391.22	383.96	375.84	376.19	380.81	378.57
Overall	37,238	37,015	239.97	251.59	229.74	247.20	233.16	248.60

Note. N represents sample size.

Table S4. Z-Scored RT Means, Standard Errors for Nonword and Word Trials by Language

	All Trials				*Z* < 2.5				*Z* < 3.0			
Language	*M* Z NW	*M* Z W	*SE* Z NW	*SE* Z W	*M* Z NW	*M* Z W	*SE* Z NW	*SE* Z W	*M* Z NW	*M* Z W	*SE* Z NW	*SE* Z W
Brazilian Portuguese	0.29	-0.26	0.08	0.06	0.12	-0.32	0.06	0.04	0.17	-0.30	0.06	0.05
Czech	0.31	-0.25	0.07	0.06	0.15	-0.31	0.05	0.04	0.19	-0.30	0.06	0.05
Danish	0.28	-0.22	0.08	0.07	0.11	-0.29	0.06	0.05	0.15	-0.27	0.06	0.05
English	0.26	-0.20	0.03	0.03	0.09	-0.28	0.02	0.02	0.13	-0.26	0.03	0.02
French	0.27	-0.23	0.08	0.06	0.12	-0.30	0.06	0.05	0.16	-0.28	0.06	0.05
German	0.26	-0.20	0.04	0.04	0.11	-0.27	0.03	0.03	0.15	-0.25	0.03	0.03
Greek	0.20	-0.14	0.09	0.07	0.05	-0.22	0.07	0.06	0.09	-0.20	0.07	0.06
Hungarian	0.18	-0.13	0.08	0.07	0.05	-0.22	0.06	0.06	0.08	-0.20	0.06	0.06
Italian	0.26	-0.24	0.07	0.06	0.12	-0.31	0.05	0.04	0.15	-0.29	0.05	0.05
Japanese	0.17	-0.13	0.07	0.06	0.04	-0.23	0.05	0.05	0.07	-0.21	0.06	0.05
Korean	0.23	-0.16	0.08	0.07	0.08	-0.26	0.06	0.05	0.11	-0.24	0.06	0.05
Polish	0.27	-0.23	0.07	0.05	0.12	-0.29	0.05	0.04	0.15	-0.28	0.05	0.04
Portuguese (European)	0.35	-0.27	0.08	0.05	0.17	-0.33	0.06	0.04	0.22	-0.31	0.06	0.04
Romanian	0.32	-0.28	0.09	0.07	0.16	-0.33	0.06	0.05	0.20	-0.32	0.07	0.05
Russian	0.21	-0.22	0.05	0.05	0.08	-0.29	0.04	0.04	0.11	-0.27	0.04	0.04
Serbian	0.36	-0.33	0.09	0.06	0.22	-0.37	0.07	0.05	0.27	-0.36	0.07	0.06
Simplified Chinese	0.23	-0.18	0.09	0.07	0.08	-0.27	0.06	0.05	0.11	-0.25	0.07	0.06
Spanish	0.29	-0.25	0.06	0.05	0.13	-0.31	0.05	0.04	0.17	-0.30	0.05	0.04
Turkish	0.22	-0.17	0.05	0.04	0.07	-0.25	0.04	0.03	0.10	-0.24	0.04	0.03
Overall	0.26	-0.21	0.07	0.06	0.11	-0.29	0.05	0.04	0.15	-0.27	0.06	0.05

Note. *M* = mean, *SE* = standard error, NW = nonwords, W = words.

Table S5. Raw RT Means, Standard Errors for Nonword and Word Trials by Language

	All Trials				*Z* < 2.5				*Z* < 3.0			
Language	*M* RT NW	*M* RT W	*SE* RT NW	*SE* RT W	*M* RT NW	*M* RT W	*SE* RT NW	*SE* RT W	*M* RT NW	*M* RT W	*SE* RT NW	*SE* RT W
Brazilian Portuguese	816.22	650.17	27.60	17.67	767.77	633.08	22.25	14.67	781.67	637.77	23.61	15.35
Czech	897.13	733.37	25.23	18.08	851.73	717.09	20.93	15.45	864.23	721.32	21.93	16.00
Danish	817.53	669.35	28.34	21.28	767.45	648.28	22.03	17.10	780.96	653.81	23.51	18.02
English	739.24	619.00	10.37	7.96	695.35	598.94	7.75	6.13	705.67	603.44	8.24	6.45
French	739.52	620.90	22.83	16.80	702.65	605.31	17.91	13.57	711.69	608.93	18.86	14.16
German	810.43	682.87	14.37	11.17	768.79	664.38	11.66	9.17	780.04	668.95	12.25	9.55
Greek	776.00	683.82	28.58	22.45	737.14	661.31	23.02	18.35	747.63	666.66	24.25	19.13
Hungarian	725.44	649.81	23.11	20.27	693.03	628.80	18.54	16.27	701.1	633.87	19.38	17.04
Italian	751.93	627.31	21.02	15.46	715.48	611.94	16.64	12.54	725.03	615.55	17.56	13.07
Japanese	810.06	726.11	24.28	19.56	773.30	701.42	20.01	15.85	782.91	706.83	20.91	16.52
Korean	728.22	636.27	23.51	19.06	690.82	613.00	17.37	14.26	699.12	617.57	18.41	14.98
Polish	803.38	672.52	21.32	16.21	763.82	655.34	17.26	13.40	774.47	659.54	18.16	13.93
Portuguese (European)	809.41	641.84	26.77	17.21	759.56	625.56	21.36	14.28	773.75	629.84	22.71	14.91
Romanian	861.56	680.25	31.06	21.20	813.79	664.80	25.78	18.07	827.71	668.92	27.13	18.74
Russian	856.69	735.68	19.24	16.07	819.06	717.05	16.33	13.80	829.75	721.88	17.02	14.29
Serbian	1017.57	768.09	37.82	26.01	971.96	754.00	34.37	23.58	988.92	758.72	35.55	24.3
Simplified Chinese	750.25	640.14	27.66	21.44	707.90	616.12	20.07	16.22	717.98	621.52	21.55	17.16
Spanish	752.31	614.08	19.41	13.47	711.27	599.41	15.16	10.99	721.6	602.99	16.04	11.47
Turkish	758.58	656.46	15.18	12.98	719.01	634.91	11.71	10.26	728.37	639.84	12.34	10.74
Overall	801.43	669.00	23.57	17.57	759.76	650.22	18.97	14.40	770.96	654.81	19.98	15.02

Supplemental Figures

Figure S1 Average priming effect distributions for raw response times. Distribution of average priming effects using raw response times (in comparison to Z-scores in Figure 1) for languages that met the minimum sample size criteria using boxplots. Order of languages is matched to Figure 1. The pre-registered language selection for the study included a requirement to ensure at least one non-Latin script within the language choices. The graph color codes these languages for convenience to highlight the diversity in included languages. This plot represents all item average data without outliers removed (*n* per language = 1000, total *n* = 19000). The minimum value was -583.64, maximum 550.39, with the median represented as a solid bar and the interquartile range as the box for the boxplot. The whiskers extend from the end of the boxplot up to 1.5 times the interquartile range.

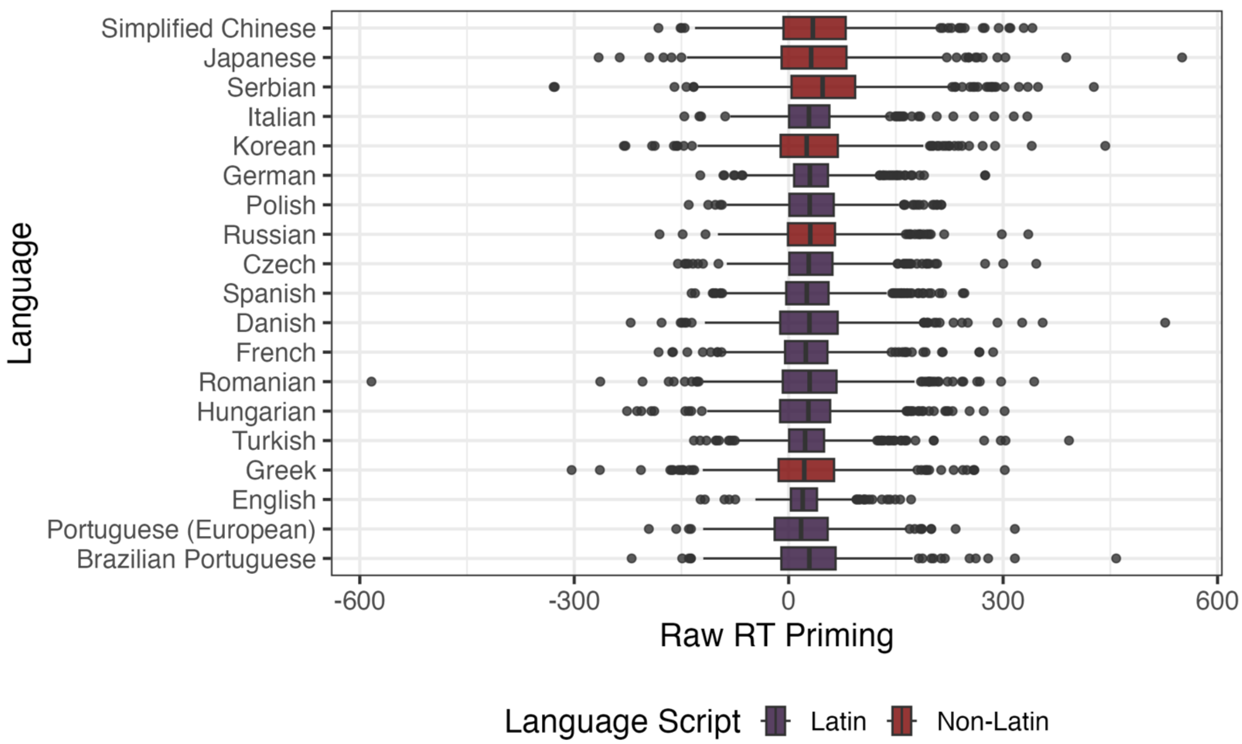

Figure S2 Priming effect sizes for raw response times. Forest plot of average priming effects for raw response times for each language ordered by priming average when no outliers are removed (least restrictive), *Z*-scores more than 2.5 are removed (most restrictive), and *Z*-scores more than 3.0 are removed. The languages are ordered in the same order as Figures 1 and 2. Sample sizes are based on item averages with *n* = 19000 item averages. Error bars represent a 95% confidence interval. Triangles represent non-Latin languages for convenience. See https://osf.io/m8kjv for the average response times, and the standard errors used to create this diagram.