ZOTERO\_TRANSFER\_DOCUMENT

The Zotero citations in this document have been converted to a format that can be safely transferred between word processors. Open this document in a supported word processor and press Refresh in the Zotero plugin to continue working with the citations.

Measuring the Semantic Priming Effect Across Many Languages

Erin M. Buchanan, Analytics, Harrisburg University, Harrisburg, PA, USA, [ebuchanan@harrisburgu.edu](mailto:ebuchanan@harrisburgu.edu)\*, @aggieerin

Kelly M. Cuccolo, Alma College, Alma, MI, USA

Nicholas Coles, Stanford University, CA, USA

Tom Heyman, Methodology and Statistics Unit, Institute of Psychology, Leiden University, The Netherlands

Aishwarya Iyer, Montfort College, KA, India

Neil Lewis Jr., Cornell University, NY, USA

Kim Peters, University of Exeter, UK

Niels van Berkel, Aalborg University, Denmark

Anna E. van ’t Veer, Leiden University, Netherlands

Jack E. Taylor, School of Psychology and Neuroscience, University of Glasgow, UK

Maria Montefinese, IRCCS San Camillo Hospital, Venice, Italy

K. D. Valentine, Massachusetts General Hospital, Boston, MA, USA; Harvard Medical School, Boston, MA, USA

Nicholas P. Maxwell, University of Southern Mississippi, Hattiesburg, MS

Belgüzar Nilay Türkan, Pamukkale University, Department of Psychology, Denizli, Turkey

Glenn P. Williams, School of Psychology, Faculty of Health Sciences and Wellbeing, University of Sunderland, UK

Juan C. Oliveros-Chacana, Centro de Investigación en Ciencias Cognitivas, Facultad de Psicología, Universidad de Talca, Chile

Jan Philipp Röer, Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany

Chiara Fini, Department of Dynamic and Clinical Psychology and Health Studies, Sapienza University of Rome, Italy

Oguz A. Acar, City, University of London, UK

Joseph P. McFall, State University of New York at Fredonia, USA

Ekaterina Pronizius, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria

Jordan W. Suchow, School of Business, Stevens Institute of Technology, USA

Luisa Batalha, Australian Catholic University, Australia

Asil Ali Özdoğru, Department of Psychology, Üsküdar University, İstanbul, Turkey

Hendrik Godbersen, FOM University of Applied Sciences, Essen, Germany

Muhammad Mussaffa Butt, Government College University, Lahore, Pakistan

Jacek Buczny, Department of Experimental and Applied Psychology, VU Amsterdam, The Netherlands

Bastian Jaeger, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, The Netherlands

Bradley J. Baker, Temple University, Philadelphia, PA, USA

Philip A. Grim II, Harrisburg University, Harrisburg, PA, USA

Zainab A. Alsuhaibani, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia

Martín Martínez, University of Navarra, Spain

John Protzko, Central Connecticut State University, USA

Dermot Lynott, Department of Psychology, Maynooth University, Ireland

Max Korbmacher, Department of Health, Western Norway University of Applied, Norway

Mehmet Peker, Ege University, Department of Psychology, Turkey

Barnaby J.W. Dixson, School of Health and Behavioural Sciences, University of the Sunshine Coast, Australia

Mahmoud M. Elsherif, Department of Psychology, University of Birmingham, Birmingham, UK

Maital Neta, Department of Psychology, University of Nebraska-Lincoln, USA

Flavio Azevedo, Cambridge University, Cambridge, UK

Paulo Roberto dos Santos Ferreira, UFGD, Brazil

Fredrik Sigfrids, Åbo Akademi University, Finland

Tiago J S Lima, Department of Work and Social Psychology, University of Brasília

Sandra J. Geiger, Environmental Psychology, Department of Cognition, Emotion, and Methods, Faculty of Psychology, University of Vienna, Vienna, Austria

Anjali Thapar, Bryn Mawr College, USA

Manuel Perea, University of València, Spain

Raluca D. Szekely-Copîndean, Romanian Academy, Cluj-Napoca, Romania

Thomas Rhys Evans, School of Psychology and Counselling, University of Greenwich, England, UK

Steven Verheyen, Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

David Moreau, University of Auckland, NZ

Ulrich S. Tran, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

Dina Abdel Salam El-Dakhs, Prince Sultan University, Saudi Arabia

Izuchukwu L. G. Ndukaihe, Alex Ekwueme Federal University Ndufu-Alike, Nigeria

Tijana Vesić Pavlović, University of Belgrade, Faculty of Mechanical Engineering, Serbia

Debora I. Burin, Universidad de Buenos Aires, Facultad de Psicologi­a / CONICET, Argentina

Patrícia Arriaga, Iscte-Instituto Universitario de Lisboa, CIS-IUL Lisbon, Portugal

Dauren Kasanov, Department Of Psychology, Ural Federal University, Ekaterinburg, Russian Federation

Jacob J. Keech, School of Health and Behavioural Sciences, University of the Sunshine Coast, QLD, Australia

María Fernández-López, University of València, Spain

Suzanne L. K. Stewart, School of Psychology, University of Chester, UK

David C. Vaidis, Université Paris Cité, France

Théo Besson, Université Paris Cité, France

Carlota Batres, Franklin and Marshall College, USA

Leigh Ann Vaughn, Ithaca College, USA

Magdalena Senderecka, Institute of Philosophy, Jagiellonian University, Krakow, Poland

Claudia Mazzuca, Department of Dynamic and Clinical Psychology and Health Studies, Sapienza University of Rome, Italy

Leticia Micheli, Institute of Psychology, Würzburg University, Würzburg, Germany

Martin R. Vasilev, Bournemouth University, Department of Psychology, UK

Kathleen Schmidt, Southern Illinois University, USA

Cameron Brick, University of Amsterdam, Department of Psychology, Amsterdam, Netherlands

Bruno Schivinski, School of Media and Communication, RMIT University, Australia

Susana Ruiz-Fernandez, FOM University of Applied Sciences, Essen, Germany

Ewa Ilczuk, Institute of Psychology, Jagiellonian University, Krakow, Poland

Carmel A Levitan, Department of Cognitive Science, Occidental College, USA

Emily Higgins, Dublin, Ireland

Gerit Pfuhl, Department of Psychology, UiT The Arctic University of Norway,

Jackson G. Lu, Massachusetts Institute of Technology, USA

Miroslav Sirota, University of Essex, UK

Zoran Pavlović, University of Belgrade, Faculty of Philosophy, Department of Psychology, Serbia, Europe

Ettore Ambrosini, Department of Neuroscience, University of Padova, Italy; Department of General Psychology, University of Padova; Padova Neuroscience Center, University of Padova

Nienke Böhm, Vrije Universiteit Amsterdam, The Netherlands, Amsterdam, The Netherlands

Aslan Karaaslan, Ege University, Turkey

Marietta Papadatou-Pastou, National and Kapodistrian University of Athens, Athens, Greece

Sezin Öner, Department of Psychology, Kadir Has University, Istanbul, Turkey

Ernest Baskin, Saint Joseph’s University, USA

Kate E. Mulgrew, School of Health and Behavioural Sciences, University of the Sunshine Coast, QLD, Australia

José Luis Ulloa, Centro de Investigación en Ciencias Cognitivas, Facultad de Psicología, Universidad de Talca, Chile

Ewa Szumowska, Institute of Psychology, Jagiellonian University, Krakow, Poland

Patricia Garrido-Vásquez, Department of Psychology, University of Concepción, Chile

Krystian Barzykowski, Institute of Psychology, Jagiellonian University, Krakow, Poland

Alexandra I. Kosachenko, Ural Federal University, Ekaterinburg, Russia

Chin Wen Cong, Department of Psychology and Counselling, Faculty of Arts and Social Science, Universiti Tunku Abdul Rahman (UTAR), Kampar, Perak, Malaysia

Claus Lamm, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

Andrei Dumbravă, G.Georgescu Institute of Cardiology Iași Romania; Alexandu Ioan Cuza University, Romania

Vanessa Era, Department of Psychology, Sapienza University, Rome, Italy; IRCCS Fondazione Santa Lucia, Rome, Italy

Luis Carlos Pereira Monteiro, Neuroscience and Cell Biology Graduate Program, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil

Peter R. Mallik, Ashland University, Department of Psychology, USA

Chris Isloi, Independent Scientist

Ali H. Al-Hoorie, Royal Commission for Jubail and Yanbu, Jubail Industrial City, Saudi Arabia

Natalia Irrazabal, Universidad de Palermo - National Scientific and Technical Research Council - Argentina

Yuri G. Pavlov, Ural Federal University, Ekaterinburg, Russia

Anna O. Kuzminska, University of Warsaw, Faculty of Management, Poland

William E. Davis, Wittenberg University, USA

Sarah E. Fisher, Ashland University, USA  
Mai Helmy, Psychology Department, College of Education, Sultan Qaboos University- Oman; Psychology department, Faculty of Arts, Menoufia University, Shebin El-Kom, Egypt  
Julia Valeiro Paterlini, Ashland University, USA  
Guanxiong Huang, Department of Media and Communication, City University of Hong Kong, Hong Kong SAR, China  
Anna M. Borghi, Department of Dynamic and Clinical Psychology and Health Studies, Sapienza University of Rome, Italy  
Balazs Aczel, ELTE, Eotvos Lorand University Budapest, Hungary  
Stefan Stieger, Karl Landsteiner University of Health Sciences, Department Psychology and Psychodynamics, Austria  
S. C. Chen, Department of Human Development and Psychology, Tzu-Chi University, Taiwan  
Laura M. Stevens, University of Birmingham, UK  
Christophe Blaison, Université de Paris Cité, France  
Abigail G. Sanders, Ashland University, USA  
Robert M. Ross, Department of Psychology, Macquarie University, Sydney, Australia  
Madeleine P. Ingham, University of Birmingham, UK  
Tia C. Bennett, University of Birmingham, UK  
Jason Geller, Rutgers University Center for Cognitive Science, USA  
Ogeday Çoker, Pamukkale University, Department of Psychology, Denizli, Turkey  
Erin Sievers, Ashland University, USA  
Christopher R. Chartier, Ashland University, Department of Psychology, USA  
Heather D. Flowe, University of Birmingham, England, UK  
Melissa F. Collof, School of Psychology, University of Birmingham, UK  
Francesco Foroni, Australian Catholic University, NSW, Australia  
Tess M. Atkinson, Central Connecticut State University, CT, USA  
Amanda Kaser, Ashland University, OH, USA  
Zdenek Meier, Palacky University Olomouc, Olomouc University Social Health Institute, Olomouc, Czech Republic  
Nwadiogo Chisom ARINZE, Alex Ekwueme Federal University Ndufu-Alike, Nigeria  
Marton A. Varga, ELTE Eotvos Lorand University Budapest, Budapest, Hungary  
David Willinger, Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria  
Rumandeep K. Hayre, University of Birmingham, School of Psychology, UK  
Miguel A. Vadillo, Universidad Autónoma de Madrid, Spain  
Otto Loberg, Department of Psychology, Faculty of Science and Technology, Bournemouth University, UK  
Aspasia Eleni Paltoglou, Department of Psychology, Faculty of Health and Education, Manchester Metropolitan University, UK

Gianni Ribeiro, The University of Queensland, Australia  
Roxana-Elena Morariu, Babeș-Bolyai University, Cluj-Napoca, Romania   
Timo B. Roettger, Universitetet i Oslo, Department of Linguistics and Scandinavian Studies, Norway  
Tolga Ergiyen, Izmir University of Economics, Turkey  
Maja Becker, CLLE, Université de Toulouse, CNRS, France  
Yoann Julliard, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LIP/PC2S, 38000 Grenoble, France  
Fatima Zakra Sahli, Mohammed V University Rabat, Morocco  
Kelly Wolfe, University of Edinburgh, UK  
Klara Malinakova, Olomouc University Social Health Institute, Palacky University in Olomouc, Olomouc, Czech Republic  
Michal Parzuchowski, Center for Research on Cognition and Behavior, SWPS University of Social Sciences and Humanities in Sopot, Poland  
Radka Zidkova, Olomouc University Social Health Institute, Palacky University Olomouc, Olomouc, Czech Republic  
Lukas Novak, Olomouc University Social Health Institute, Palacky University in Olomouc, Olomouc, Czech Republic  
Sarah E MacPherson, Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, UK  
Christopher L Aberson, Cal Poly Humboldt, USA  
Wolf Vanpaemel, University of Leuven, Belgium  
Bernhard Angele, Bournemouth University, UK  
Dominique Muller, Univ. Grenoble Alpes; Institut Universitaire de France, France  
Elif Gizem Demirag Burak, Koc University, Turkey  
Peter Tavel, Olomouc University Social Health Institute, Palacky University in Olomouc, Olomouc, Czech Republic  
Günce Yavuz-Ergiyen, Izmir University of Economics, Turkey

Savannah C. Lewis, Ashland University, OH, USA

Abstract

Semantic priming has been studied for nearly 50 years across various experimental manipulations and theoretical frameworks. These studies provide insight into the cognitive underpinnings of semantic representations in both healthy and clinical populations; however, they have suffered from several issues including generally low sample sizes and a lack of diversity in linguistic implementations. Here, we will test the size and the variability of the semantic priming effect across ten languages by creating a large database of semantic priming values, based on an adaptive sampling procedure. Differences in response latencies between related word-pair conditions and unrelated word-pair conditions (i.e., difference score confidence interval is greater than zero) will allow quantifying evidence for semantic priming, whereas improvements in model fit with the addition of a random intercept for language will provide support for variability in semantic priming across languages.

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[ITEM CSL\_CITATION {"citationID":"HvXLhE51","properties":{"formattedCitation":"\\super 1\\nosupersub{}","plainCitation":"1","noteIndex":0},"citationItems":[{"id":773,"uris":["http://zotero.org/groups/2911188/items/UVW8VIXW"],"itemData":{"id":773,"type":"article-journal","container-title":"Journal of Experimental Psychology","DOI":"10.1037/h0031564","ISSN":"0022-1015","issue":"2","journalAbbreviation":"Journal of Experimental Psychology","language":"en","page":"227-234","source":"DOI.org (Crossref)","title":"Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations.","title-short":"Facilitation in recognizing pairs of words","volume":"90","author":[{"family":"Meyer","given":"David E."},{"family":"Schvaneveldt","given":"Roger W."}],"issued":{"date-parts":[["1971"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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[ITEM CSL\_CITATION {"citationID":"nMNoGDXQ","properties":{"formattedCitation":"\\super 1\\nosupersub{}","plainCitation":"1","noteIndex":0},"citationItems":[{"id":773,"uris":["http://zotero.org/groups/2911188/items/UVW8VIXW"],"itemData":{"id":773,"type":"article-journal","container-title":"Journal of Experimental Psychology","DOI":"10.1037/h0031564","ISSN":"0022-1015","issue":"2","journalAbbreviation":"Journal of Experimental Psychology","language":"en","page":"227-234","source":"DOI.org (Crossref)","title":"Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations.","title-short":"Facilitation in recognizing pairs of words","volume":"90","author":[{"family":"Meyer","given":"David E."},{"family":"Schvaneveldt","given":"Roger W."}],"issued":{"date-parts":[["1971"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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[ITEM CSL\_CITATION {"citationID":"ANnR8kBr","properties":{"formattedCitation":"\\super 2\\nosupersub{}","plainCitation":"2","noteIndex":0},"citationItems":[{"id":812,"uris":["http://zotero.org/groups/2911188/items/8NU73LIM"],"itemData":{"id":812,"type":"book","edition":"0","ISBN":"978-1-135-43255-3","language":"en","note":"DOI: 10.4324/9780203338001","publisher":"Psychology Press","source":"DOI.org (Crossref)","title":"Semantic Priming","URL":"https://www.taylorfrancis.com/books/9781135432553","author":[{"family":"McNamara","given":"Timothy P."}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2005",9,8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), often by using results from these studies to develop theoretical and computational models that capture empirical effects[ITEM CSL\_CITATION {"citationID":"a2dfokisnj5","properties":{"formattedCitation":"\\super 3\\uc0\\u8211{}6\\nosupersub{}","plainCitation":"3–6","noteIndex":0},"citationItems":[{"id":776,"uris":["http://zotero.org/groups/2911188/items/QI7MMVNX"],"itemData":{"id":776,"type":"article-journal","container-title":"Journal of Memory and Language","DOI":"10.1016/j.jml.2016.04.001","ISSN":"0749596X","journalAbbreviation":"Journal of Memory and Language","language":"en","page":"57-78","source":"DOI.org (Crossref)","title":"Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation","title-short":"Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting","volume":"92","author":[{"family":"Mandera","given":"Paweł"},{"family":"Keuleers","given":"Emmanuel"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2017",2]]}}},{"id":787,"uris":["http://zotero.org/groups/2911188/items/7ZKWUIZQ"],"itemData":{"id":787,"type":"chapter","container-title":"The Cambridge Handbook of Psycholinguistics","edition":"1","ISBN":"978-1-139-02937-7","note":"DOI: 10.1017/CBO9781139029377.014","page":"259-282","publisher":"Cambridge University Press","source":"DOI.org (Crossref)","title":"Computational Models of Semantic Memory","URL":"https://www.cambridge.org/core/product/identifier/9781139029377%23c86064-13-1/type/book\_part","editor":[{"family":"Spivey","given":"Michael"},{"family":"McRae","given":"Ken"},{"family":"Joanisse","given":"Marc"}],"author":[{"family":"Cree","given":"George S."},{"family":"Armstrong","given":"Blair C."}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2012",8,20]]}}},{"id":813,"uris":["http://zotero.org/groups/2911188/items/XTNP2K7S"],"itemData":{"id":813,"type":"book","note":"DOI: 10.1093/oxfordhb/9780195376746.013.0014","publisher":"Oxford University Press","source":"DOI.org (Crossref)","title":"Semantic Memory","URL":"http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780195376746.001.0001/oxfordhb-9780195376746-e-14","author":[{"family":"McRae","given":"Ken"},{"family":"Jones","given":"Michael"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2013",3,11]]}}},{"id":816,"uris":["http://zotero.org/groups/2911188/items/NSSXZK73"],"itemData":{"id":816,"type":"chapter","container-title":"The Cambridge Handbook of Computational Psychology","edition":"1","ISBN":"978-0-511-81677-2","note":"DOI: 10.1017/CBO9780511816772.012","page":"226-266","publisher":"Cambridge University Press","source":"DOI.org (Crossref)","title":"Computational Models of Semantic Memory","URL":"https://www.cambridge.org/core/product/identifier/9780511816772%23c85741-ch8/type/book\_part","editor":[{"family":"Sun","given":"Ron"}],"author":[{"family":"Rogers","given":"Timothy T."}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2001",1,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Priming has also been used in studies of attention[ITEM CSL\_CITATION {"citationID":"WHPjz7gw","properties":{"formattedCitation":"\\super 7,8\\nosupersub{}","plainCitation":"7,8","noteIndex":0},"citationItems":[{"id":793,"uris":["http://zotero.org/groups/2911188/items/KLEY75KM"],"itemData":{"id":793,"type":"article-journal","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/s13423-015-0841-4","ISSN":"1069-9384, 1531-5320","issue":"6","journalAbbreviation":"Psychon Bull Rev","language":"en","page":"1577-1597","source":"DOI.org (Crossref)","title":"The negative priming paradigm: An update and implications for selective attention","title-short":"The negative priming paradigm","volume":"22","author":[{"family":"Frings","given":"Christian"},{"family":"Schneider","given":"Katja Kerstin"},{"family":"Fox","given":"Elaine"}],"issued":{"date-parts":[["2015",12]]}}},{"id":820,"uris":["http://zotero.org/groups/2911188/items/VHE8HVFA"],"itemData":{"id":820,"type":"article-journal","container-title":"Cognition","DOI":"10.1016/j.cognition.2011.08.017","ISSN":"00100277","issue":"1","journalAbbreviation":"Cognition","language":"en","page":"91-95","source":"DOI.org (Crossref)","title":"Unconscious semantic activation depends on feature-specific attention allocation","volume":"122","author":[{"family":"Spruyt","given":"Adriaan"},{"family":"De Houwer","given":"Jan"},{"family":"Everaert","given":"Tom"},{"family":"Hermans","given":"Dirk"}],"issued":{"date-parts":[["2012",1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), studies of bi/multilingual people[ITEM CSL\_CITATION {"citationID":"1oZGrbot","properties":{"formattedCitation":"\\super 9,10\\nosupersub{}","plainCitation":"9,10","noteIndex":0},"citationItems":[{"id":811,"uris":["http://zotero.org/groups/2911188/items/3D2FG4JM"],"itemData":{"id":811,"type":"book","edition":"0","ISBN":"978-1-135-59341-4","language":"en","note":"DOI: 10.4324/9780203880944","publisher":"Routledge","source":"DOI.org (Crossref)","title":"Using Priming Methods in Second Language Research","URL":"https://www.taylorfrancis.com/books/9781135593414","author":[{"family":"McDonough","given":"Kim"},{"family":"Trofimovich","given":"Pavel"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2011",2,25]]}}},{"id":819,"uris":["http://zotero.org/groups/2911188/items/SBH2A9CV"],"itemData":{"id":819,"type":"article-journal","container-title":"Child Development","DOI":"10.1111/cdev.12133","ISSN":"00093920","issue":"2","journalAbbreviation":"Child Dev","language":"en","page":"755-766","source":"DOI.org (Crossref)","title":"One World, Two Languages: Cross-Language Semantic Priming in Bilingual Toddlers","title-short":"One World, Two Languages","volume":"85","author":[{"family":"Singh","given":"Leher"}],"issued":{"date-parts":[["2014",3]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), on neurodivergent individuals such as those affected by Parkinson’s disease, aphasia, or schizophrenia[ITEM CSL\_CITATION {"citationID":"FT4AIbLf","properties":{"formattedCitation":"\\super 11\\uc0\\u8211{}13\\nosupersub{}","plainCitation":"11–13","noteIndex":0},"citationItems":[{"id":786,"uris":["http://zotero.org/groups/2911188/items/TUFA67TF"],"itemData":{"id":786,"type":"article-journal","abstract":"The impact of basal ganglia dysfunction on semantic processing was \ninvestigated by comparing the performance of individuals with \nnonthalamic subcortical (NS) vascular lesions, Parkinson's disease \n(PD), cortical lesions, and matched controls on a semantic priming \ntask. Unequibiased lexical ambiguity primes were used in auditory \nprime-target pairs comprising 4 critical conditions; dominant related \n(e.g.,\n bank–money\n ), subordinate related (e.g.,\n bank–river\n ), dominant unrelated (e.g.,\n foot–money\n ) and subordinate unrelated (e.g.,\n bat–river\n ). Participants made speeded lexical decisions \n(word/nonword) on targets using a go–no-go response. When a \nshort prime–target interstimulus interval (ISI) of 200 ms was \nemployed, all groups demonstrated priming for dominant and subordinate \nconditions, indicating nonselective meaning facilitation and intact \nautomatic lexical processing. Differences emerged at the long ISI (1250 \nms), where control and cortical lesion participants evidenced selective \nfacilitation of the dominant meaning, whereas NS and PD groups \ndemonstrated a protracted period of nonselective meaning facilitation. \nThis finding suggests a circumscribed deficit in the selective \nattentional engagement of the semantic network on the basis of meaning \nfrequency, possibly implicating a disturbance of \nfrontal–subcortical systems influencing inhibitory semantic \nmechanisms. (\n JINS\n , 2003,\n 9\n , 1041–1052.)","container-title":"Journal of the International Neuropsychological Society","DOI":"10.1017/S1355617703970081","ISSN":"1355-6177, 1469-7661","issue":"7","journalAbbreviation":"J Int Neuropsychol Soc","language":"en","page":"1041-1052","source":"DOI.org (Crossref)","title":"The basal ganglia and semantic engagement: Potential insights from semantic priming in individuals with subcortical vascular lesions, Parkinson's disease, and cortical lesions","title-short":"The basal ganglia and semantic engagement","volume":"9","author":[{"family":"Copland","given":"David"}],"issued":{"date-parts":[["2003",11]]}}},{"id":795,"uris":["http://zotero.org/groups/2911188/items/84JK567T"],"itemData":{"id":795,"type":"chapter","container-title":"Twenty-First Century Psycholinguistics: Four Cornerstones","edition":"1st","page":"57-68","publisher":"Routledge/Taylor & Francis Group","title":"Linguistic Representation and Language Use in Aphasia","URL":"https://www.taylorfrancis.com/chapters/edit/10.4324/9781315084503-5/linguistic-representation-language-use-aphasia-marco-haverkort","author":[{"family":"Haverkort","given":"Marco"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2005"]]}}},{"id":822,"uris":["http://zotero.org/groups/2911188/items/A5BKTIF8"],"itemData":{"id":822,"type":"article-journal","abstract":"Abstract\n \n Aberrant semantic processing has been linked to the etiology of formal thought disorder (TD) symptoms in schizophrenia. In this cross-sectional study, two prominent theories, overactivation and disorganized structure of semantic memory (SM), were examined in relation to TD symptoms using the continuum approach across two established semantic tasks (direct/indirect semantic priming and categorical fluency). The aim was to examine the validity of the two TD theories in relation to TD symptoms in schizophrenia. Greater direct and indirect priming, fluency productivity and category errors were expected if the data supported the overactivation theory. Reduced fluency productivity and increased category errors would be characteristic of disorganized storage. Fifty-seven schizophrenia/schizoaffective disorder patients and 48 controls completed a clinical assessment and the semantic tasks. There was significantly reduced direct priming in patients compared to controls (\n p\n <.05), while indirect priming was not significantly different; there was no association between TD and degree of priming. Patients produced more category-inappropriate words (\n p\n <.005) than controls, which was related to increasing severity of circumstantiality. The pattern of results was more indicative of a disorganized SM storage problem in this sample. This phenomenon may underlie some TD symptoms in general schizophrenia. The findings strengthen the relationship between SM deficits and TD symptoms, though this appears to differ between individual symptoms. The authors discuss the value of the continuum approach in addressing research questions in TD etiology. Given low levels of TD in this study, replication of these findings in a sample with greater TD is desirable. (\n JINS\n , 2015,\n 21\n , 629–638)","container-title":"Journal of the International Neuropsychological Society","DOI":"10.1017/S1355617715000648","ISSN":"1355-6177, 1469-7661","issue":"8","journalAbbreviation":"J Int Neuropsychol Soc","language":"en","page":"629-638","source":"DOI.org (Crossref)","title":"Assessing the Relationship between Semantic Processing and Thought Disorder Symptoms in Schizophrenia","volume":"21","author":[{"family":"Tan","given":"Eric Josiah"},{"family":"Neill","given":"Erica"},{"family":"Rossell","given":"Susan Lee"}],"issued":{"date-parts":[["2015",9]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), and in a large body of neuroscience studies[ITEM CSL\_CITATION {"citationID":"UcMIQHHM","properties":{"formattedCitation":"\\super 14\\uc0\\u8211{}16\\nosupersub{}","plainCitation":"14–16","noteIndex":0},"citationItems":[{"id":805,"uris":["http://zotero.org/groups/2911188/items/ABHPHY3S"],"itemData":{"id":805,"type":"article-journal","container-title":"Advances in Cognitive Psychology","DOI":"10.2478/v10053-008-0090-4","ISSN":"1895-1171","issue":"-1","journalAbbreviation":"Advances in Cognitive Psychology","page":"55-67","source":"DOI.org (Crossref)","title":"Neuro-cognitive mechanisms of conscious and unconscious visual perception: From a plethora of phenomena to general principles","title-short":"Neuro-cognitive mechanisms of conscious and unconscious visual perception","volume":"7","author":[{"family":"Kiefer","given":"Markus"},{"family":"Ansorge","given":"Ulrich"},{"family":"Haynes","given":"John-Dylan"},{"family":"Hamker","given":"Fred"},{"family":"Mattler","given":"Uwe"},{"family":"Verleger","given":"Rolf"},{"family":"Niedeggen","given":"Michael"}],"issued":{"date-parts":[["2011",1,1]]}}},{"id":821,"uris":["http://zotero.org/groups/2911188/items/LHB5YZCT"],"itemData":{"id":821,"type":"article-journal","container-title":"Neuroscience Letters","DOI":"10.1016/j.neulet.2017.05.007","ISSN":"03043940","journalAbbreviation":"Neuroscience Letters","language":"en","page":"192-197","source":"DOI.org (Crossref)","title":"The priming of priming: Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory","title-short":"The priming of priming","volume":"651","author":[{"family":"Steinhauer","given":"Karsten"},{"family":"Royle","given":"Phaedra"},{"family":"Drury","given":"John E."},{"family":"Fromont","given":"Lauren A."}],"issued":{"date-parts":[["2017",6]]}}},{"id":807,"uris":["http://zotero.org/groups/2911188/items/XP3QYQ2R"],"itemData":{"id":807,"type":"article-journal","container-title":"Neuroscience","DOI":"10.1016/j.neuroscience.2013.02.010","ISSN":"03064522","journalAbbreviation":"Neuroscience","language":"en","page":"319-326","source":"DOI.org (Crossref)","title":"Correlation between prime duration and semantic priming effect: Evidence from N400 effect","title-short":"Correlation between prime duration and semantic priming effect","volume":"238","author":[{"family":"Liu","given":"B."},{"family":"Wu","given":"G."},{"family":"Meng","given":"X."},{"family":"Dang","given":"J."}],"issued":{"date-parts":[["2013",5]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). The purpose of this study is to leverage the power and network of the Psychological Science Accelerator (PSA)[ITEM CSL\_CITATION {"citationID":"u3xbUQZi","properties":{"formattedCitation":"\\super 17\\nosupersub{}","plainCitation":"17","noteIndex":0},"citationItems":[{"id":1401,"uris":["http://zotero.org/groups/2911188/items/5PIDILIS"],"itemData":{"id":1401,"type":"article-journal","abstract":"Concerns about the veracity of psychological research have been growing. Many findings in psychological science are based on studies with insufficient statistical power and nonrepresentative samples, or may otherwise be limited to specific, ungeneralizable settings or populations. Crowdsourced research, a type of large-scale collaboration in which one or more research projects are conducted across multiple lab sites, offers a pragmatic solution to these and other current methodological challenges. The Psychological Science Accelerator (PSA) is a distributed network of laboratories designed to enable and support crowdsourced research projects. These projects can focus on novel research questions or replicate prior research in large, diverse samples. The PSA’s mission is to accelerate the accumulation of reliable and generalizable evidence in psychological science. Here, we describe the background, structure, principles, procedures, benefits, and challenges of the PSA. In contrast to other crowdsourced research networks, the PSA is ongoing (as opposed to time limited), efficient (in that structures and principles are reused for different projects), decentralized, diverse (in both subjects and researchers), and inclusive (of proposals, contributions, and other relevant input from anyone inside or outside the network). The PSA and other approaches to crowdsourced psychological science will advance understanding of mental processes and behaviors by enabling rigorous research and systematic examination of its generalizability.","container-title":"Advances in Methods and Practices in Psychological Science","DOI":"10.1177/2515245918797607","ISSN":"2515-2459, 2515-2467","issue":"4","journalAbbreviation":"Advances in Methods and Practices in Psychological Science","language":"en","page":"501-515","source":"DOI.org (Crossref)","title":"The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network","title-short":"The Psychological Science Accelerator","volume":"1","author":[{"family":"Moshontz","given":"Hannah"},{"family":"Campbell","given":"Lorne"},{"family":"Ebersole","given":"Charles R."},{"family":"IJzerman","given":"Hans"},{"family":"Urry","given":"Heather L."},{"family":"Forscher","given":"Patrick S."},{"family":"Grahe","given":"Jon E."},{"family":"McCarthy","given":"Randy J."},{"family":"Musser","given":"Erica D."},{"family":"Antfolk","given":"Jan"},{"family":"Castille","given":"Christopher M."},{"family":"Evans","given":"Thomas Rhys"},{"family":"Fiedler","given":"Susann"},{"family":"Flake","given":"Jessica Kay"},{"family":"Forero","given":"Diego A."},{"family":"Janssen","given":"Steve M. J."},{"family":"Keene","given":"Justin Robert"},{"family":"Protzko","given":"John"},{"family":"Aczel","given":"Balazs"},{"family":"Álvarez Solas","given":"Sara"},{"family":"Ansari","given":"Daniel"},{"family":"Awlia","given":"Dana"},{"family":"Baskin","given":"Ernest"},{"family":"Batres","given":"Carlota"},{"family":"Borras-Guevara","given":"Martha Lucia"},{"family":"Brick","given":"Cameron"},{"family":"Chandel","given":"Priyanka"},{"family":"Chatard","given":"Armand"},{"family":"Chopik","given":"William J."},{"family":"Clarance","given":"David"},{"family":"Coles","given":"Nicholas A."},{"family":"Corker","given":"Katherine S."},{"family":"Dixson","given":"Barnaby James Wyld"},{"family":"Dranseika","given":"Vilius"},{"family":"Dunham","given":"Yarrow"},{"family":"Fox","given":"Nicholas W."},{"family":"Gardiner","given":"Gwendolyn"},{"family":"Garrison","given":"S. Mason"},{"family":"Gill","given":"Tripat"},{"family":"Hahn","given":"Amanda C."},{"family":"Jaeger","given":"Bastian"},{"family":"Kačmár","given":"Pavol"},{"family":"Kaminski","given":"Gwenaël"},{"family":"Kanske","given":"Philipp"},{"family":"Kekecs","given":"Zoltan"},{"family":"Kline","given":"Melissa"},{"family":"Koehn","given":"Monica A."},{"family":"Kujur","given":"Pratibha"},{"family":"Levitan","given":"Carmel A."},{"family":"Miller","given":"Jeremy K."},{"family":"Okan","given":"Ceylan"},{"family":"Olsen","given":"Jerome"},{"family":"Oviedo-Trespalacios","given":"Oscar"},{"family":"Özdoğru","given":"Asil Ali"},{"family":"Pande","given":"Babita"},{"family":"Parganiha","given":"Arti"},{"family":"Parveen","given":"Noorshama"},{"family":"Pfuhl","given":"Gerit"},{"family":"Pradhan","given":"Sraddha"},{"family":"Ropovik","given":"Ivan"},{"family":"Rule","given":"Nicholas O."},{"family":"Saunders","given":"Blair"},{"family":"Schei","given":"Vidar"},{"family":"Schmidt","given":"Kathleen"},{"family":"Singh","given":"Margaret Messiah"},{"family":"Sirota","given":"Miroslav"},{"family":"Steltenpohl","given":"Crystal N."},{"family":"Stieger","given":"Stefan"},{"family":"Storage","given":"Daniel"},{"family":"Sullivan","given":"Gavin Brent"},{"family":"Szabelska","given":"Anna"},{"family":"Tamnes","given":"Christian K."},{"family":"Vadillo","given":"Miguel A."},{"family":"Valentova","given":"Jaroslava V."},{"family":"Vanpaemel","given":"Wolf"},{"family":"Varella","given":"Marco A. C."},{"family":"Vergauwe","given":"Evie"},{"family":"Verschoor","given":"Mark"},{"family":"Vianello","given":"Michelangelo"},{"family":"Voracek","given":"Martin"},{"family":"Williams","given":"Glenn P."},{"family":"Wilson","given":"John Paul"},{"family":"Zickfeld","given":"Janis H."},{"family":"Arnal","given":"Jack D."},{"family":"Aydin","given":"Burak"},{"family":"Chen","given":"Sau-Chin"},{"family":"DeBruine","given":"Lisa M."},{"family":"Fernandez","given":"Ana Maria"},{"family":"Horstmann","given":"Kai T."},{"family":"Isager","given":"Peder M."},{"family":"Jones","given":"Benedict"},{"family":"Kapucu","given":"Aycan"},{"family":"Lin","given":"Hause"},{"family":"Mensink","given":"Michael C."},{"family":"Navarrete","given":"Gorka"},{"family":"Silan","given":"Miguel A."},{"family":"Chartier","given":"Christopher R."}],"issued":{"date-parts":[["2018",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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[ITEM CSL\_CITATION {"citationID":"EN2YbitO","properties":{"formattedCitation":"\\super 18\\nosupersub{}","plainCitation":"18","noteIndex":0},"citationItems":[{"id":752,"uris":["http://zotero.org/groups/2911188/items/9TM7RT3K"],"itemData":{"id":752,"type":"article-journal","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/BF03212999","ISSN":"1069-9384, 1531-5320","issue":"4","journalAbbreviation":"Psychonomic Bulletin & Review","language":"en","page":"618-630","source":"DOI.org (Crossref)","title":"Semantic priming without association: A meta-analytic review","title-short":"Semantic priming without association","volume":"7","author":[{"family":"Lucas","given":"Margery"}],"issued":{"date-parts":[["2000",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) and Hutchison[ITEM CSL\_CITATION {"citationID":"x5doGB90","properties":{"formattedCitation":"\\super 19\\nosupersub{}","plainCitation":"19","noteIndex":0},"citationItems":[{"id":750,"uris":["http://zotero.org/groups/2911188/items/SBLGEPIT"],"itemData":{"id":750,"type":"article-journal","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/BF03196544","ISSN":"1069-9384, 1531-5320","issue":"4","journalAbbreviation":"Psychonomic Bulletin & Review","language":"en","page":"785-813","source":"DOI.org (Crossref)","title":"Is semantic priming due to association strength or feature overlap? A microanalytic review","title-short":"Is semantic priming due to association strength or feature overlap?","volume":"10","author":[{"family":"Hutchison","given":"Keith A."}],"issued":{"date-parts":[["2003",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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[ITEM CSL\_CITATION {"citationID":"OfBqLGQf","properties":{"formattedCitation":"\\super 20\\nosupersub{}","plainCitation":"20","noteIndex":0},"citationItems":[{"id":753,"uris":["http://zotero.org/groups/2911188/items/HSGWUQ8Y"],"itemData":{"id":753,"type":"article-journal","abstract":"This article presents the Linguistic Annotated Bibliography (LAB) as a searchable Web portal to quickly and easily access reliable database norms, related programs, and variable calculations. These publications were coded by language, number of stimuli, stimuli type (i.e., words, pictures, symbols), keywords (i.e., frequency, semantics, valence), and other useful information. This tool not only allows researchers to search for the specific type of stimuli needed for experiments but also permits the exploration of publication trends across 100 years of research. Details about the portal creation and use are outlined, as well as various analyses of change in publication rates and keywords. In general, advances in computational power have allowed for the increase in dataset size in the recent decades, in addition to an increase in the number of linguistic variables provided in each publication.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1130-8","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1878-1888","source":"Springer Link","title":"LAB: Linguistic Annotated Bibliography – a searchable portal for normed database information","title-short":"LAB","volume":"51","author":[{"family":"Buchanan","given":"Erin M."},{"family":"Valentine","given":"K. D."},{"family":"Maxwell","given":"Nicholas P."}],"issued":{"date-parts":[["2019",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). These datasets include lexical frequency[ITEM CSL\_CITATION {"citationID":"0Dt40dqw","properties":{"formattedCitation":"\\super 21,22\\nosupersub{}","plainCitation":"21,22","noteIndex":0},"citationItems":[{"id":815,"uris":["http://zotero.org/groups/2911188/items/8MC8HFKJ"],"itemData":{"id":815,"type":"article-journal","abstract":"We examine the use of film subtitles as an approximation of word frequencies in human interactions. Because subtitle files are widely available on the Internet, they may present a fast and easy way to obtain word frequency measures in language registers other than text writing. We compiled a corpus of 52 million French words, coming from a variety of films. Frequency measures based on this corpus compared well to other spoken and written frequency measures, and explained variance in lexical decision times in addition to what is accounted for by the available French written frequency measures.","container-title":"Applied Psycholinguistics","DOI":"10.1017/S014271640707035X","ISSN":"0142-7164, 1469-1817","issue":"4","journalAbbreviation":"Applied Psycholinguistics","language":"en","page":"661-677","source":"DOI.org (Crossref)","title":"The use of film subtitles to estimate word frequencies","volume":"28","author":[{"family":"New","given":"Boris"},{"family":"Brysbaert","given":"Marc"},{"family":"Veronis","given":"Jean"},{"family":"Pallier","given":"Christophe"}],"issued":{"date-parts":[["2007",10]]}}},{"id":770,"uris":["http://zotero.org/groups/2911188/items/MYRKNVW5"],"itemData":{"id":770,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-015-0621-0","ISSN":"1554-3528","issue":"3","journalAbbreviation":"Behav Res","language":"en","page":"963-972","source":"DOI.org (Crossref)","title":"Worldlex: Twitter and blog word frequencies for 66 languages","title-short":"Worldlex","volume":"48","author":[{"family":"Gimenes","given":"Manuel"},{"family":"New","given":"Boris"}],"issued":{"date-parts":[["2016",9]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), large collections of text (e.g., corpora)[ITEM CSL\_CITATION {"citationID":"JaZ6qOsY","properties":{"formattedCitation":"\\super 23\\nosupersub{}","plainCitation":"23","noteIndex":0},"citationItems":[{"id":829,"uris":["http://zotero.org/groups/2911188/items/7ZGGXV5T"],"itemData":{"id":829,"type":"article-journal","note":"publisher: European Language Resources Association","title":"Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles","author":[{"family":"Lison","given":"Pierre"},{"family":"Tiedemann","given":"Jörg"}],"issued":{"date-parts":[["2016"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), response latencies,[ITEM CSL\_CITATION {"citationID":"ZEo7mvu0","properties":{"formattedCitation":"\\super 24\\uc0\\u8211{}26\\nosupersub{}","plainCitation":"24–26","noteIndex":0},"citationItems":[{"id":774,"uris":["http://zotero.org/groups/2911188/items/GCWZAB9X"],"itemData":{"id":774,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-012-0304-z","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1099-1114","source":"DOI.org (Crossref)","title":"The semantic priming project","volume":"45","author":[{"family":"Hutchison","given":"Keith A."},{"family":"Balota","given":"David A."},{"family":"Neely","given":"James H."},{"family":"Cortese","given":"Michael J."},{"family":"Cohen-Shikora","given":"Emily R."},{"family":"Tse","given":"Chi-Shing"},{"family":"Yap","given":"Melvin J."},{"family":"Bengson","given":"Jesse J."},{"family":"Niemeyer","given":"Dale"},{"family":"Buchanan","given":"Erin"}],"issued":{"date-parts":[["2013",12]]}}},{"id":782,"uris":["http://zotero.org/groups/2911188/items/URNUPYHW"],"itemData":{"id":782,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BF03193014","ISSN":"1554-351X, 1554-3528","issue":"3","journalAbbreviation":"Behavior Research Methods","language":"en","page":"445-459","source":"DOI.org (Crossref)","title":"The English Lexicon Project","volume":"39","author":[{"family":"Balota","given":"David A."},{"family":"Yap","given":"Melvin J."},{"family":"Hutchison","given":"Keith A."},{"family":"Cortese","given":"Michael J."},{"family":"Kessler","given":"Brett"},{"family":"Loftis","given":"Bjorn"},{"family":"Neely","given":"James H."},{"family":"Nelson","given":"Douglas L."},{"family":"Simpson","given":"Greg B."},{"family":"Treiman","given":"Rebecca"}],"issued":{"date-parts":[["2007",8]]}}},{"id":781,"uris":["http://zotero.org/groups/2911188/items/8QU8WI3W"],"itemData":{"id":781,"type":"article-journal","container-title":"Frontiers in Psychology","DOI":"10.3389/fpsyg.2018.02156","ISSN":"1664-1078","journalAbbreviation":"Front. Psychol.","page":"2156","source":"DOI.org (Crossref)","title":"SPALEX: A Spanish Lexical Decision Database From a Massive Online Data Collection","title-short":"SPALEX","volume":"9","author":[{"family":"Aguasvivas","given":"Jose Armando"},{"family":"Carreiras","given":"Manuel"},{"family":"Brysbaert","given":"Marc"},{"family":"Mandera","given":"Paweł"},{"family":"Keuleers","given":"Emmanuel"},{"family":"Duñabeitia","given":"Jon Andoni"}],"issued":{"date-parts":[["2018",11,12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) and subjective ratings from participants on semantic dimensions such as emotion[ITEM CSL\_CITATION {"citationID":"NM7smRls","properties":{"formattedCitation":"\\super 27\\uc0\\u8211{}29\\nosupersub{}","plainCitation":"27–29","noteIndex":0},"citationItems":[{"id":861,"uris":["http://zotero.org/groups/2911188/items/66LM44QH"],"itemData":{"id":861,"type":"article-journal","container-title":"Journal of Behavior Therapy and Experimental Psychiatry","DOI":"10.1016/0005-7916(94)90063-9","ISSN":"00057916","issue":"1","journalAbbreviation":"Journal of Behavior Therapy and Experimental Psychiatry","language":"en","page":"49-59","source":"DOI.org (Crossref)","title":"Measuring emotion: The self-assessment manikin and the semantic differential","title-short":"Measuring emotion","volume":"25","author":[{"family":"Bradley","given":"Margaret M."},{"family":"Lang","given":"Peter J."}],"issued":{"date-parts":[["1994",3]]}}},{"id":766,"uris":["http://zotero.org/groups/2911188/items/QJ84D4TI"],"itemData":{"id":766,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-012-0314-x","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1191-1207","source":"DOI.org (Crossref)","title":"Norms of valence, arousal, and dominance for 13,915 English lemmas","volume":"45","author":[{"family":"Warriner","given":"Amy Beth"},{"family":"Kuperman","given":"Victor"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2013",12]]}}},{"id":862,"uris":["http://zotero.org/groups/2911188/items/RIY3KN6I"],"itemData":{"id":862,"type":"report","publisher":"Technical report C-1, the center for research in psychophysiology …","title":"Affective norms for English words (ANEW): Instruction manual and affective ratings","author":[{"family":"Bradley","given":"Margaret M"},{"family":"Lang","given":"Peter J"}],"issued":{"date-parts":[["1999"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), concreteness[ITEM CSL\_CITATION {"citationID":"kI1LBNO1","properties":{"formattedCitation":"\\super 30\\nosupersub{}","plainCitation":"30","noteIndex":0},"citationItems":[{"id":764,"uris":["http://zotero.org/groups/2911188/items/8AAEFNB5"],"itemData":{"id":764,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-013-0403-5","ISSN":"1554-3528","issue":"3","journalAbbreviation":"Behav Res","language":"en","page":"904-911","source":"DOI.org (Crossref)","title":"Concreteness ratings for 40 thousand generally known English word lemmas","volume":"46","author":[{"family":"Brysbaert","given":"Marc"},{"family":"Warriner","given":"Amy Beth"},{"family":"Kuperman","given":"Victor"}],"issued":{"date-parts":[["2014",9]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), or familiarity[ITEM CSL\_CITATION {"citationID":"1NCBG0QE","properties":{"formattedCitation":"\\super 31\\nosupersub{}","plainCitation":"31","noteIndex":0},"citationItems":[{"id":850,"uris":["http://zotero.org/groups/2911188/items/CD7TV3PS"],"itemData":{"id":850,"type":"article-journal","abstract":"Ratings of age of acquisition (AoA), imageability, and familiarity were collected for 1,526 words. The methodology made use of a modular approach, in which the full sample of words was divided into five separate blocks. Within each block, each word was rated on each of the three variables by 20 participants (undergraduate students from the University of Bristol). Analyses comparing these ratings to existing norm databases demonstrated that this methodology resulted in high reliability (assessed by Cronbach’s α) and validity. The ratings were also transformed to be compatible with the Gilhooly and Logie (1980) norms. This transformation resulted in a set of norms for 3,394 words, which is by far the largest database of ratings for AoA, imageability, and familiarity to date. The resulting database should be useful for researchers interested in manipulating or controlling these factors in word recognition, neuropsychological, or memory studies. These norms can be downloaded from language.psy.bris .ac.uk/bristol\_norms.html.","container-title":"Behavior Research Methods","DOI":"10.3758/BF03193891","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behavior Research Methods","language":"en","page":"598-605","source":"Springer Link","title":"The Bristol norms for age of acquisition, imageability, and familiarity","volume":"38","author":[{"family":"Stadthagen-Gonzalez","given":"Hans"},{"family":"Davis","given":"Colin J."}],"issued":{"date-parts":[["2006",11,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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[ITEM CSL\_CITATION {"citationID":"sx99MXkn","properties":{"formattedCitation":"\\super 32\\uc0\\u8211{}35\\nosupersub{}","plainCitation":"32–35","noteIndex":0},"citationItems":[{"id":755,"uris":["http://zotero.org/groups/2911188/items/DVY399LK"],"itemData":{"id":755,"type":"article-journal","container-title":"Journal of Experimental Psychology: General","DOI":"10.1037/xge0000192","ISSN":"1939-2222, 0096-3445","issue":"9","journalAbbreviation":"Journal of Experimental Psychology: General","language":"en","page":"1228-1254","source":"DOI.org (Crossref)","title":"Structure at every scale: A semantic network account of the similarities between unrelated concepts.","title-short":"Structure at every scale","volume":"145","author":[{"family":"De Deyne","given":"Simon"},{"family":"Navarro","given":"Daniel J."},{"family":"Perfors","given":"Amy"},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2016",9]]}}},{"id":754,"uris":["http://zotero.org/groups/2911188/items/AQRLJE4H"],"itemData":{"id":754,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1115-7","ISSN":"1554-3528","issue":"3","journalAbbreviation":"Behav Res","language":"en","page":"987-1006","source":"DOI.org (Crossref)","title":"The “Small World of Words” English word association norms for over 12,000 cue words","volume":"51","author":[{"family":"De Deyne","given":"Simon"},{"family":"Navarro","given":"Danielle J."},{"family":"Perfors","given":"Amy"},{"family":"Brysbaert","given":"Marc"},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2019",6]]}}},{"id":756,"uris":["http://zotero.org/groups/2911188/items/4T4P7VQA"],"itemData":{"id":756,"type":"article-journal","container-title":"Journal of Cognition","DOI":"10.5334/joc.50","ISSN":"2514-4820","issue":"1","language":"en","page":"45","source":"DOI.org (Crossref)","title":"Predicting Lexical Norms: A Comparison between a Word Association Model and Text-Based Word Co-occurrence Models","title-short":"Predicting Lexical Norms","volume":"1","author":[{"family":"Vankrunkelsven","given":"Hendrik"},{"family":"Verheyen","given":"Steven"},{"family":"Storms","given":"Gert"},{"family":"De Deyne","given":"Simon"}],"issued":{"date-parts":[["2018",11,27]]}}},{"id":757,"uris":["http://zotero.org/groups/2911188/items/G9NUW9IP"],"itemData":{"id":757,"type":"article-journal","abstract":"Abstract\n Network science is an emerging discipline drawing from sociology, computer science, physics and a number of other fields to examine complex systems in economical, biological, social, and technological domains. To examine these complex systems, nodes are used to represent individual entities, and links are used to represent relationships between entities, forming a web-like structure, or network, of the entire system. The structure that emerges in these complex networks influences the dynamics of that system. We provide a short review of how this mathematical approach has been used to examine the structure found in the phonological lexicon, and of how subsequent psycholinguistic investigations demonstrate that several of the structural characteristics of the phonological network influence various language-related processes, including word retrieval during the recognition and production of spoken words, recovery from instances of failed lexical retrieval, and the acquisition of word-forms. This approach allows researchers to examine the lexicon at the micro-, meso-, and macro-levels, holding much promise for increasing our understanding of language-related processes and representations.","container-title":"Yearbook of the Poznan Linguistic Meeting","DOI":"10.1515/yplm-2015-0007","ISSN":"2449-7525","issue":"1","language":"en","page":"119-138","source":"DOI.org (Crossref)","title":"Using complex networks to understand the mental lexicon","volume":"1","author":[{"family":"Vitevitch","given":"Michael S."},{"family":"Goldstein","given":"Rutherford"},{"family":"Siew","given":"Cynthia S.Q."},{"family":"Castro","given":"Nichol"}],"issued":{"date-parts":[["2014",12,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/).

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[ITEM CSL\_CITATION {"citationID":"I4nE1NXT","properties":{"formattedCitation":"\\super 36\\nosupersub{}","plainCitation":"36","noteIndex":0},"citationItems":[{"id":1402,"uris":["http://zotero.org/groups/2911188/items/K6G6PQGU"],"itemData":{"id":1402,"type":"article-journal","abstract":"This article considers a practice in scientific communication termed HARKing (Hypothesizing After the Results are Known). HARKing is defined as presenting a post hoc hypothesis (i.e., one based on or informed by one's results) in one's research report as if it were, in fact, an a priori hypotheses. Several forms of HARKing are identified and survey data are presented that suggests that at least some forms of HARKing are widely practiced and widely seen as inappropriate. I identify several reasons why scientists might HARK. Then I discuss several reasons why scientists ought not to HARK. It is conceded that the question of whether HARKing's costs exceed its benefits is a complex one that ought to be addressed through research, open discussion, and debate. To help stimulate such discussion (and for those such as myself who suspect that HARKing's costs do exceed its benefits), I conclude the article with some suggestions for deterring HARKing.","container-title":"Personality and Social Psychology Review","DOI":"10.1207/s15327957pspr0203\_4","ISSN":"1088-8683, 1532-7957","issue":"3","journalAbbreviation":"Pers Soc Psychol Rev","language":"en","page":"196-217","source":"DOI.org (Crossref)","title":"HARKing: Hypothesizing After the Results are Known","title-short":"HARKing","volume":"2","author":[{"family":"Kerr","given":"Norbert L."}],"issued":{"date-parts":[["1998",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Semantic similarity has various conceptual and computational definitions that all generally describe the shared meaning between two words or texts[ITEM CSL\_CITATION {"citationID":"IWQ59J8U","properties":{"formattedCitation":"\\super 5\\nosupersub{}","plainCitation":"5","noteIndex":0},"citationItems":[{"id":813,"uris":["http://zotero.org/groups/2911188/items/XTNP2K7S"],"itemData":{"id":813,"type":"book","note":"DOI: 10.1093/oxfordhb/9780195376746.013.0014","publisher":"Oxford University Press","source":"DOI.org (Crossref)","title":"Semantic Memory","URL":"http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780195376746.001.0001/oxfordhb-9780195376746-e-14","author":[{"family":"McRae","given":"Ken"},{"family":"Jones","given":"Michael"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2013",3,11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). The most common forms of similarity are feature-based similarity (i.e., number of shared features between words)[ITEM CSL\_CITATION {"citationID":"F3R6DBTO","properties":{"formattedCitation":"\\super 37\\uc0\\u8211{}39\\nosupersub{}","plainCitation":"37–39","noteIndex":0},"citationItems":[{"id":849,"uris":["http://zotero.org/groups/2911188/items/VWXBUIIN"],"itemData":{"id":849,"type":"article-journal","abstract":"An attractor network was trained to compute from word form to semantic representations that were based on subject-generated features. The model was driven largely by higher-order semantic structure. The network simulated two recent experiments that employed items included in its training set (McRae and Boisvert, 1998). In Simulation 1, short stimulus onset asynchrony priming was demonstrated for semantically similar items. Simulation 2 reproduced subtle effects obtained by varying degree of similarity. Two predictions from the model were then tested on human subjects. In Simulation 3 and Experiment 1, the items from Simulation 1 were reversed, and both the network and subjects showed minimally different priming effects in the two directions. In Experiment 2, consistent with attractor networks but contrary to a key aspect of hierarchical spreading activation accounts priming was determined by featural similarity rather than shared superordinate category. It is concluded that semantic-similarity priming is due to featural overlap that is a natural consequence of distributed representations of word meaning.","container-title":"Cognitive Science","DOI":"10.1207/s15516709cog2303\_4","ISSN":"1551-6709","issue":"3","language":"en","note":"\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog2303\_4","page":"371-414","source":"Wiley Online Library","title":"An Attractor Model of Lexical Conceptual Processing: Simulating Semantic Priming","title-short":"An Attractor Model of Lexical Conceptual Processing","volume":"23","author":[{"family":"Cree","given":"George S."},{"family":"McRae","given":"Ken"},{"family":"McNorgan","given":"Chris"}],"issued":{"date-parts":[["1999"]]}}},{"id":848,"uris":["http://zotero.org/groups/2911188/items/KZX4KNJY"],"itemData":{"id":848,"type":"article-journal","abstract":"Current feature-based semantic memory models assume that the semantic representations of concepts differ systematically across living and nonliving categories and that such differences account for the emergence of category-specific semantic deficits in brain-damaged people. To assess some of the different models' main assumptions about structural differences at the semantic feature level in the two major semantic domains, we administrated a feature-listing task to normal young volunteers on 64 concepts drawn from living and nonliving semantic categories. We investigated whether feature correlation, a variable with a crucial role in the emergence of category-specific deficits, should be computed as a concept-dependent or as a concept-independent measure, and we chose the former. We also addressed the issue of a psychological counterpart of feature production frequency. Finally, we analysed the database obtained from the feature-listing tasks, looked at cross-domain differences for correlation, feature frequency, distinctiveness, and feature type, and discussed the implications of these findings for current semantic memory models.","container-title":"Cognitive Neuropsychology","DOI":"10.1080/02643290542000067","ISSN":"0264-3294","issue":"4","note":"publisher: Routledge\n\_eprint: https://doi.org/10.1080/02643290542000067\nPMID: 21049342","page":"515-540","source":"Taylor and Francis+NEJM","title":"Analysis of the semantic representations of living and nonliving concepts: A normative study","title-short":"Analysis of the semantic representations of living and nonliving concepts","volume":"23","author":[{"family":"Zannino","given":"Gian Daniele"},{"family":"Perri","given":"Roberta"},{"family":"Pasqualetti","given":"Patrizio"},{"family":"Caltagirone","given":"Carlo"},{"family":"Carlesimo","given":"Giovanni A."}],"issued":{"date-parts":[["2006",6,1]]}}},{"id":758,"uris":["http://zotero.org/groups/2911188/items/2R8B7YRW"],"itemData":{"id":758,"type":"article-journal","abstract":"A limiting factor in understanding memory and language is often the availability of large numbers of stimuli to use and explore in experimental studies. In this study, we expand on three previous databases of concepts to over 4000 words including nouns, verbs, adjectives, and other parts of speech. Participants in the study were asked to provide lists of features for each concept presented (a semantic feature production task), which were combined with previous research in this area. These feature lists for each concept were then coded into their root word form and affixes (i.e., cat and s for cats) to explore the impact of word form on semantic similarity measures, which are often calculated by comparing concept feature lists (feature overlap). All concept features, coding, and calculated similarity information is provided in a searchable database for easy access and utilization for future researchers when designing experiments that use word stimuli. The final database of word pairs was combined with the Semantic Priming Project to examine the relation of semantic similarity statistics on semantic priming in tandem with other psycholinguistic variables.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-019-01243-z","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1849-1863","source":"Springer Link","title":"English semantic feature production norms: An extended database of 4436 concepts","title-short":"English semantic feature production norms","volume":"51","author":[{"family":"Buchanan","given":"Erin M."},{"family":"Valentine","given":"K. D."},{"family":"Maxwell","given":"Nicholas P."}],"issued":{"date-parts":[["2019",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), association strength (i.e., the probability of a first word eliciting a second word when simply shown the first word)[ITEM CSL\_CITATION {"citationID":"9c4Yi45C","properties":{"formattedCitation":"\\super 33,40\\nosupersub{}","plainCitation":"33,40","noteIndex":0},"citationItems":[{"id":754,"uris":["http://zotero.org/groups/2911188/items/AQRLJE4H"],"itemData":{"id":754,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1115-7","ISSN":"1554-3528","issue":"3","journalAbbreviation":"Behav Res","language":"en","page":"987-1006","source":"DOI.org (Crossref)","title":"The “Small World of Words” English word association norms for over 12,000 cue words","volume":"51","author":[{"family":"De Deyne","given":"Simon"},{"family":"Navarro","given":"Danielle J."},{"family":"Perfors","given":"Amy"},{"family":"Brysbaert","given":"Marc"},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2019",6]]}}},{"id":522,"uris":["http://zotero.org/groups/2860599/items/MS33ZQQ6"],"itemData":{"id":522,"type":"article-journal","abstract":"Preexisting word knowledge is accessed in many cognitive tasks, and this article offers a means for indexing this knowledge so that it can be manipulated or controlled. We offer free association data for 72,000 word pairs, along with over a million entries of related data, such as forward and backward strength, number of competing associates, and printed frequency. A separate file contains the 5,019 normed words, their statistics, and thousands of independently normed rhyme, stem, and fragment cues. Other files providen × n associative networks for more than 4,000 words and a list of idiosyncratic responses for each normed word. The database will be useful for investigators interested in cuing, priming, recognition, network theory, linguistics, and implicit testing applications. They also will be useful for evaluating the predictive value of free association probabilities as compared with other measures, such as similarity ratings and co-occurrence norms. Of several procedures for measuring preexisting strength between two words, the best remains to be determined. The norms may be downloaded fromwww.psychonomic.org/archive/.","container-title":"Behavior Research Methods, Instruments, & Computers","DOI":"10.3758/BF03195588","ISSN":"1532-5970","issue":"3","journalAbbreviation":"Behavior Research Methods, Instruments, & Computers","language":"en","page":"402-407","source":"Springer Link","title":"The University of South Florida free association, rhyme, and word fragment norms","volume":"36","author":[{"family":"Nelson","given":"Douglas L."},{"family":"McEvoy","given":"Cathy L."},{"family":"Schreiber","given":"Thomas A."}],"issued":{"date-parts":[["2004",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), or text co-occurrence (i.e., words are similar because they frequently appear in proximity to one another)[ITEM CSL\_CITATION {"citationID":"5E021NY9","properties":{"formattedCitation":"\\super 41\\uc0\\u8211{}43\\nosupersub{}","plainCitation":"41–43","noteIndex":0},"citationItems":[{"id":847,"uris":["http://zotero.org/groups/2911188/items/RWLDFII2"],"itemData":{"id":847,"type":"article-journal","abstract":"Latent Semantic Analysis (LSA) is a theory and method for extracting and representing the contextual‐usage meaning of words by statistical computations applied to a large corpus of text (Landauer & Dumais, 1997). The underlying idea is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other. The adequacy of LSA's reflection of human knowledge has been established in a variety of ways. For example, its scores overlap those of humans on standard vocabulary and subject matter tests; it mimics human word sorting and category judgments; it simulates word‐word and passage‐word lexical priming data; and, as reported in 3 following articles in this issue, it accurately estimates passage coherence, learnability of passages by individual students, and the quality and quantity of knowledge contained in an essay.","container-title":"Discourse Processes","DOI":"10.1080/01638539809545028","ISSN":"0163-853X","issue":"2-3","note":"publisher: Routledge\n\_eprint: https://doi.org/10.1080/01638539809545028","page":"259-284","source":"Taylor and Francis+NEJM","title":"An introduction to latent semantic analysis","volume":"25","author":[{"family":"Landauer","given":"Thomas K"},{"family":"Foltz","given":"Peter W."},{"family":"Laham","given":"Darrell"}],"issued":{"date-parts":[["1998",1,1]]}}},{"id":845,"uris":["http://zotero.org/groups/2911188/items/L8YIH346"],"itemData":{"id":845,"type":"article-journal","container-title":"Psychological Review","DOI":"10.1037/0033-295X.104.2.211","ISSN":"1939-1471, 0033-295X","issue":"2","journalAbbreviation":"Psychological Review","language":"en","page":"211-240","source":"DOI.org (Crossref)","title":"A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge.","title-short":"A solution to Plato's problem","volume":"104","author":[{"family":"Landauer","given":"Thomas K."},{"family":"Dumais","given":"Susan T."}],"issued":{"date-parts":[["1997"]]}}},{"id":844,"uris":["http://zotero.org/groups/2911188/items/GV83FGZS"],"itemData":{"id":844,"type":"article-journal","abstract":"A procedure that processes a corpus of text and produces numeric vectors containing information about its meanings for each word is presented. This procedure is applied to a large corpus of natural language text taken from Usenet, and the resulting vectors are examined to determine what information is contained within them. These vectors provide the coordinates in a high-dimensional space in which word relationships can be analyzed. Analyses of both vector similarity and multidimensional scaling demonstrate that there is significant semantic information carried in the vectors. A comparison of vector similarity with human reaction times in a single-word priming experiment is presented. These vectors provide the basis for a representational model of semantic memory, hyperspace analogue to language (HAL).","container-title":"Behavior Research Methods, Instruments, & Computers","DOI":"10.3758/BF03204766","ISSN":"1532-5970","issue":"2","journalAbbreviation":"Behavior Research Methods, Instruments, & Computers","language":"en","page":"203-208","source":"Springer Link","title":"Producing high-dimensional semantic spaces from lexical co-occurrence","volume":"28","author":[{"family":"Lund","given":"Kevin"},{"family":"Burgess","given":"Curt"}],"issued":{"date-parts":[["1996",6,1]]}},"locator":"-"}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Each of these computational definitions of similarity can be calculated from normed datasets or text corpora to provide a continuous measure of similarity distance 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[ITEM CSL\_CITATION {"citationID":"jJIUdVrU","properties":{"formattedCitation":"\\super 24\\nosupersub{}","plainCitation":"24","noteIndex":0},"citationItems":[{"id":774,"uris":["http://zotero.org/groups/2911188/items/GCWZAB9X"],"itemData":{"id":774,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-012-0304-z","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1099-1114","source":"DOI.org (Crossref)","title":"The semantic priming project","volume":"45","author":[{"family":"Hutchison","given":"Keith A."},{"family":"Balota","given":"David A."},{"family":"Neely","given":"James H."},{"family":"Cortese","given":"Michael J."},{"family":"Cohen-Shikora","given":"Emily R."},{"family":"Tse","given":"Chi-Shing"},{"family":"Yap","given":"Melvin J."},{"family":"Bengson","given":"Jesse J."},{"family":"Niemeyer","given":"Dale"},{"family":"Buchanan","given":"Erin"}],"issued":{"date-parts":[["2013",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). This project provided data for lexical decision and naming tasks for 1,661 English words and non-words, along with other psycholinguistic measures for future research. The results of the Semantic Priming Project showed 23 ms to 25 ms decreases in word response latencies (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.[ITEM CSL\_CITATION {"citationID":"4xo4C9Vt","properties":{"formattedCitation":"\\super 44\\nosupersub{}","plainCitation":"44","noteIndex":0},"citationItems":[{"id":1896,"uris":["http://zotero.org/groups/2911188/items/LPPKP2N3"],"itemData":{"id":1896,"type":"article-journal","abstract":"The current study examines the underlying processes of semantic priming using the largest priming database available (i.e., Semantic Priming Project, Hutchison et al. Behavior Research Methods, 45(4), 1099–1114, 2013). Specifically, it compares priming effects in two tasks: lexical decision and pronunciation. Task similarities were assessed at two different stimulus onset asynchronies (SOAs) (i.e., 200 and 1,200 ms) and for both primary and other associates. To evaluate how consistent priming is across these two tasks, item-level priming effects obtained in each task were correlated for each condition separately. The results revealed significant correlations at the short SOA for both primary and other associates. The correlations at the long SOA were significantly smaller and only reached significance when z-transformed response times were used. Furthermore, this pattern remained essentially the same when only asymmetric forward associates (e.g., panda-bear) were considered, suggesting that the cross-task stability at the short SOA was not merely caused by retrospective processes such as semantic matching. Instead, these findings provide evidence for a rapidly operating, item-based, relational characteristic such as spreading activation.","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/s13423-015-0932-2","ISSN":"1531-5320","issue":"2","journalAbbreviation":"Psychon Bull Rev","language":"en","page":"540-547","source":"Springer Link","title":"Uncovering underlying processes of semantic priming by correlating item-level effects","volume":"23","author":[{"family":"Heyman","given":"Tom"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2016",4,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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[ITEM CSL\_CITATION {"citationID":"FFwsZEgY","properties":{"formattedCitation":"\\super 45\\nosupersub{}","plainCitation":"45","noteIndex":0},"citationItems":[{"id":799,"uris":["http://zotero.org/groups/2911188/items/35XIQGAZ"],"itemData":{"id":799,"type":"article-journal","abstract":"The current study explores a set of variables that have the potential to predict semantic priming effects for 300 prime–target associates at the item level. Young and older adults performed either lexical decision (LDT) or naming tasks. A multiple regression procedure was used to predict priming based upon prime characteristics, target characteristics, and prime–target semantic similarity. Results indicate that semantic priming (a) can be reliably predicted at an item level; (b) is equivalent in magnitude across standardized measures of priming in LDTs and naming tasks; (c) is greater following quickly recognized primes; (d) is greater in LDTs for targets that produce slow lexical decision latencies; (e) is greater for pairs high in forward associative strength across tasks and across stimulus onset asynchronies (SOAs); (f) is greater for pairs high in backward associative strength in both tasks, but only at a long SOA; and (g) does not vary as a function of estimates from latent semantic analysis (LSA). Based upon these results, it is suggested that researchers take extreme caution in comparing priming effects across different item sets. Moreover, the current findings lend support to spreading activation and feature overlap theories of priming, but do not support priming based upon contextual similarity as captured by LSA.","container-title":"Quarterly Journal of Experimental Psychology","DOI":"10.1080/17470210701438111","ISSN":"1747-0218, 1747-0226","issue":"7","journalAbbreviation":"Quarterly Journal of Experimental Psychology","language":"en","page":"1036-1066","source":"DOI.org (Crossref)","title":"Predicting Semantic Priming at the Item Level","volume":"61","author":[{"family":"Hutchison","given":"Keith A."},{"family":"Balota","given":"David A."},{"family":"Cortese","given":"Michael J."},{"family":"Watson","given":"Jason M."}],"issued":{"date-parts":[["2008",7]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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.[ITEM CSL\_CITATION {"citationID":"pF13SJAo","properties":{"formattedCitation":"\\super 46\\nosupersub{}","plainCitation":"46","noteIndex":0},"citationItems":[{"id":777,"uris":["http://zotero.org/groups/2911188/items/FCGMR7DI"],"itemData":{"id":777,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1040-9","ISSN":"1554-3528","issue":"6","journalAbbreviation":"Behav Res","language":"en","page":"2173-2183","source":"DOI.org (Crossref)","title":"The (un)reliability of item-level semantic priming effects","volume":"50","author":[{"family":"Heyman","given":"Tom"},{"family":"Bruninx","given":"Anke"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2018",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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[ITEM CSL\_CITATION {"citationID":"PUz8c4VO","properties":{"formattedCitation":"\\super 47,48\\nosupersub{}","plainCitation":"47,48","noteIndex":0},"citationItems":[{"id":1898,"uris":["http://zotero.org/groups/2911188/items/QDNYCSFA"],"itemData":{"id":1898,"type":"article-journal","abstract":"Abstract\n Talk of linguistic universals has given cognitive scientists the impression that languages are all built to a common pattern. In fact, there are vanishingly few universals of language in the direct sense that all languages exhibit them. Instead, diversity can be found at almost every level of linguistic organization. This fundamentally changes the object of enquiry from a cognitive science perspective. This target article summarizes decades of cross-linguistic work by typologists and descriptive linguists, showing just how few and unprofound the universal characteristics of language are, once we honestly confront the diversity offered to us by the world's 6,000 to 8,000 languages. After surveying the various uses of “universal,” we illustrate the ways languages vary radically in sound, meaning, and syntactic organization, and then we examine in more detail the core grammatical machinery of recursion, constituency, and grammatical relations. Although there are significant recurrent patterns in organization, these are better explained as stable engineering solutions satisfying multiple design constraints, reflecting both cultural-historical factors and the constraints of human cognition.\n Linguistic diversity then becomes the crucial datum for cognitive science: we are the only species with a communication system that is fundamentally variable at all levels. Recognizing the true extent of structural diversity in human language opens up exciting new research directions for cognitive scientists, offering thousands of different natural experiments given by different languages, with new opportunities for dialogue with biological paradigms concerned with change and diversity, and confronting us with the extraordinary plasticity of the highest human skills.","container-title":"Behavioral and Brain Sciences","DOI":"10.1017/S0140525X0999094X","ISSN":"0140-525X, 1469-1825","issue":"5","journalAbbreviation":"Behav Brain Sci","language":"en","page":"429-448","source":"DOI.org (Crossref)","title":"The myth of language universals: Language diversity and its importance for cognitive science","title-short":"The myth of language universals","volume":"32","author":[{"family":"Evans","given":"Nicholas"},{"family":"Levinson","given":"Stephen C."}],"issued":{"date-parts":[["2009",10]]}}},{"id":1901,"uris":["http://zotero.org/groups/2911188/items/4N89NPQV"],"itemData":{"id":1901,"type":"article-journal","container-title":"Language and Cognitive Processes","DOI":"10.1080/01690960143000164","ISSN":"0169-0965, 1464-0732","issue":"5-6","journalAbbreviation":"Language and Cognitive Processes","language":"en","page":"699-708","source":"DOI.org (Crossref)","title":"Access to lexical representations: Cross-linguistic issues","title-short":"Access to lexical representations","volume":"16","author":[{"family":"Marslen-Wilson","given":"William D."}],"issued":{"date-parts":[["2001",10]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Related concepts that one may consider universal, such as LEFT and RIGHT, are not coded into all languages[ITEM CSL\_CITATION {"citationID":"1oBvuqW7","properties":{"formattedCitation":"\\super 49\\nosupersub{}","plainCitation":"49","noteIndex":0},"citationItems":[{"id":1897,"uris":["http://zotero.org/groups/2911188/items/P7HBRD7Y"],"itemData":{"id":1897,"type":"article-journal","abstract":"Abstract\n The linguistic and cognitive sciences have severely underestimated the degree of linguistic diversity in the world. Part of the reason for this is that we have projected assumptions based on English and familiar languages onto the rest. We focus on some distortions this has introduced, especially in the study of semantics.","container-title":"Behavioral and Brain Sciences","DOI":"10.1017/S0140525X1000018X","ISSN":"0140-525X, 1469-1825","issue":"2-3","journalAbbreviation":"Behav Brain Sci","language":"en","page":"103-103","source":"DOI.org (Crossref)","title":"WEIRD languages have misled us, too","volume":"33","author":[{"family":"Majid","given":"Asifa"},{"family":"Levinson","given":"Stephen C."}],"issued":{"date-parts":[["2010",6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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[ITEM CSL\_CITATION {"citationID":"DVjWBlE8","properties":{"formattedCitation":"\\super 50,51\\nosupersub{}","plainCitation":"50,51","noteIndex":0},"citationItems":[{"id":1900,"uris":["http://zotero.org/groups/2911188/items/S3UPJURF"],"itemData":{"id":1900,"type":"article-journal","container-title":"Journal of Memory and Language","DOI":"10.1016/j.jml.2008.01.003","ISSN":"0749596X","issue":"4","journalAbbreviation":"Journal of Memory and Language","language":"en","page":"916-930","source":"DOI.org (Crossref)","title":"Masked associative/semantic priming effects across languages with highly proficient bilinguals","volume":"58","author":[{"family":"Perea","given":"Manuel"},{"family":"Duñabeitia","given":"Jon Andoni"},{"family":"Carreiras","given":"Manuel"}],"issued":{"date-parts":[["2008",5]]}}},{"id":1902,"uris":["http://zotero.org/groups/2911188/items/7QKRE42S"],"itemData":{"id":1902,"type":"article-journal","container-title":"Journal of Cognitive Psychology","DOI":"10.1080/20445911.2011.589382","ISSN":"2044-5911, 2044-592X","issue":"8","journalAbbreviation":"Journal of Cognitive Psychology","language":"en","page":"942-961","source":"DOI.org (Crossref)","title":"Effects of the degree of meaning similarity on cross-language semantic priming in highly proficient bilinguals","volume":"23","author":[{"family":"Guasch","given":"Marc"},{"family":"Sánchez-Casas","given":"Rosa"},{"family":"Ferré","given":"Pilar"},{"family":"García-Albea","given":"José E."}],"issued":{"date-parts":[["2011",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Therefore, claims about human language are often based on a small set of languages, limiting the generalizability of these claims[ITEM CSL\_CITATION {"citationID":"3SU166RZ","properties":{"formattedCitation":"\\super 52\\nosupersub{}","plainCitation":"52","noteIndex":0},"citationItems":[{"id":1899,"uris":["http://zotero.org/groups/2911188/items/WVA85W78"],"itemData":{"id":1899,"type":"article-journal","container-title":"Language Sciences","DOI":"10.1016/j.langsci.2018.05.010","ISSN":"03880001","journalAbbreviation":"Language Sciences","language":"en","page":"101173","source":"DOI.org (Crossref)","title":"Biases we live by: Anglocentrism in linguistics and cognitive sciences","title-short":"Biases we live by","volume":"76","author":[{"family":"Levisen","given":"Carsten"}],"issued":{"date-parts":[["2019",11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Even with the increase in publication of normed datasets in non-English languages[ITEM CSL\_CITATION {"citationID":"DDv3TzFc","properties":{"formattedCitation":"\\super 20\\nosupersub{}","plainCitation":"20","noteIndex":0},"citationItems":[{"id":753,"uris":["http://zotero.org/groups/2911188/items/HSGWUQ8Y"],"itemData":{"id":753,"type":"article-journal","abstract":"This article presents the Linguistic Annotated Bibliography (LAB) as a searchable Web portal to quickly and easily access reliable database norms, related programs, and variable calculations. These publications were coded by language, number of stimuli, stimuli type (i.e., words, pictures, symbols), keywords (i.e., frequency, semantics, valence), and other useful information. This tool not only allows researchers to search for the specific type of stimuli needed for experiments but also permits the exploration of publication trends across 100 years of research. Details about the portal creation and use are outlined, as well as various analyses of change in publication rates and keywords. In general, advances in computational power have allowed for the increase in dataset size in the recent decades, in addition to an increase in the number of linguistic variables provided in each publication.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1130-8","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1878-1888","source":"Springer Link","title":"LAB: Linguistic Annotated Bibliography – a searchable portal for normed database information","title-short":"LAB","volume":"51","author":[{"family":"Buchanan","given":"Erin M."},{"family":"Valentine","given":"K. D."},{"family":"Maxwell","given":"Nicholas P."}],"issued":{"date-parts":[["2019",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), 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[ITEM CSL\_CITATION {"citationID":"WIrAcPK8","properties":{"formattedCitation":"\\super 53,54\\nosupersub{}","plainCitation":"53,54","noteIndex":0},"citationItems":[{"id":827,"uris":["http://zotero.org/groups/2911188/items/HCF9XE2N"],"itemData":{"id":827,"type":"paper-conference","container-title":"Advances in neural information processing systems","page":"3111–3119","title":"Distributed representations of words and phrases and their compositionality","author":[{"family":"Mikolov","given":"Tomas"},{"family":"Sutskever","given":"Ilya"},{"family":"Chen","given":"Kai"},{"family":"Corrado","given":"Greg S"},{"family":"Dean","given":"Jeff"}],"issued":{"date-parts":[["2013"]]}}},{"id":828,"uris":["http://zotero.org/groups/2911188/items/NWD22SPB"],"itemData":{"id":828,"type":"article-journal","abstract":"We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e. it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.","container-title":"arXiv:1301.3781 [cs]","note":"arXiv: 1301.3781","source":"arXiv.org","title":"Efficient Estimation of Word Representations in Vector Space","URL":"http://arxiv.org/abs/1301.3781","author":[{"family":"Mikolov","given":"Tomas"},{"family":"Chen","given":"Kai"},{"family":"Corrado","given":"Greg"},{"family":"Dean","given":"Jeffrey"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2013",9,6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) and recently published corpora[ITEM CSL\_CITATION {"citationID":"b9bSzGVA","properties":{"formattedCitation":"\\super 23,55\\nosupersub{}","plainCitation":"23,55","noteIndex":0},"citationItems":[{"id":829,"uris":["http://zotero.org/groups/2911188/items/7ZGGXV5T"],"itemData":{"id":829,"type":"article-journal","note":"publisher: European Language Resources Association","title":"Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles","author":[{"family":"Lison","given":"Pierre"},{"family":"Tiedemann","given":"Jörg"}],"issued":{"date-parts":[["2016"]]}}},{"id":858,"uris":["http://zotero.org/groups/2911188/items/KRBMAICU"],"itemData":{"id":858,"type":"article-journal","abstract":"This paper introduces a novel collection of word embeddings, numerical representations of lexical semantics, in 55 languages, trained on a large corpus of pseudo-conversational speech transcriptions from television shows and movies. The embeddings were trained on the OpenSubtitles corpus using the fastText implementation of the skipgram algorithm. Performance comparable with (and in some cases exceeding) embeddings trained on non-conversational (Wikipedia) text is reported on standard benchmark evaluation datasets. A novel evaluation method of particular relevance to psycholinguists is also introduced: prediction of experimental lexical norms in multiple languages. The models, as well as code for reproducing the models and all analyses reported in this paper (implemented as a user-friendly Python package), are freely available at: https://github.com/jvparidon/subs2vec.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-020-01406-3","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"629-655","source":"Springer Link","title":"subs2vec: Word embeddings from subtitles in 55 languages","title-short":"subs2vec","volume":"53","author":[{"family":"Paridon","given":"Jeroen","non-dropping-particle":"van"},{"family":"Thompson","given":"Bill"}],"issued":{"date-parts":[["2021",4,1]]}},"locator":"2"}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), a broader coverage dataset in up to 43 languages is possible. Therefore, this study aims to provide data that complements and extends 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 this 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[ITEM CSL\_CITATION {"citationID":"7lo9l8ub","properties":{"formattedCitation":"\\super 56\\nosupersub{}","plainCitation":"56","noteIndex":0},"citationItems":[{"id":1889,"uris":["http://zotero.org/groups/2911188/items/PKPWKL3M"],"itemData":{"id":1889,"type":"article-journal","abstract":"The model selection literature has been generally poor at reflecting the deep foundations of the Akaike information criterion (AIC) and at making appropriate comparisons to the Bayesian information criterion (BIC). There is a clear philosophy, a sound criterion based in information theory, and a rigorous statistical foundation for AIC. AIC can be justified as Bayesian using a “savvy” prior on models that is a function of sample size and the number of model parameters. Furthermore, BIC can be derived as a non-Bayesian result. Therefore, arguments about using AIC versus BIC for model selection cannot be from a Bayes versus frequentist perspective. The philosophical context of what is assumed about reality, approximating models, and the intent of model-based inference should determine whether AIC or BIC is used. Various facets of such multimodel inference are presented here, particularly methods of model averaging.","container-title":"Sociological Methods & Research","DOI":"10.1177/0049124104268644","ISSN":"0049-1241, 1552-8294","issue":"2","journalAbbreviation":"Sociological Methods & Research","language":"en","page":"261-304","source":"DOI.org (Crossref)","title":"Multimodel Inference: Understanding AIC and BIC in Model Selection","title-short":"Multimodel Inference","volume":"33","author":[{"family":"Burnham","given":"Kenneth P."},{"family":"Anderson","given":"David R."}],"issued":{"date-parts":[["2004",11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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[ITEM CSL\_CITATION {"citationID":"B1rV1iPt","properties":{"formattedCitation":"\\super 57,58\\nosupersub{}","plainCitation":"57,58","noteIndex":0},"citationItems":[{"id":1404,"uris":["http://zotero.org/groups/2911188/items/QPL9GB9Z"],"itemData":{"id":1404,"type":"article-journal","container-title":"European Journal of Cognitive Psychology","DOI":"10.1080/09541449208406187","ISSN":"0954-1446, 1464-0635","issue":"4","journalAbbreviation":"European Journal of Cognitive Psychology","language":"en","page":"253-272","source":"DOI.org (Crossref)","title":"Components of the between-language semantic priming effect","volume":"4","author":[{"family":"Tzelgov","given":"Joseph"},{"family":"Eben-ezra","given":"Sigal"}],"issued":{"date-parts":[["1992",10]]}}},{"id":1405,"uris":["http://zotero.org/groups/2911188/items/BPJA4IN4"],"itemData":{"id":1405,"type":"article-journal","container-title":"Journal of Verbal Learning and Verbal Behavior","DOI":"10.1016/S0022-5371(84)90336-0","ISSN":"00225371","issue":"4","journalAbbreviation":"Journal of Verbal Learning and Verbal Behavior","language":"en","page":"519-539","source":"DOI.org (Crossref)","title":"The bilingual lexicon: Language-specific units in an integrated network","title-short":"The bilingual lexicon","volume":"23","author":[{"family":"Kirsner","given":"Kim"},{"family":"Smith","given":"Marilyn C."},{"family":"Lockhart","given":"R.S."},{"family":"King","given":"M.L."},{"family":"Jain","given":"M."}],"issued":{"date-parts":[["1984",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/).

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[ITEM CSL\_CITATION {"citationID":"19i1VIoP","properties":{"formattedCitation":"\\super 59,60\\nosupersub{}","plainCitation":"59,60","noteIndex":0},"citationItems":[{"id":801,"uris":["http://zotero.org/groups/2911188/items/N3B3EU8F"],"itemData":{"id":801,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BF03192966","ISSN":"1554-351X, 1554-3528","issue":"4","journalAbbreviation":"Behavior Research Methods","language":"en","page":"755-766","source":"DOI.org (Crossref)","title":"Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach","volume":"39","author":[{"family":"Kelley","given":"Ken"}],"issued":{"date-parts":[["2007",11]]}}},{"id":802,"uris":["http://zotero.org/groups/2911188/items/BMSD9AH8"],"itemData":{"id":802,"type":"article-journal","container-title":"Psychological Methods","DOI":"10.1037/met0000127","ISSN":"1939-1463, 1082-989X","issue":"2","journalAbbreviation":"Psychological Methods","language":"en","page":"226-243","source":"DOI.org (Crossref)","title":"Accuracy in parameter estimation for a general class of effect sizes: A sequential approach.","title-short":"Accuracy in parameter estimation for a general class of effect sizes","volume":"23","author":[{"family":"Kelley","given":"Ken"},{"family":"Darku","given":"Francis Bilson"},{"family":"Chattopadhyay","given":"Bhargab"}],"issued":{"date-parts":[["2018",6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) to address the known reliability issues in item-level responding[ITEM CSL\_CITATION {"citationID":"4GLn9dyf","properties":{"formattedCitation":"\\super 44,46\\nosupersub{}","plainCitation":"44,46","noteIndex":0},"citationItems":[{"id":1896,"uris":["http://zotero.org/groups/2911188/items/LPPKP2N3"],"itemData":{"id":1896,"type":"article-journal","abstract":"The current study examines the underlying processes of semantic priming using the largest priming database available (i.e., Semantic Priming Project, Hutchison et al. Behavior Research Methods, 45(4), 1099–1114, 2013). Specifically, it compares priming effects in two tasks: lexical decision and pronunciation. Task similarities were assessed at two different stimulus onset asynchronies (SOAs) (i.e., 200 and 1,200 ms) and for both primary and other associates. To evaluate how consistent priming is across these two tasks, item-level priming effects obtained in each task were correlated for each condition separately. The results revealed significant correlations at the short SOA for both primary and other associates. The correlations at the long SOA were significantly smaller and only reached significance when z-transformed response times were used. Furthermore, this pattern remained essentially the same when only asymmetric forward associates (e.g., panda-bear) were considered, suggesting that the cross-task stability at the short SOA was not merely caused by retrospective processes such as semantic matching. Instead, these findings provide evidence for a rapidly operating, item-based, relational characteristic such as spreading activation.","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/s13423-015-0932-2","ISSN":"1531-5320","issue":"2","journalAbbreviation":"Psychon Bull Rev","language":"en","page":"540-547","source":"Springer Link","title":"Uncovering underlying processes of semantic priming by correlating item-level effects","volume":"23","author":[{"family":"Heyman","given":"Tom"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2016",4,1]]}}},{"id":777,"uris":["http://zotero.org/groups/2911188/items/FCGMR7DI"],"itemData":{"id":777,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1040-9","ISSN":"1554-3528","issue":"6","journalAbbreviation":"Behav Res","language":"en","page":"2173-2183","source":"DOI.org (Crossref)","title":"The (un)reliability of item-level semantic priming effects","volume":"50","author":[{"family":"Heyman","given":"Tom"},{"family":"Bruninx","given":"Anke"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2018",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) to support Hypothesis 1. The items will be selected using new computational techniques for addressing semantic similarity[ITEM CSL\_CITATION {"citationID":"I5ZHdTeM","properties":{"formattedCitation":"\\super 53,54\\nosupersub{}","plainCitation":"53,54","noteIndex":0},"citationItems":[{"id":827,"uris":["http://zotero.org/groups/2911188/items/HCF9XE2N"],"itemData":{"id":827,"type":"paper-conference","container-title":"Advances in neural information processing systems","page":"3111–3119","title":"Distributed representations of words and phrases and their compositionality","author":[{"family":"Mikolov","given":"Tomas"},{"family":"Sutskever","given":"Ilya"},{"family":"Chen","given":"Kai"},{"family":"Corrado","given":"Greg S"},{"family":"Dean","given":"Jeff"}],"issued":{"date-parts":[["2013"]]}}},{"id":828,"uris":["http://zotero.org/groups/2911188/items/NWD22SPB"],"itemData":{"id":828,"type":"article-journal","abstract":"We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e. it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.","container-title":"arXiv:1301.3781 [cs]","note":"arXiv: 1301.3781","source":"arXiv.org","title":"Efficient Estimation of Word Representations in Vector Space","URL":"http://arxiv.org/abs/1301.3781","author":[{"family":"Mikolov","given":"Tomas"},{"family":"Chen","given":"Kai"},{"family":"Corrado","given":"Greg"},{"family":"Dean","given":"Jeffrey"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2013",9,6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) with recently available large corpora of movie subtitles[ITEM CSL\_CITATION {"citationID":"H5SkNuLN","properties":{"formattedCitation":"\\super 23\\nosupersub{}","plainCitation":"23","noteIndex":0},"citationItems":[{"id":829,"uris":["http://zotero.org/groups/2911188/items/7ZGGXV5T"],"itemData":{"id":829,"type":"article-journal","note":"publisher: European Language Resources Association","title":"Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles","author":[{"family":"Lison","given":"Pierre"},{"family":"Tiedemann","given":"Jörg"}],"issued":{"date-parts":[["2016"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) to appropriately match comparable items across languages. As noted in Buchanan et al.[ITEM CSL\_CITATION {"citationID":"SVPmOKkb","properties":{"formattedCitation":"\\super 20\\nosupersub{}","plainCitation":"20","noteIndex":0},"citationItems":[{"id":753,"uris":["http://zotero.org/groups/2911188/items/HSGWUQ8Y"],"itemData":{"id":753,"type":"article-journal","abstract":"This article presents the Linguistic Annotated Bibliography (LAB) as a searchable Web portal to quickly and easily access reliable database norms, related programs, and variable calculations. These publications were coded by language, number of stimuli, stimuli type (i.e., words, pictures, symbols), keywords (i.e., frequency, semantics, valence), and other useful information. This tool not only allows researchers to search for the specific type of stimuli needed for experiments but also permits the exploration of publication trends across 100 years of research. Details about the portal creation and use are outlined, as well as various analyses of change in publication rates and keywords. In general, advances in computational power have allowed for the increase in dataset size in the recent decades, in addition to an increase in the number of linguistic variables provided in each publication.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1130-8","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1878-1888","source":"Springer Link","title":"LAB: Linguistic Annotated Bibliography – a searchable portal for normed database information","title-short":"LAB","volume":"51","author":[{"family":"Buchanan","given":"Erin M."},{"family":"Valentine","given":"K. D."},{"family":"Maxwell","given":"Nicholas P."}],"issued":{"date-parts":[["2019",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), 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.

**Method**

**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 a background on how we plan to estimate 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 is 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[ITEM CSL\_CITATION {"citationID":"CAuBDCcM","properties":{"formattedCitation":"\\super 46\\nosupersub{}","plainCitation":"46","noteIndex":0},"citationItems":[{"id":777,"uris":["http://zotero.org/groups/2911188/items/FCGMR7DI"],"itemData":{"id":777,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1040-9","ISSN":"1554-3528","issue":"6","journalAbbreviation":"Behav Res","language":"en","page":"2173-2183","source":"DOI.org (Crossref)","title":"The (un)reliability of item-level semantic priming effects","volume":"50","author":[{"family":"Heyman","given":"Tom"},{"family":"Bruninx","given":"Anke"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2018",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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 will focus 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 non-word (e.g., GAT). The dependent variable in this study is response latency, and we will use lexical decision data from the English Lexicon Project[ITEM CSL\_CITATION {"citationID":"ah1dSEIE","properties":{"formattedCitation":"\\super 25\\nosupersub{}","plainCitation":"25","noteIndex":0},"citationItems":[{"id":782,"uris":["http://zotero.org/groups/2911188/items/URNUPYHW"],"itemData":{"id":782,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BF03193014","ISSN":"1554-351X, 1554-3528","issue":"3","journalAbbreviation":"Behavior Research Methods","language":"en","page":"445-459","source":"DOI.org (Crossref)","title":"The English Lexicon Project","volume":"39","author":[{"family":"Balota","given":"David A."},{"family":"Yap","given":"Melvin J."},{"family":"Hutchison","given":"Keith A."},{"family":"Cortese","given":"Michael J."},{"family":"Kessler","given":"Brett"},{"family":"Loftis","given":"Bjorn"},{"family":"Neely","given":"James H."},{"family":"Nelson","given":"Douglas L."},{"family":"Simpson","given":"Greg B."},{"family":"Treiman","given":"Rebecca"}],"issued":{"date-parts":[["2007",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) and the Semantic Priming Project[ITEM CSL\_CITATION {"citationID":"CVyBeDg2","properties":{"formattedCitation":"\\super 24\\nosupersub{}","plainCitation":"24","noteIndex":0},"citationItems":[{"id":774,"uris":["http://zotero.org/groups/2911188/items/GCWZAB9X"],"itemData":{"id":774,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-012-0304-z","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1099-1114","source":"DOI.org (Crossref)","title":"The semantic priming project","volume":"45","author":[{"family":"Hutchison","given":"Keith A."},{"family":"Balota","given":"David A."},{"family":"Neely","given":"James H."},{"family":"Cortese","given":"Michael J."},{"family":"Cohen-Shikora","given":"Emily R."},{"family":"Tse","given":"Chi-Shing"},{"family":"Yap","given":"Melvin J."},{"family":"Bengson","given":"Jesse J."},{"family":"Niemeyer","given":"Dale"},{"family":"Buchanan","given":"Erin"}],"issued":{"date-parts":[["2013",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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[ITEM CSL\_CITATION {"citationID":"eaiFyY0j","properties":{"formattedCitation":"\\super 46\\nosupersub{}","plainCitation":"46","noteIndex":0},"citationItems":[{"id":777,"uris":["http://zotero.org/groups/2911188/items/FCGMR7DI"],"itemData":{"id":777,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1040-9","ISSN":"1554-3528","issue":"6","journalAbbreviation":"Behav Res","language":"en","page":"2173-2183","source":"DOI.org (Crossref)","title":"The (un)reliability of item-level semantic priming effects","volume":"50","author":[{"family":"Heyman","given":"Tom"},{"family":"Bruninx","given":"Anke"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2018",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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 will be 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”[ITEM CSL\_CITATION {"citationID":"dEeq0jhY","properties":{"formattedCitation":"\\super 59\\uc0\\u8211{}61\\nosupersub{}","plainCitation":"59–61","noteIndex":0},"citationItems":[{"id":801,"uris":["http://zotero.org/groups/2911188/items/N3B3EU8F"],"itemData":{"id":801,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BF03192966","ISSN":"1554-351X, 1554-3528","issue":"4","journalAbbreviation":"Behavior Research Methods","language":"en","page":"755-766","source":"DOI.org (Crossref)","title":"Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach","volume":"39","author":[{"family":"Kelley","given":"Ken"}],"issued":{"date-parts":[["2007",11]]}}},{"id":802,"uris":["http://zotero.org/groups/2911188/items/BMSD9AH8"],"itemData":{"id":802,"type":"article-journal","container-title":"Psychological Methods","DOI":"10.1037/met0000127","ISSN":"1939-1463, 1082-989X","issue":"2","journalAbbreviation":"Psychological Methods","language":"en","page":"226-243","source":"DOI.org (Crossref)","title":"Accuracy in parameter estimation for a general class of effect sizes: A sequential approach.","title-short":"Accuracy in parameter estimation for a general class of effect sizes","volume":"23","author":[{"family":"Kelley","given":"Ken"},{"family":"Darku","given":"Francis Bilson"},{"family":"Chattopadhyay","given":"Bhargab"}],"issued":{"date-parts":[["2018",6]]}}},{"id":810,"uris":["http://zotero.org/groups/2911188/items/NM37K5U7"],"itemData":{"id":810,"type":"article-journal","container-title":"Annual Review of Psychology","DOI":"10.1146/annurev.psych.59.103006.093735","ISSN":"0066-4308, 1545-2085","issue":"1","journalAbbreviation":"Annu. Rev. Psychol.","language":"en","page":"537-563","source":"DOI.org (Crossref)","title":"Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation","volume":"59","author":[{"family":"Maxwell","given":"Scott E."},{"family":"Kelley","given":"Ken"},{"family":"Rausch","given":"Joseph R."}],"issued":{"date-parts":[["2008",1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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 will pair accuracy in parameter estimation with a sequential testing procedure to adequately sample each item, rather than estimate an overall effect size. Therefore, we will use 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 will be assessed to determine if the item has met the goals for accuracy in parameter estimation as our stopping rule. If so, the item will be 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.

***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[ITEM CSL\_CITATION {"citationID":"VcdF1lbt","properties":{"formattedCitation":"\\super 62\\nosupersub{}","plainCitation":"62","noteIndex":0},"citationItems":[{"id":792,"uris":["http://zotero.org/groups/2911188/items/IN5W7TYG"],"itemData":{"id":792,"type":"article-journal","container-title":"Psychological Bulletin","DOI":"10.1037/0033-2909.125.6.777","ISSN":"1939-1455, 0033-2909","issue":"6","journalAbbreviation":"Psychological Bulletin","language":"en","page":"777-799","source":"DOI.org (Crossref)","title":"Individual differences in information-processing rate and amount: Implications for group differences in response latency.","title-short":"Individual differences in information-processing rate and amount","volume":"125","author":[{"family":"Faust","given":"Mark E."},{"family":"Balota","given":"David A."},{"family":"Spieler","given":"Daniel H."},{"family":"Ferraro","given":"F. Richard"}],"issued":{"date-parts":[["1999"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). The data was then subset for only real word trials that were correctly answered. The average sample size before data reduction 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 are skewed due to the large number of infrequent words in the English Lexicon Project, and we will use 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 are 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 would represent our potentially stopping rule.

The data was 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 will need approximately 50 participants per word. This value will be used as our minimum sample size for a lexical decision task, and the accuracy standard error level will potentially be 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 are expected to be variable, as this priming strength should be predicted by other psycholinguistic variables, such as word relatedness. Therefore, we aim to achieve an accurate representation of lexical decision times, from which priming can then be calculated. However, it should be noted that accurately measured response latencies do not necessarily imply “reliable” priming or difference score data[ITEM CSL\_CITATION {"citationID":"aZFCIJ06","properties":{"formattedCitation":"\\super 63\\nosupersub{}","plainCitation":"63","noteIndex":0},"citationItems":[{"id":1893,"uris":["http://zotero.org/groups/2911188/items/JTL7R4GY"],"itemData":{"id":1893,"type":"article-journal","container-title":"Psychological Bulletin","DOI":"10.1037/h0076158","ISSN":"0033-2909","issue":"1","journalAbbreviation":"Psychological Bulletin","language":"en","page":"85-86","source":"DOI.org (Crossref)","title":"Unreliability of difference scores: A paradox for measurement of change.","title-short":"Unreliability of difference scores","volume":"82","author":[{"family":"Overall","given":"John E."},{"family":"Woodward","given":"J. Arthur"}],"issued":{"date-parts":[["1975"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), but larger sample sizes should provide more evidence of the picture of item-level reliability. We used this 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 is 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 find the maximum sample size of 320 participants for 80% of the items to reach the smaller standard error of 0.06. Therefore, we will use 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 will complete approximately 800 lexical decision trials per session, and each participant only completes 150 of these concepts (75 targets in the related condition, 75 targets in the unrelated condition as cue words are not analyzed) that are 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 usable lexical decision trials per participant times the data retention rate (a conservative estimate of 90%). The final estimate for sample size per language is 741 to 4741 [(1000\*50\*2) / (150\*.90); (1000\*320\*2) / 150\*.90]. The complete code and description of this process are detailed at: <https://osf.io/rxgkf/>.

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 is quite large, we will employ a stopping procedure to ensure participant time and effort is maximized, and data collection is optimized. To summarize, the minimum sample size will be 50 participants per word and the maximum will be 320, which results in 741 to 4741 participants per language based on expected usable trials. Therefore, the total sample size will range from 7410 to 47410 participants for ten languages. After 50 participants, each concept will be examined for standard error, and data collection for that concept will be decreased in probability when the standard error reaches our average criterion of 0.09. Item probability for selection will also be decreased when they reach the maximum proposed sample size (*n* = 320). This process will be automated online and checked in a scheduled subroutine.

While 43 languages have been identified for possible data collection, we plan 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 will 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 will first explain the types of word-pair conditions in a semantic priming study (i.e., related, unrelated, and non-word). Next, we will 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 are shuffled so that the cue word is combined with a target word with which it has a negligible cosine distance similarity (i.e., < .15).

Finally, non-words pair conditions are created by using the Wuggy-like algorithm[ITEM CSL\_CITATION {"citationID":"Jcivz6sZ","properties":{"formattedCitation":"\\super 64\\nosupersub{}","plainCitation":"64","noteIndex":0},"citationItems":[{"id":803,"uris":["http://zotero.org/groups/2911188/items/CJTQA4H7"],"itemData":{"id":803,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BRM.42.3.627","ISSN":"1554-351X, 1554-3528","issue":"3","journalAbbreviation":"Behavior Research Methods","language":"en","page":"627-633","source":"DOI.org (Crossref)","title":"Wuggy: A multilingual pseudoword generator","title-short":"Wuggy","volume":"42","author":[{"family":"Keuleers","given":"Emmanuel"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2010",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) for non-logographic languages. We will consult with at least two native speakers to change one stroke or radical such that the character(s) are a pronounceable word with no meaning by starting from known non-word lists[ITEM CSL\_CITATION {"citationID":"Lg3mDMsR","properties":{"formattedCitation":"\\super 65\\nosupersub{}","plainCitation":"65","noteIndex":0},"citationItems":[{"id":1895,"uris":["http://zotero.org/groups/2911188/items/W8M8HHIJ"],"itemData":{"id":1895,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-016-0810-5","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1503-1519","source":"DOI.org (Crossref)","title":"The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words","title-short":"The Chinese Lexicon Project","volume":"49","author":[{"family":"Tse","given":"Chi-Shing"},{"family":"Yap","given":"Melvin J."},{"family":"Chan","given":"Yuen-Lai"},{"family":"Sze","given":"Wei Ping"},{"family":"Shaoul","given":"Cyrus"},{"family":"Lin","given":"Dan"}],"issued":{"date-parts":[["2017",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Any disagreements between native speakers will be resolved with discussion between these speakers. Each cue and target word were first hyphenated using the syllypackage and LaTeX style hyphenation[ITEM CSL\_CITATION {"citationID":"pU0c6ChM","properties":{"formattedCitation":"\\super 66\\nosupersub{}","plainCitation":"66","noteIndex":0},"citationItems":[{"id":1894,"uris":["http://zotero.org/groups/2911188/items/LUAX74UN"],"itemData":{"id":1894,"type":"book","abstract":"Provides the hyphenation algorithm used for 'TeX'/'LaTeX' and similar software, as proposed by Liang (1983, <https://tug.org/docs/liang/>). Mainly contains the function hyphen() to be used for hyphenation/syllable counting of text objects. It was originally developed for and part of the 'koRpus' package, but later released as a separate package so it's lighter to have this particular functionality available for other packages. Support for various languages needs be added on-the-fly or by plugin packages (<https://undocumeantit.github.io/repos/>); this package does not include any language specific data. Due to some restrictions on CRAN, the full package sources are only available from the project homepage. To ask for help, report bugs, request features, or discuss the development of the package, please subscribe to the koRpus-dev mailing list (<http://korpusml.reaktanz.de>).","source":"R-Packages","title":"sylly: Hyphenation and Syllable Counting for Text Analysis","title-short":"sylly","URL":"https://CRAN.R-project.org/package=sylly","version":"0.1-6","author":[{"family":"Michalke","given":"Meik"}],"accessed":{"date-parts":[["2022",3,7]]},"issued":{"date-parts":[["2020",9,20]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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 will examine each programmatically generated word to ensure they are pronounceable (i.e., phonologically valid) and not pseudo-homophones (i.e., wherein the pronunciation sounds like a real word, KEEP → KEAP)[ITEM CSL\_CITATION {"citationID":"KirAiL3y","properties":{"formattedCitation":"\\super 64\\nosupersub{}","plainCitation":"64","noteIndex":0},"citationItems":[{"id":803,"uris":["http://zotero.org/groups/2911188/items/CJTQA4H7"],"itemData":{"id":803,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BRM.42.3.627","ISSN":"1554-351X, 1554-3528","issue":"3","journalAbbreviation":"Behavior Research Methods","language":"en","page":"627-633","source":"DOI.org (Crossref)","title":"Wuggy: A multilingual pseudoword generator","title-short":"Wuggy","volume":"42","author":[{"family":"Keuleers","given":"Emmanuel"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2010",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). In cases of disagreement, the native speakers will discuss and resolve these inconsistencies. When they have marked a non-word for exclusion, a new non-word will be generated until speakers agree it meets the rules for non-words. Native speakers may also suggest alternatives, which the lead author will check to ensure match to desired non-word characteristics.

To control the ability of participants to anticipate or guess the answers, we will ensure that half the trials will be answered with a word and half with a nonword. Therefore, we will use 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 can be 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 will be 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 is 50% for words-nonwords across all trials, and the relatedness proportion for pairs is 18.8%.

**Similarity calculation**

***Corpora.*** As described in the introduction, the choice of related words based on similarity is key for the study. There are multiple measures of semantic similarity including the cosine between overlapping features[ITEM CSL\_CITATION {"citationID":"8GaaMWu6","properties":{"formattedCitation":"\\super 39\\nosupersub{}","plainCitation":"39","noteIndex":0},"citationItems":[{"id":758,"uris":["http://zotero.org/groups/2911188/items/2R8B7YRW"],"itemData":{"id":758,"type":"article-journal","abstract":"A limiting factor in understanding memory and language is often the availability of large numbers of stimuli to use and explore in experimental studies. In this study, we expand on three previous databases of concepts to over 4000 words including nouns, verbs, adjectives, and other parts of speech. Participants in the study were asked to provide lists of features for each concept presented (a semantic feature production task), which were combined with previous research in this area. These feature lists for each concept were then coded into their root word form and affixes (i.e., cat and s for cats) to explore the impact of word form on semantic similarity measures, which are often calculated by comparing concept feature lists (feature overlap). All concept features, coding, and calculated similarity information is provided in a searchable database for easy access and utilization for future researchers when designing experiments that use word stimuli. The final database of word pairs was combined with the Semantic Priming Project to examine the relation of semantic similarity statistics on semantic priming in tandem with other psycholinguistic variables.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-019-01243-z","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1849-1863","source":"Springer Link","title":"English semantic feature production norms: An extended database of 4436 concepts","title-short":"English semantic feature production norms","volume":"51","author":[{"family":"Buchanan","given":"Erin M."},{"family":"Valentine","given":"K. D."},{"family":"Maxwell","given":"Nicholas P."}],"issued":{"date-parts":[["2019",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), free association probabilities[ITEM CSL\_CITATION {"citationID":"do66vwr0","properties":{"formattedCitation":"\\super 33,40,67\\nosupersub{}","plainCitation":"33,40,67","noteIndex":0},"citationItems":[{"id":522,"uris":["http://zotero.org/groups/2860599/items/MS33ZQQ6"],"itemData":{"id":522,"type":"article-journal","abstract":"Preexisting word knowledge is accessed in many cognitive tasks, and this article offers a means for indexing this knowledge so that it can be manipulated or controlled. We offer free association data for 72,000 word pairs, along with over a million entries of related data, such as forward and backward strength, number of competing associates, and printed frequency. A separate file contains the 5,019 normed words, their statistics, and thousands of independently normed rhyme, stem, and fragment cues. Other files providen × n associative networks for more than 4,000 words and a list of idiosyncratic responses for each normed word. The database will be useful for investigators interested in cuing, priming, recognition, network theory, linguistics, and implicit testing applications. They also will be useful for evaluating the predictive value of free association probabilities as compared with other measures, such as similarity ratings and co-occurrence norms. Of several procedures for measuring preexisting strength between two words, the best remains to be determined. The norms may be downloaded fromwww.psychonomic.org/archive/.","container-title":"Behavior Research Methods, Instruments, & Computers","DOI":"10.3758/BF03195588","ISSN":"1532-5970","issue":"3","journalAbbreviation":"Behavior Research Methods, Instruments, & Computers","language":"en","page":"402-407","source":"Springer Link","title":"The University of South Florida free association, rhyme, and word fragment norms","volume":"36","author":[{"family":"Nelson","given":"Douglas L."},{"family":"McEvoy","given":"Cathy L."},{"family":"Schreiber","given":"Thomas A."}],"issued":{"date-parts":[["2004",8,1]]}}},{"id":754,"uris":["http://zotero.org/groups/2911188/items/AQRLJE4H"],"itemData":{"id":754,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1115-7","ISSN":"1554-3528","issue":"3","journalAbbreviation":"Behav Res","language":"en","page":"987-1006","source":"DOI.org (Crossref)","title":"The “Small World of Words” English word association norms for over 12,000 cue words","volume":"51","author":[{"family":"De Deyne","given":"Simon"},{"family":"Navarro","given":"Danielle J."},{"family":"Perfors","given":"Amy"},{"family":"Brysbaert","given":"Marc"},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2019",6]]}}},{"id":790,"uris":["http://zotero.org/groups/2911188/items/UKZ5IDEM"],"itemData":{"id":790,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-012-0260-7","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"480-498","source":"DOI.org (Crossref)","title":"Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations","volume":"45","author":[{"family":"De Deyne","given":"Simon"},{"family":"Navarro","given":"Daniel J."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2013",6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), and local/global coherence values from network models[ITEM CSL\_CITATION {"citationID":"KZhZXagU","properties":{"formattedCitation":"\\super 35,68\\nosupersub{}","plainCitation":"35,68","noteIndex":0},"citationItems":[{"id":757,"uris":["http://zotero.org/groups/2911188/items/G9NUW9IP"],"itemData":{"id":757,"type":"article-journal","abstract":"Abstract\n Network science is an emerging discipline drawing from sociology, computer science, physics and a number of other fields to examine complex systems in economical, biological, social, and technological domains. To examine these complex systems, nodes are used to represent individual entities, and links are used to represent relationships between entities, forming a web-like structure, or network, of the entire system. The structure that emerges in these complex networks influences the dynamics of that system. We provide a short review of how this mathematical approach has been used to examine the structure found in the phonological lexicon, and of how subsequent psycholinguistic investigations demonstrate that several of the structural characteristics of the phonological network influence various language-related processes, including word retrieval during the recognition and production of spoken words, recovery from instances of failed lexical retrieval, and the acquisition of word-forms. This approach allows researchers to examine the lexicon at the micro-, meso-, and macro-levels, holding much promise for increasing our understanding of language-related processes and representations.","container-title":"Yearbook of the Poznan Linguistic Meeting","DOI":"10.1515/yplm-2015-0007","ISSN":"2449-7525","issue":"1","language":"en","page":"119-138","source":"DOI.org (Crossref)","title":"Using complex networks to understand the mental lexicon","volume":"1","author":[{"family":"Vitevitch","given":"Michael S."},{"family":"Goldstein","given":"Rutherford"},{"family":"Siew","given":"Cynthia S.Q."},{"family":"Castro","given":"Nichol"}],"issued":{"date-parts":[["2014",12,1]]}}},{"id":818,"uris":["http://zotero.org/groups/2911188/items/LCBL9DGL"],"itemData":{"id":818,"type":"article-journal","container-title":"Journal of Experimental Psychology: Learning, Memory, and Cognition","DOI":"10.1037/xlm0000139","ISSN":"1939-1285, 0278-7393","issue":"3","journalAbbreviation":"Journal of Experimental Psychology: Learning, Memory, and Cognition","language":"en","page":"394-410","source":"DOI.org (Crossref)","title":"Spoken word recognition and serial recall of words from components in the phonological network.","volume":"42","author":[{"family":"Siew","given":"Cynthia S. Q."},{"family":"Vitevitch","given":"Michael S."}],"issued":{"date-parts":[["2016"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). However, the underlying data for these calculations is inconsistent across languages. Therefore, one solution is to use the data present in the OpenSubtitles datasets[ITEM CSL\_CITATION {"citationID":"n5M2F8xw","properties":{"formattedCitation":"\\super 23\\nosupersub{}","plainCitation":"23","noteIndex":0},"citationItems":[{"id":829,"uris":["http://zotero.org/groups/2911188/items/7ZGGXV5T"],"itemData":{"id":829,"type":"article-journal","note":"publisher: European Language Resources Association","title":"Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles","author":[{"family":"Lison","given":"Pierre"},{"family":"Tiedemann","given":"Jörg"}],"issued":{"date-parts":[["2016"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) (i.e., a large collection of movie subtitles) to calculate word frequency and cosine distance 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[ITEM CSL\_CITATION {"citationID":"77QEec0l","properties":{"formattedCitation":"\\super 21,69\\uc0\\u8211{}76\\nosupersub{}","plainCitation":"21,69–76","noteIndex":0},"citationItems":[{"id":815,"uris":["http://zotero.org/groups/2911188/items/8MC8HFKJ"],"itemData":{"id":815,"type":"article-journal","abstract":"We examine the use of film subtitles as an approximation of word frequencies in human interactions. Because subtitle files are widely available on the Internet, they may present a fast and easy way to obtain word frequency measures in language registers other than text writing. We compiled a corpus of 52 million French words, coming from a variety of films. Frequency measures based on this corpus compared well to other spoken and written frequency measures, and explained variance in lexical decision times in addition to what is accounted for by the available French written frequency measures.","container-title":"Applied Psycholinguistics","DOI":"10.1017/S014271640707035X","ISSN":"0142-7164, 1469-1817","issue":"4","journalAbbreviation":"Applied Psycholinguistics","language":"en","page":"661-677","source":"DOI.org (Crossref)","title":"The use of film subtitles to estimate word frequencies","volume":"28","author":[{"family":"New","given":"Boris"},{"family":"Brysbaert","given":"Marc"},{"family":"Veronis","given":"Jean"},{"family":"Pallier","given":"Christophe"}],"issued":{"date-parts":[["2007",10]]}}},{"id":769,"uris":["http://zotero.org/groups/2911188/items/4VMU33UE"],"itemData":{"id":769,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BRM.41.4.977","ISSN":"1554-351X, 1554-3528","issue":"4","journalAbbreviation":"Behavior Research Methods","language":"en","page":"977-990","source":"DOI.org (Crossref)","title":"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","title-short":"Moving beyond Kučera and Francis","volume":"41","author":[{"family":"Brysbaert","given":"Marc"},{"family":"New","given":"Boris"}],"issued":{"date-parts":[["2009",11]]}}},{"id":824,"uris":["http://zotero.org/groups/2911188/items/C6YGRPXJ"],"itemData":{"id":824,"type":"article-journal","abstract":"We present word frequencies based on subtitles of British television programmes. We show that the SUBTLEX-UK word frequencies explain more of the variance in the lexical decision times of the British Lexicon Project than the word frequencies based on the British National Corpus and the SUBTLEX-US frequencies. In addition to the word form frequencies, we also present measures of contextual diversity part-of-speech specific word frequencies, word frequencies in children programmes, and word bigram frequencies, giving researchers of British English access to the full range of norms recently made available for other languages. Finally, we introduce a new measure of word frequency, the Zipf scale, which we hope will stop the current misunderstandings of the word frequency effect.","container-title":"Quarterly Journal of Experimental Psychology","DOI":"10.1080/17470218.2013.850521","ISSN":"1747-0218, 1747-0226","issue":"6","journalAbbreviation":"Quarterly Journal of Experimental Psychology","language":"en","page":"1176-1190","source":"DOI.org (Crossref)","title":"Subtlex-UK: A New and Improved Word Frequency Database for British English","title-short":"Subtlex-UK","volume":"67","author":[{"family":"Heuven","given":"Walter J. B.","non-dropping-particle":"van"},{"family":"Mandera","given":"Pawel"},{"family":"Keuleers","given":"Emmanuel"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2014",6]]}}},{"id":804,"uris":["http://zotero.org/groups/2911188/items/K7XP4HKZ"],"itemData":{"id":804,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BRM.42.3.643","ISSN":"1554-351X, 1554-3528","issue":"3","journalAbbreviation":"Behavior Research Methods","language":"en","page":"643-650","source":"DOI.org (Crossref)","title":"SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles","title-short":"SUBTLEX-NL","volume":"42","author":[{"family":"Keuleers","given":"Emmanuel"},{"family":"Brysbaert","given":"Marc"},{"family":"New","given":"Boris"}],"issued":{"date-parts":[["2010",8]]}}},{"id":784,"uris":["http://zotero.org/groups/2911188/items/X7I3CHYV"],"itemData":{"id":784,"type":"article-journal","container-title":"PLoS ONE","DOI":"10.1371/journal.pone.0010729","ISSN":"1932-6203","issue":"6","journalAbbreviation":"PLoS ONE","language":"en","page":"e10729","source":"DOI.org (Crossref)","title":"SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles","title-short":"SUBTLEX-CH","volume":"5","author":[{"family":"Cai","given":"Qing"},{"family":"Brysbaert","given":"Marc"}],"editor":[{"family":"Rodriguez-Fornells","given":"Antoni"}],"issued":{"date-parts":[["2010",6,2]]}},"locator":"-"},{"id":783,"uris":["http://zotero.org/groups/2911188/items/L7PN6WRY"],"itemData":{"id":783,"type":"article-journal","abstract":"We review recent evidence indicating that researchers in experimental psychology may have used suboptimal estimates of word frequency. Word frequency measures should be based on a corpus of at least 20 million words that contains language participants in psychology experiments are likely to have been exposed to. In addition, the quality of word frequency measures should be ascertained by correlating them with behavioral word processing data. When we apply these criteria to the word frequency measures available for the German language, we find that the commonly used Celex frequencies are the least powerful to predict lexical decision times. Better results are obtained with the Leipzig frequencies, the dlexDB frequencies, and the Google Books 2000–2009 frequencies. However, as in other languages the best performance is observed with subtitle-based word frequencies. The SUBTLEX-DE word frequencies collected for the present ms are made available in easy-to-use files and are free for educational purposes.","container-title":"Experimental Psychology","DOI":"10.1027/1618-3169/a000123","ISSN":"1618-3169, 2190-5142","issue":"5","journalAbbreviation":"Experimental Psychology","language":"en","page":"412-424","source":"DOI.org (Crossref)","title":"The Word Frequency Effect: A Review of Recent Developments and Implications for the Choice of Frequency Estimates in German","title-short":"The Word Frequency Effect","volume":"58","author":[{"family":"Brysbaert","given":"Marc"},{"family":"Buchmeier","given":"Matthias"},{"family":"Conrad","given":"Markus"},{"family":"Jacobs","given":"Arthur M."},{"family":"Bölte","given":"Jens"},{"family":"Böhl","given":"Andrea"}],"issued":{"date-parts":[["2011",7,1]]}}},{"id":791,"uris":["http://zotero.org/groups/2911188/items/GEUZKB7R"],"itemData":{"id":791,"type":"article-journal","container-title":"Frontiers in Psychology","DOI":"10.3389/fpsyg.2010.00218","ISSN":"1664-1078","journalAbbreviation":"Front. Psychology","source":"DOI.org (Crossref)","title":"Subtitle-Based Word Frequencies as the Best Estimate of Reading Behavior: The Case of Greek","title-short":"Subtitle-Based Word Frequencies as the Best Estimate of Reading Behavior","URL":"http://journal.frontiersin.org/article/10.3389/fpsyg.2010.00218/abstract","volume":"1","author":[{"family":"Dimitropoulou","given":"Maria"},{"family":"Duñabeitia","given":"Jon Andoni"},{"family":"Avilés","given":"Alberto"},{"family":"Corral","given":"José"},{"family":"Carreiras","given":"Manuel"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2010"]]}}},{"id":809,"uris":["http://zotero.org/groups/2911188/items/5EUV39TM"],"itemData":{"id":809,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-014-0489-4","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"471-483","source":"DOI.org (Crossref)","title":"Subtlex-pl: subtitle-based word frequency estimates for Polish","title-short":"Subtlex-pl","volume":"47","author":[{"family":"Mandera","given":"Paweł"},{"family":"Keuleers","given":"Emmanuel"},{"family":"Wodniecka","given":"Zofia"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2015",6]]}}},{"id":1886,"uris":["http://zotero.org/groups/2911188/items/DGJ4AUDN"],"itemData":{"id":1886,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-013-0326-1","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1246-1258","source":"DOI.org (Crossref)","title":"EsPal: One-stop shopping for Spanish word properties","title-short":"EsPal","volume":"45","author":[{"family":"Duchon","given":"Andrew"},{"family":"Perea","given":"Manuel"},{"family":"Sebastián-Gallés","given":"Nuria"},{"family":"Martí","given":"Antonia"},{"family":"Carreiras","given":"Manuel"}],"issued":{"date-parts":[["2013",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Cosine distance 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 (i.e., Chinese-English mix). We will use 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[ITEM CSL\_CITATION {"citationID":"vKgEo74s","properties":{"formattedCitation":"\\super 77\\nosupersub{}","plainCitation":"77","noteIndex":0},"citationItems":[{"id":859,"uris":["http://zotero.org/groups/2911188/items/9H5KP4VC"],"itemData":{"id":859,"type":"book","abstract":"This natural language processing toolkit provides language-agnostic 'tokenization', 'parts of speech tagging', 'lemmatization' and 'dependency parsing' of raw text. Next to text parsing, the package also allows you to train annotation models based on data of 'treebanks' in 'CoNLL-U' format as provided at <https://universaldependencies.org/format.html>. The techniques are explained in detail in the paper: 'Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe', available at <doi:10.18653/v1/K17-3009>. The toolkit also contains functionalities for commonly used data manipulations on texts which are enriched with the output of the parser. Namely functionalities and algorithms for collocations, token co-occurrence, document term matrix handling, term frequency inverse document frequency calculations, information retrieval metrics (Okapi BM25), handling of multi-word expressions, keyword detection (Rapid Automatic Keyword Extraction, noun phrase extraction, syntactical patterns) sentiment scoring and semantic similarity analysis.","source":"R-Packages","title":"udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit","title-short":"udpipe","URL":"https://CRAN.R-project.org/package=udpipe","version":"0.8.6","author":[{"family":"Wijffels","given":"Jan"},{"family":"BNOSAC","given":""},{"family":"Linguistics","given":"Institute of Formal and Applied"},{"family":"Physics","given":"Faculty of Mathematics","dropping-particle":"and"},{"family":"Prague","given":"Charles University","dropping-particle":"in"},{"family":"Republic","given":"Czech"},{"family":"Straka","given":"Milan"},{"family":"Straková","given":"Jana"}],"accessed":{"date-parts":[["2021",8,24]]},"issued":{"date-parts":[["2021",6,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) 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 allows for the possibility of 43 different languages in this project. See Figure 1 for the model selection process.

***Modeling.*** The subs2vec project[ITEM CSL\_CITATION {"citationID":"bqQw5Ovk","properties":{"formattedCitation":"\\super 55\\nosupersub{}","plainCitation":"55","noteIndex":0},"citationItems":[{"id":858,"uris":["http://zotero.org/groups/2911188/items/KRBMAICU"],"itemData":{"id":858,"type":"article-journal","abstract":"This paper introduces a novel collection of word embeddings, numerical representations of lexical semantics, in 55 languages, trained on a large corpus of pseudo-conversational speech transcriptions from television shows and movies. The embeddings were trained on the OpenSubtitles corpus using the fastText implementation of the skipgram algorithm. Performance comparable with (and in some cases exceeding) embeddings trained on non-conversational (Wikipedia) text is reported on standard benchmark evaluation datasets. A novel evaluation method of particular relevance to psycholinguists is also introduced: prediction of experimental lexical norms in multiple languages. The models, as well as code for reproducing the models and all analyses reported in this paper (implemented as a user-friendly Python package), are freely available at: https://github.com/jvparidon/subs2vec.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-020-01406-3","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"629-655","source":"Springer Link","title":"subs2vec: Word embeddings from subtitles in 55 languages","title-short":"subs2vec","volume":"53","author":[{"family":"Paridon","given":"Jeroen","non-dropping-particle":"van"},{"family":"Thompson","given":"Bill"}],"issued":{"date-parts":[["2021",4,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) used the OpenSubtitles data to create fastText[ITEM CSL\_CITATION {"citationID":"byp9DXDU","properties":{"formattedCitation":"\\super 78\\nosupersub{}","plainCitation":"78","noteIndex":0},"citationItems":[{"id":857,"uris":["http://zotero.org/groups/2911188/items/MFQIKRWP"],"itemData":{"id":857,"type":"article-journal","container-title":"arXiv preprint arXiv:1607.04606","title":"Enriching Word Vectors with Subword Information","author":[{"family":"Bojanowski","given":"Piotr"},{"family":"Grave","given":"Edouard"},{"family":"Joulin","given":"Armand"},{"family":"Mikolov","given":"Tomas"}],"issued":{"date-parts":[["2016"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) computational representation for 55 languages. fastText is a distributional vector space model, an extension of word2vec[ITEM CSL\_CITATION {"citationID":"aU88XiFz","properties":{"formattedCitation":"\\super 53,54\\nosupersub{}","plainCitation":"53,54","noteIndex":0},"citationItems":[{"id":827,"uris":["http://zotero.org/groups/2911188/items/HCF9XE2N"],"itemData":{"id":827,"type":"paper-conference","container-title":"Advances in neural information processing systems","page":"3111–3119","title":"Distributed representations of words and phrases and their compositionality","author":[{"family":"Mikolov","given":"Tomas"},{"family":"Sutskever","given":"Ilya"},{"family":"Chen","given":"Kai"},{"family":"Corrado","given":"Greg S"},{"family":"Dean","given":"Jeff"}],"issued":{"date-parts":[["2013"]]}}},{"id":828,"uris":["http://zotero.org/groups/2911188/items/NWD22SPB"],"itemData":{"id":828,"type":"article-journal","abstract":"We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e. it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.","container-title":"arXiv:1301.3781 [cs]","note":"arXiv: 1301.3781","source":"arXiv.org","title":"Efficient Estimation of Word Representations in Vector Space","URL":"http://arxiv.org/abs/1301.3781","author":[{"family":"Mikolov","given":"Tomas"},{"family":"Chen","given":"Kai"},{"family":"Corrado","given":"Greg"},{"family":"Dean","given":"Jeffrey"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2013",9,6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/)**,** 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[ITEM CSL\_CITATION {"citationID":"OOPoLT1v","properties":{"formattedCitation":"\\super 4,5\\nosupersub{}","plainCitation":"4,5","noteIndex":0},"citationItems":[{"id":787,"uris":["http://zotero.org/groups/2911188/items/7ZKWUIZQ"],"itemData":{"id":787,"type":"chapter","container-title":"The Cambridge Handbook of Psycholinguistics","edition":"1","ISBN":"978-1-139-02937-7","note":"DOI: 10.1017/CBO9781139029377.014","page":"259-282","publisher":"Cambridge University Press","source":"DOI.org (Crossref)","title":"Computational Models of Semantic Memory","URL":"https://www.cambridge.org/core/product/identifier/9781139029377%23c86064-13-1/type/book\_part","editor":[{"family":"Spivey","given":"Michael"},{"family":"McRae","given":"Ken"},{"family":"Joanisse","given":"Marc"}],"author":[{"family":"Cree","given":"George S."},{"family":"Armstrong","given":"Blair C."}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2012",8,20]]}}},{"id":813,"uris":["http://zotero.org/groups/2911188/items/XTNP2K7S"],"itemData":{"id":813,"type":"book","note":"DOI: 10.1093/oxfordhb/9780195376746.013.0014","publisher":"Oxford University Press","source":"DOI.org (Crossref)","title":"Semantic Memory","URL":"http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780195376746.001.0001/oxfordhb-9780195376746-e-14","author":[{"family":"McRae","given":"Ken"},{"family":"Jones","given":"Michael"}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2013",3,11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). We will use the existing models from subs2vec to extract related word concepts for the most frequent concepts identified using the top cosine distance between word vectors.

***Cue selection procedure.*** The procedure for stimuli selection can be viewed at <https://osf.io/s9h3z/> and is displayed graphically in Figure 1. If the language is available via subs2vec, the provided subtitle frequency counts will be examined. If the language has more than 50,000 unique concepts represented in the subtitle data, we will use the subtitle model only. If the subtitles do not provide enough linguistic information (i.e., fewer than 50,000 concepts in the corpus), we will use the combined Wikipedia and subtitle model[ITEM CSL\_CITATION {"citationID":"pbjQmXLL","properties":{"formattedCitation":"\\super 55\\nosupersub{}","plainCitation":"55","noteIndex":0},"citationItems":[{"id":858,"uris":["http://zotero.org/groups/2911188/items/KRBMAICU"],"itemData":{"id":858,"type":"article-journal","abstract":"This paper introduces a novel collection of word embeddings, numerical representations of lexical semantics, in 55 languages, trained on a large corpus of pseudo-conversational speech transcriptions from television shows and movies. The embeddings were trained on the OpenSubtitles corpus using the fastText implementation of the skipgram algorithm. Performance comparable with (and in some cases exceeding) embeddings trained on non-conversational (Wikipedia) text is reported on standard benchmark evaluation datasets. A novel evaluation method of particular relevance to psycholinguists is also introduced: prediction of experimental lexical norms in multiple languages. The models, as well as code for reproducing the models and all analyses reported in this paper (implemented as a user-friendly Python package), are freely available at: https://github.com/jvparidon/subs2vec.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-020-01406-3","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"629-655","source":"Springer Link","title":"subs2vec: Word embeddings from subtitles in 55 languages","title-short":"subs2vec","volume":"53","author":[{"family":"Paridon","given":"Jeroen","non-dropping-particle":"van"},{"family":"Thompson","given":"Bill"}],"issued":{"date-parts":[["2021",4,1]]}},"locator":"2"}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). subs2veccontains 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[ITEM CSL\_CITATION {"citationID":"SU83WMdg","properties":{"formattedCitation":"\\super 21,69\\nosupersub{}","plainCitation":"21,69","noteIndex":0},"citationItems":[{"id":815,"uris":["http://zotero.org/groups/2911188/items/8MC8HFKJ"],"itemData":{"id":815,"type":"article-journal","abstract":"We examine the use of film subtitles as an approximation of word frequencies in human interactions. Because subtitle files are widely available on the Internet, they may present a fast and easy way to obtain word frequency measures in language registers other than text writing. We compiled a corpus of 52 million French words, coming from a variety of films. Frequency measures based on this corpus compared well to other spoken and written frequency measures, and explained variance in lexical decision times in addition to what is accounted for by the available French written frequency measures.","container-title":"Applied Psycholinguistics","DOI":"10.1017/S014271640707035X","ISSN":"0142-7164, 1469-1817","issue":"4","journalAbbreviation":"Applied Psycholinguistics","language":"en","page":"661-677","source":"DOI.org (Crossref)","title":"The use of film subtitles to estimate word frequencies","volume":"28","author":[{"family":"New","given":"Boris"},{"family":"Brysbaert","given":"Marc"},{"family":"Veronis","given":"Jean"},{"family":"Pallier","given":"Christophe"}],"issued":{"date-parts":[["2007",10]]}}},{"id":769,"uris":["http://zotero.org/groups/2911188/items/4VMU33UE"],"itemData":{"id":769,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BRM.41.4.977","ISSN":"1554-351X, 1554-3528","issue":"4","journalAbbreviation":"Behavior Research Methods","language":"en","page":"977-990","source":"DOI.org (Crossref)","title":"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","title-short":"Moving beyond Kučera and Francis","volume":"41","author":[{"family":"Brysbaert","given":"Marc"},{"family":"New","given":"Boris"}],"issued":{"date-parts":[["2009",11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/); 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 will then be tagged for part-of-speech using udpipe*,* selecting tokens that are tagged as nouns, adjectives, adverbs, and verbs. From the udpipeoutput, 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 will be 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[ITEM CSL\_CITATION {"citationID":"5vk2liHQ","properties":{"formattedCitation":"\\super 79\\nosupersub{}","plainCitation":"79","noteIndex":0},"citationItems":[{"id":854,"uris":["http://zotero.org/groups/2911188/items/RVNK88BW"],"itemData":{"id":854,"type":"book","abstract":"Provides multiple sources of stopwords, for use in text analysis and natural language processing.","source":"R-Packages","title":"stopwords: Multilingual Stopword Lists","title-short":"stopwords","URL":"https://CRAN.R-project.org/package=stopwords","version":"2.2","author":[{"family":"Benoit","given":"Kenneth"},{"family":"Muhr","given":"David"},{"family":"Watanabe","given":"Kohei"}],"accessed":{"date-parts":[["2021",9,21]]},"issued":{"date-parts":[["2021",2,10]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). This procedure will cover all but two languages in our list of 43 possible languages. For the final two languages, we will use *udpipe* to tag the OpenSubtitles directly and calculate word frequency. Additionally, fastText model using the same parameters as subs2vec will be trained for similarity calculation. The 10,000 most frequent concepts will be selected at this point.

***Target selection procedure.*** Using the fastText models for each language, we will select the top five cosine distance similarity values for each concept in each language independently, resulting in 50,000 possible cue-target pairs. These will be cross-referenced across languages using Google Translate to create a master list of potential cue-target pairings. The related word pairs (*n* = 1000) will be selected from this list using each cue or target only once, favoring pairs with translations in most languages. Therefore, the selection procedure will be based on the most common cue-target pairs across languages, rather than selecting similar words in one language and then translating. This procedure is programmatic, using Google Translate, which may not produce the most appropriate translation for a word. Therefore, native speakers will ensure 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 will suggest a more common or appropriate word for items they think are unusual, and in cases of disagreement, group discussion between the two translators will be used. In some instances, translation may indicate that a particular language does not have separate concepts for the cue-target pairing. In this instance, we will change the cue word to a related word for that language from the five selected in the original list. Thus, all targetsare matched across languages, and as many cues as possible while avoiding repetition within a cue-target pair.

**Procedure**

We will 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 2. First, participants will complete 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 we plan to combine with the current dataset.

**Implementation**

***Timing software.*** While participants will be naïve to the word pairings, the principal investigator will know the pair combinations during data collection and analysis. A small demonstration of the experiment can be found at: <https://psa007.psysciacc.org/>. The study will be programmed using lab.js[ITEM CSL\_CITATION {"citationID":"g9536bhJ","properties":{"formattedCitation":"\\super 80\\nosupersub{}","plainCitation":"80","noteIndex":0},"citationItems":[{"id":797,"uris":["http://zotero.org/groups/2911188/items/279I436T"],"itemData":{"id":797,"type":"report","abstract":"Web-based data collection is increasingly popular in both experimental and survey-based research, because it is flexible, efficient and location-independent. While dedicated software for laboratory-based experimentation and online surveys is commonplace, researchers looking to implement experiments in the browser have, heretofore, often had to manually construct their studies’ content and logic using code. We introduce lab.js, a free, open-source experiment builder that makes it easy to build experiments for both online and in-laboratory data collection. Through its visual interface, stimuli can be designed and combined into a study without programming, though studies’ appearance and behavior can be fully customized using HTML, CSS and JavaScript code if required. Presentation and response times are kept and measured with high accuracy and precision heretofore unmatched in browser-based studies. Experiments constructed with lab.js can be run directly on a local computer, and published online with ease, with direct deployment to cloud hosting, export to any web server, and integration with popular data collection tools. Studies can also be shared in an editable format, archived, re-used and adapted, enabling effortless, transparent replications, and thus facilitating open, cumulative science. The software is provided free of charge under an open-source license; further information, code and extensive documentation are available from https://lab.js.org/.","genre":"preprint","note":"DOI: 10.31234/osf.io/fqr49","publisher":"PsyArXiv","source":"DOI.org (Crossref)","title":"lab.js: A free, open, online study builder","title-short":"lab.js","URL":"https://osf.io/fqr49","author":[{"family":"Henninger","given":"Felix"},{"family":"Shevchenko","given":"Yury"},{"family":"Mertens","given":"Ulf Kai"},{"family":"Kieslich","given":"Pascal J."},{"family":"Hilbig","given":"Benjamin E."}],"accessed":{"date-parts":[["2021",8,12]]},"issued":{"date-parts":[["2019",1,16]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), which is an online, open-source, study-building software. Precise timing measurement is required for this study, and the lab.js team has documented the accuracy of measurement within their framework[ITEM CSL\_CITATION {"citationID":"8MdEzldE","properties":{"formattedCitation":"\\super 81\\nosupersub{}","plainCitation":"81","noteIndex":0},"citationItems":[{"id":796,"uris":["http://zotero.org/groups/2911188/items/NY5DJ2X4"],"itemData":{"id":796,"type":"speech","event":"Society for Computers in Psychology","event-place":"New Orleans, LA","genre":"Presentation","publisher-place":"New Orleans, LA","title":"Who said browser-based experiments can’t have proper timing? Implementing accurate presentation and response timing in browser","URL":"https://lab.js.org/resources/performance/","author":[{"family":"Henninger","given":"Felix"},{"family":"Shevchenko","given":"Y"},{"family":"Mertens","given":"U. K."},{"family":"Kieslich","given":"P"},{"family":"Hilbig","given":"B. E."}],"issued":{"date-parts":[["2018",11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), and previous work has shown no differences between lab and web-based data collection for response latencies[ITEM CSL\_CITATION {"citationID":"Zr7jpA5S","properties":{"formattedCitation":"\\super 82\\nosupersub{}","plainCitation":"82","noteIndex":0},"citationItems":[{"id":798,"uris":["http://zotero.org/groups/2911188/items/PLKFCTR2"],"itemData":{"id":798,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-015-0678-9","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1718-1724","source":"DOI.org (Crossref)","title":"Reaction time effects in lab- versus Web-based research: Experimental evidence","title-short":"Reaction time effects in lab- versus Web-based research","volume":"48","author":[{"family":"Hilbig","given":"Benjamin E."}],"issued":{"date-parts":[["2016",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). In addition, SPALEX, a large lexical decision database in Spanish, was collected completely online[ITEM CSL\_CITATION {"citationID":"Ifew2YXR","properties":{"formattedCitation":"\\super 26\\nosupersub{}","plainCitation":"26","noteIndex":0},"citationItems":[{"id":781,"uris":["http://zotero.org/groups/2911188/items/8QU8WI3W"],"itemData":{"id":781,"type":"article-journal","container-title":"Frontiers in Psychology","DOI":"10.3389/fpsyg.2018.02156","ISSN":"1664-1078","journalAbbreviation":"Front. Psychol.","page":"2156","source":"DOI.org (Crossref)","title":"SPALEX: A Spanish Lexical Decision Database From a Massive Online Data Collection","title-short":"SPALEX","volume":"9","author":[{"family":"Aguasvivas","given":"Jose Armando"},{"family":"Carreiras","given":"Manuel"},{"family":"Brysbaert","given":"Marc"},{"family":"Mandera","given":"Paweł"},{"family":"Keuleers","given":"Emmanuel"},{"family":"Duñabeitia","given":"Jon Andoni"}],"issued":{"date-parts":[["2018",11,12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). We will recommend 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 are saved when participants take the experiment to examine for potential implementation differences.

Participants will be directed to an online web portal to complete the study, and all data will be retained in the online platform with nightly backups to the server. Participants will be 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 will indicate important aspects about screen size, browser, and timing accuracy. In order to enforce this requirement, participants will be asked to hit the spacebar to continue the study.

***Adaptive stimuli selection.*** At the start of data collection, all presented items will be randomly selected from the larger item pool by equalizing the probability of inclusion equal for all words and non-words (*p* = 1/1000 concepts). After the minimum sample size is collected, each word’s standard error will be checked to determine if the sample size for that item has reached our accuracy criteria. If so, the probability of sampling that item will be decreased by half. Once a concept has reached the maximum required sample size, the probability of sampling will also be decreased by half. This procedure will allow for random sampling of the items that still need participants without eliminating words from the item pool. Therefore, we will ensure that there are always words to randomly select from (i.e., to keep the same procedure and number of trials for all participants) and that the randomization is 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 have reached the stopping criteria or maximum sample size, the probabilities will be equalized. We have set minimum, maximum, and a stopping rule for the initial data collection; however, we will allow data collection after these have been reached and will post updates to the data using a DOI service to allow researchers to cite the specific dataset they used for their research (modeled after the Small World of Words Project[ITEM CSL\_CITATION {"citationID":"wAhmWGsV","properties":{"formattedCitation":"\\super 33\\nosupersub{}","plainCitation":"33","noteIndex":0},"citationItems":[{"id":754,"uris":["http://zotero.org/groups/2911188/items/AQRLJE4H"],"itemData":{"id":754,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1115-7","ISSN":"1554-3528","issue":"3","journalAbbreviation":"Behav Res","language":"en","page":"987-1006","source":"DOI.org (Crossref)","title":"The “Small World of Words” English word association norms for over 12,000 cue words","volume":"51","author":[{"family":"De Deyne","given":"Simon"},{"family":"Navarro","given":"Danielle J."},{"family":"Perfors","given":"Amy"},{"family":"Brysbaert","given":"Marc"},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2019",6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), which is ongoing). All data will be included in our dataset, and the analysis section describes how we will indicate potential data for exclusion. Therefore, data collection will occur in a repeated-measures design in which participants do not see all of the possible stimuli, but do see all the possible conditions (related, unrelated, and non-word pairs). They are blind to the condition each pair is presented in.

**Study Procedure**

***Demographics.*** Participants will be directed to select their first language, which will then direct them to the appropriate translation of the experiment. Participants will be 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) for demographic variables. A flow chart of the procedure is provided in Figure 2.

***Lexical Decision Task.*** Instructions on how to complete a lexical decision task will be shown on the next screen, followed by 10 practice trials. Each trial starts with a fixation cross (+) in the middle of the screen for 500 ms. The stimulus item will then be displayed in the middle of the screen in uppercase san-serif 18-point font (i.e., Arial font, DOG). On the bottom of the screen the possible responses will be 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). Response keys will be mapped such that the “nonword” response option is on the non-dominant hand side of the keyboard, and the “word” response option is on the dominant hand side[ITEM CSL\_CITATION {"citationID":"K1HqdP0A","properties":{"formattedCitation":"\\super 83\\nosupersub{}","plainCitation":"83","noteIndex":0},"citationItems":[{"id":1888,"uris":["http://zotero.org/groups/2911188/items/PWNRBLSK"],"itemData":{"id":1888,"type":"article-journal","container-title":"Psychological Bulletin","DOI":"10.1037/0033-2909.132.3.416","ISSN":"1939-1455, 0033-2909","issue":"3","journalAbbreviation":"Psychological Bulletin","language":"en","page":"416-442","source":"DOI.org (Crossref)","title":"Polarity correspondence: A general principle for performance of speeded binary classification tasks.","title-short":"Polarity correspondence","volume":"132","author":[{"family":"Proctor","given":"Robert W."},{"family":"Cho","given":"Yang Seok"}],"issued":{"date-parts":[["2006"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Participants will be asked for the dominant hand at the beginning of the study to determine the response mapping for their study. Participants will make their choice for each concept, and during the practice trials, they will receive feedback if their answer was correct or incorrect. The next stimulus will appear 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 is showing). Responses will time out after three seconds and move on to the next trial. After 10 trials, participants will see the instruction screen again with a reminder that they will now be doing the real task.

After 100 trials, the participants will be shown a short break screen with the option to continue by hitting the spacebar after 10 seconds. After eight blocks of 100 trials (800 word-nonword decisions), the experiment will end with a thank you screen. On this screen, participants will indicate what type of credit they are receiving for the study (e.g., course credit, payment, no compensation, prize drawing), and they will be given instructions on how to indicate that they have completed the study to the appropriate lab. Participants will be allowed to take the study multiple times as items are randomly selected for inclusion. An estimate for the time required for the study is approximately 30 minutes inclusive of practice trials, reading all instructions, and breaks. This estimate is based on previous studies of lexical decision times[ITEM CSL\_CITATION {"citationID":"Ab5OfrPn","properties":{"formattedCitation":"\\super 25\\nosupersub{}","plainCitation":"25","noteIndex":0},"citationItems":[{"id":782,"uris":["http://zotero.org/groups/2911188/items/URNUPYHW"],"itemData":{"id":782,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/BF03193014","ISSN":"1554-351X, 1554-3528","issue":"3","journalAbbreviation":"Behavior Research Methods","language":"en","page":"445-459","source":"DOI.org (Crossref)","title":"The English Lexicon Project","volume":"39","author":[{"family":"Balota","given":"David A."},{"family":"Yap","given":"Melvin J."},{"family":"Hutchison","given":"Keith A."},{"family":"Cortese","given":"Michael J."},{"family":"Kessler","given":"Brett"},{"family":"Loftis","given":"Bjorn"},{"family":"Neely","given":"James H."},{"family":"Nelson","given":"Douglas L."},{"family":"Simpson","given":"Greg B."},{"family":"Treiman","given":"Rebecca"}],"issued":{"date-parts":[["2007",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), and pilot testing will be used to determine if the number of trials should be reduced to accommodate the 30-minute expected time.

***Comparison to the Semantic Priming Project.*** This procedure is a single stream lexical decision task wherein every concept (cue and target) is judged for lexicality (i.e., word/non-word). 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[ITEM CSL\_CITATION {"citationID":"KBbFNJBY","properties":{"formattedCitation":"\\super 84,85\\nosupersub{}","plainCitation":"84,85","noteIndex":0},"citationItems":[{"id":851,"uris":["http://zotero.org/groups/2911188/items/CSC7PUGM"],"itemData":{"id":851,"type":"article-journal","container-title":"Journal of Experimental Psychology: Learning, Memory, and Cognition","DOI":"10.1037/0278-7393.15.6.1003","ISSN":"1939-1285, 0278-7393","issue":"6","journalAbbreviation":"Journal of Experimental Psychology: Learning, Memory, and Cognition","language":"en","page":"1003-1019","source":"DOI.org (Crossref)","title":"Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching.","title-short":"Semantic priming in the lexical decision task","volume":"15","author":[{"family":"Neely","given":"James H."},{"family":"Keefe","given":"Dennis E."},{"family":"Ross","given":"Kent L."}],"issued":{"date-parts":[["1989"]]}}},{"id":1887,"uris":["http://zotero.org/groups/2911188/items/W2MN22MM"],"itemData":{"id":1887,"type":"article-journal","container-title":"Journal of Experimental Psychology: Learning, Memory, and Cognition","DOI":"10.1037/0278-7393.18.6.1191","ISSN":"1939-1285, 0278-7393","issue":"6","journalAbbreviation":"Journal of Experimental Psychology: Learning, Memory, and Cognition","language":"en","page":"1191-1210","source":"DOI.org (Crossref)","title":"How semantic is automatic semantic priming?","volume":"18","author":[{"family":"Shelton","given":"Jennifer R."},{"family":"Martin","given":"Randi C."}],"issued":{"date-parts":[["1992"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). 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[ITEM CSL\_CITATION {"citationID":"a2e3vopk51j","properties":{"formattedCitation":"\\super 18\\nosupersub{}","plainCitation":"18","noteIndex":0},"citationItems":[{"id":752,"uris":["http://zotero.org/groups/2911188/items/9TM7RT3K"],"itemData":{"id":752,"type":"article-journal","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/BF03212999","ISSN":"1069-9384, 1531-5320","issue":"4","journalAbbreviation":"Psychonomic Bulletin & Review","language":"en","page":"618-630","source":"DOI.org (Crossref)","title":"Semantic priming without association: A meta-analytic review","title-short":"Semantic priming without association","volume":"7","author":[{"family":"Lucas","given":"Margery"}],"issued":{"date-parts":[["2000",12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) provides evidence that priming effect sizes are relatively equal across task type (continuous, masked, paired, and naming), and therefore, we should expect similar results.

***Additional data.*** We will combine available lexical and subject rating data with the priming data. 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. Subjective measures are concept characteristics that are rated by participants, and we will include age of acquisition[ITEM CSL\_CITATION {"citationID":"XGKznuwB","properties":{"formattedCitation":"\\super 86\\uc0\\u8211{}89\\nosupersub{}","plainCitation":"86–89","noteIndex":0},"citationItems":[{"id":866,"uris":["http://zotero.org/groups/2911188/items/PKLNRS6Z"],"itemData":{"id":866,"type":"article-journal","container-title":"Visual Cognition","DOI":"10.1080/13506280544000066","ISSN":"1350-6285, 1464-0716","issue":"7-8","journalAbbreviation":"Visual Cognition","language":"en","page":"789-845","source":"DOI.org (Crossref)","title":"Age of acquisition and lexical processing","volume":"13","author":[{"family":"Johnston","given":"Robert A."},{"family":"Barry","given":"Christopher"}],"issued":{"date-parts":[["2006",5]]}}},{"id":867,"uris":["http://zotero.org/groups/2911188/items/7DYDIBDQ"],"itemData":{"id":867,"type":"article-journal","container-title":"Acta Psychologica","DOI":"10.1016/j.actpsy.2003.11.002","ISSN":"00016918","issue":"1","journalAbbreviation":"Acta Psychologica","language":"en","page":"43-67","source":"DOI.org (Crossref)","title":"Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation","title-short":"Age of acquisition and the cumulative-frequency hypothesis","volume":"115","author":[{"family":"Ghyselinck","given":"Mandy"},{"family":"Lewis","given":"Michael B"},{"family":"Brysbaert","given":"Marc"}],"issued":{"date-parts":[["2004",1]]}}},{"id":1406,"uris":["http://zotero.org/groups/2911188/items/Y776Y2SB"],"itemData":{"id":1406,"type":"article-journal","container-title":"Psychological Bulletin","DOI":"10.1037/0033-2909.131.5.684","ISSN":"1939-1455, 0033-2909","issue":"5","journalAbbreviation":"Psychological Bulletin","language":"en","page":"684-712","source":"DOI.org (Crossref)","title":"Age-of-Acquisition Effects in Word and Picture Identification.","volume":"131","author":[{"family":"Juhasz","given":"Barbara J."}],"issued":{"date-parts":[["2005"]]}}},{"id":1407,"uris":["http://zotero.org/groups/2911188/items/PLQ89P8T"],"itemData":{"id":1407,"type":"article-journal","container-title":"Aphasiology","DOI":"10.1080/02687038.2015.1106439","ISSN":"0268-7038, 1464-5041","issue":"11","journalAbbreviation":"Aphasiology","language":"en","page":"1240-1263","source":"DOI.org (Crossref)","title":"Aphasia and age of acquisition: are early-learned words more resilient?","title-short":"Aphasia and age of acquisition","volume":"30","author":[{"family":"Brysbaert","given":"Marc"},{"family":"Ellis","given":"Andrew W."}],"issued":{"date-parts":[["2016",11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) (approximate age you learned a concept), imageability[ITEM CSL\_CITATION {"citationID":"TGx3l12S","properties":{"formattedCitation":"\\super 90,91\\nosupersub{}","plainCitation":"90,91","noteIndex":0},"citationItems":[{"id":864,"uris":["http://zotero.org/groups/2911188/items/X9SN4ZHW"],"itemData":{"id":864,"type":"article-journal","container-title":"Bulletin of the Psychonomic Society","DOI":"10.3758/BF03337237","ISSN":"0090-5054","issue":"5","journalAbbreviation":"Bull. Psychon. Soc.","language":"en","page":"429-431","source":"DOI.org (Crossref)","title":"Imageability and concreteness","volume":"7","author":[{"family":"Richardson","given":"John T. E."}],"issued":{"date-parts":[["1976",5]]}}},{"id":865,"uris":["http://zotero.org/groups/2911188/items/B7SWE7LU"],"itemData":{"id":865,"type":"article-journal","abstract":"Previous research has shown that the positive effect of imageability upon recall is confined to abstract items. In Experiment I it was found that imageability would affect the recall of concrete items if subjects were instructed to use imagery in their memorizing. This suggested that imagery is not usually employed in remembering concrete items. In Experiment II subjects were asked to categorize items on the basis of their meaning. A majority showed sorting related to the concreteness of the items, but very few showed sorting related to imageability. In Experiment III it was found that the concreteness of an item correlated with the time taken to produce a free associate to it, but that its imageability did not. It was concluded that concreteness is a feature of lexical organization, and not a measure of the image-arousing quality of verbal material.","container-title":"Quarterly Journal of Experimental Psychology","DOI":"10.1080/14640747508400483","ISSN":"0033-555X","issue":"2","journalAbbreviation":"Quarterly Journal of Experimental Psychology","language":"en","page":"235-249","source":"DOI.org (Crossref)","title":"Concreteness and Imageability","volume":"27","author":[{"family":"Richardson","given":"John T. E."}],"issued":{"date-parts":[["1975",5]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) (how easy the concept comes to mind), concreteness[ITEM CSL\_CITATION {"citationID":"GDdrg9Xp","properties":{"formattedCitation":"\\super 92\\nosupersub{}","plainCitation":"92","noteIndex":0},"citationItems":[{"id":863,"uris":["http://zotero.org/groups/2911188/items/F8V5YI66"],"itemData":{"id":863,"type":"article-journal","container-title":"Journal of Experimental Psychology: Learning, Memory, and Cognition","DOI":"10.1037/0278-7393.20.5.1196","ISSN":"1939-1285, 0278-7393","issue":"5","journalAbbreviation":"Journal of Experimental Psychology: Learning, Memory, and Cognition","language":"en","page":"1196-1204","source":"DOI.org (Crossref)","title":"Concreteness effects on memory: When and why?","title-short":"Concreteness effects on memory","volume":"20","author":[{"family":"Paivio","given":"Allan"},{"family":"Walsh","given":"Mary"},{"family":"Bons","given":"Trudy"}],"issued":{"date-parts":[["1994"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) (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)[ITEM CSL\_CITATION {"citationID":"qPFrcgrE","properties":{"formattedCitation":"\\super 27,29\\nosupersub{}","plainCitation":"27,29","noteIndex":0},"citationItems":[{"id":861,"uris":["http://zotero.org/groups/2911188/items/66LM44QH"],"itemData":{"id":861,"type":"article-journal","container-title":"Journal of Behavior Therapy and Experimental Psychiatry","DOI":"10.1016/0005-7916(94)90063-9","ISSN":"00057916","issue":"1","journalAbbreviation":"Journal of Behavior Therapy and Experimental Psychiatry","language":"en","page":"49-59","source":"DOI.org (Crossref)","title":"Measuring emotion: The self-assessment manikin and the semantic differential","title-short":"Measuring emotion","volume":"25","author":[{"family":"Bradley","given":"Margaret M."},{"family":"Lang","given":"Peter J."}],"issued":{"date-parts":[["1994",3]]}}},{"id":862,"uris":["http://zotero.org/groups/2911188/items/RIY3KN6I"],"itemData":{"id":862,"type":"report","publisher":"Technical report C-1, the center for research in psychophysiology …","title":"Affective norms for English words (ANEW): Instruction manual and affective ratings","author":[{"family":"Bradley","given":"Margaret M"},{"family":"Lang","given":"Peter J"}],"issued":{"date-parts":[["1999"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), and familiarity (how well a person knows a concept)[ITEM CSL\_CITATION {"citationID":"ihE48NZs","properties":{"formattedCitation":"\\super 93\\nosupersub{}","plainCitation":"93","noteIndex":0},"citationItems":[{"id":860,"uris":["http://zotero.org/groups/2911188/items/ZBQWE7L2"],"itemData":{"id":860,"type":"article-journal","container-title":"Behavior Research Methods, Instruments, & Computers","DOI":"10.3758/BF03202594","ISSN":"0743-3808, 1532-5970","issue":"1","journalAbbreviation":"Behavior Research Methods, Instruments, & Computers","language":"en","page":"6-10","source":"DOI.org (Crossref)","title":"MRC psycholinguistic database: Machine-usable dictionary, version 2.00","title-short":"MRC psycholinguistic database","volume":"20","author":[{"family":"Wilson","given":"Michael"}],"issued":{"date-parts":[["1988",1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). These variables were selected from the list of most published databases for linguistic data[ITEM CSL\_CITATION {"citationID":"pA5lLhkv","properties":{"formattedCitation":"\\super 20\\nosupersub{}","plainCitation":"20","noteIndex":0},"citationItems":[{"id":753,"uris":["http://zotero.org/groups/2911188/items/HSGWUQ8Y"],"itemData":{"id":753,"type":"article-journal","abstract":"This article presents the Linguistic Annotated Bibliography (LAB) as a searchable Web portal to quickly and easily access reliable database norms, related programs, and variable calculations. These publications were coded by language, number of stimuli, stimuli type (i.e., words, pictures, symbols), keywords (i.e., frequency, semantics, valence), and other useful information. This tool not only allows researchers to search for the specific type of stimuli needed for experiments but also permits the exploration of publication trends across 100 years of research. Details about the portal creation and use are outlined, as well as various analyses of change in publication rates and keywords. In general, advances in computational power have allowed for the increase in dataset size in the recent decades, in addition to an increase in the number of linguistic variables provided in each publication.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-018-1130-8","ISSN":"1554-3528","issue":"4","journalAbbreviation":"Behav Res","language":"en","page":"1878-1888","source":"Springer Link","title":"LAB: Linguistic Annotated Bibliography – a searchable portal for normed database information","title-short":"LAB","volume":"51","author":[{"family":"Buchanan","given":"Erin M."},{"family":"Valentine","given":"K. D."},{"family":"Maxwell","given":"Nicholas P."}],"issued":{"date-parts":[["2019",8,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/).

**Analysis Plan**

An example of the data and processing for English can be found at <https://osf.io/6jmzk/>. Each of the sections described below in the descriptive statistics are available as files for raw and processed data from our OSF page. All data will be archived on our server, and we will use Zenodo (<https://zenodo.org/>) to release versions of the data with citable DOIs given the planned continuation of the project after the initial PSA support.

**Descriptive statistics**

**Participant level data**

We will present descriptive statistics on the participants involved in the study including percentages of gender, education levels, native language, and average age. Information about the device used to complete the study will include percentages of computer operating system, the web browser, and the language locale (i.e., the language the browser defaults to using). Finally, the sample sizes collected by the collaborating labs will be provided. Each of these statistics will be provided for the overall data and the data separated by language.

**Trial level data**

Each language will be saved in a separate file with an item specific trial identification number to allow for matching concepts across languages (i.e., CAT [English] → KATZE [German] → GATTA [Italian]). If a participant leaves the study early (e.g., Internet disconnection, computer crash, closes the study), the data past this point in the study is not recorded, and therefore, the trial level data represents all trials displayed during the experiment. Participants are expected to incorrectly answer trials, and these trials will be marked for exclusion. All timeout trials will be marked as missing values in the final data. No missing values will be imputed.

We will mark for exclusion minimum response latencies of less than 160 ms[ITEM CSL\_CITATION {"citationID":"fkX3Jxlc","properties":{"formattedCitation":"\\super 94\\nosupersub{}","plainCitation":"94","noteIndex":0},"citationItems":[{"id":1892,"uris":["http://zotero.org/groups/2911188/items/2WJ52XM6"],"itemData":{"id":1892,"type":"article-journal","abstract":"In 1952, W. E. Hick published an article in the Quarterly Journal of Experimental Psychology, “On the rate of gain of information.” It played a seminal role in the cognitive revolution and established one of the few widely acknowledged laws in psychology, relating choice reaction time to the number of stimulus–response alternatives (or amount of uncertainty) in a task. We review the historical context in which Hick conducted his study and describe his experiments and theoretical analyses. We discuss the article’s immediate impact on researchers, as well as challenges to and shortcomings of Hick’s law and his analysis, including effects of stimulus–response compatibility, practice, very large set sizes and sequential dependencies. Contemporary modeling developments are also described in detail. Perhaps most impressive about Hick’s law is that it continues to spawn research efforts to the present and that it is regarded as a fundamental law of interface design for human–computer interaction using technologies that did not exist at the time of Hick’s research.","container-title":"Quarterly Journal of Experimental Psychology","DOI":"10.1080/17470218.2017.1322622","ISSN":"1747-0218, 1747-0226","issue":"6","journalAbbreviation":"Quarterly Journal of Experimental Psychology","language":"en","page":"1281-1299","source":"DOI.org (Crossref)","title":"Hick’s law for choice reaction time: A review","title-short":"Hick’s law for choice reaction time","volume":"71","author":[{"family":"Proctor","given":"Robert W"},{"family":"Schneider","given":"Darryl W"}],"issued":{"date-parts":[["2018",6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) (i.e., all trials will be presented in the trial level data for openness, but these will be excluded for analysis and calculations listed below). The response latencies from each participant’s session will then be *z*-scored in line with recommendations from Faust et al.[ITEM CSL\_CITATION {"citationID":"gqGWDZ9V","properties":{"formattedCitation":"\\super 62\\nosupersub{}","plainCitation":"62","noteIndex":0},"citationItems":[{"id":792,"uris":["http://zotero.org/groups/2911188/items/IN5W7TYG"],"itemData":{"id":792,"type":"article-journal","container-title":"Psychological Bulletin","DOI":"10.1037/0033-2909.125.6.777","ISSN":"1939-1455, 0033-2909","issue":"6","journalAbbreviation":"Psychological Bulletin","language":"en","page":"777-799","source":"DOI.org (Crossref)","title":"Individual differences in information-processing rate and amount: Implications for group differences in response latency.","title-short":"Individual differences in information-processing rate and amount","volume":"125","author":[{"family":"Faust","given":"Mark E."},{"family":"Balota","given":"David A."},{"family":"Spieler","given":"Daniel H."},{"family":"Ferraro","given":"F. Richard"}],"issued":{"date-parts":[["1999"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). We will not collect enough data to note if a person takes the experiment multiple times for privacy reasons, but as these would be considered different sessions, the recommended *z*-score procedure should control for participant variability at this level. Therefore, repeated participation would not be detrimental to data collection. Finally, participants' overall proportion of correct answers will be calculated, and participants who do not correctly answer at least 80% of 100 minimum trials will be excluded for item data, priming data, and analysis. The average error in the Semantic Priming Project ranged from 4% to 5%, and this criterion was chosen to include participants who were focused on the task.

We will provide descriptive statistics on the average time to complete the study, the number of trials by word type (word, nonword), the accuracy by word type, and average *z*-scored response latencies by word type (overall, excluding *Z* > 2.5, excluding *Z* > 3.0; see below). These values will be provided for overall results and separated by language.

**Item level data**

The item file will contain lexical information about all stimuli calculated from the OpenSubtitles[ITEM CSL\_CITATION {"citationID":"Aq8ZnYbJ","properties":{"formattedCitation":"\\super 23\\nosupersub{}","plainCitation":"23","noteIndex":0},"citationItems":[{"id":829,"uris":["http://zotero.org/groups/2911188/items/7ZGGXV5T"],"itemData":{"id":829,"type":"article-journal","note":"publisher: European Language Resources Association","title":"Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles","author":[{"family":"Lison","given":"Pierre"},{"family":"Tiedemann","given":"Jörg"}],"issued":{"date-parts":[["2016"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) and subs2vec[ITEM CSL\_CITATION {"citationID":"I5z4Dfuf","properties":{"formattedCitation":"\\super 55\\nosupersub{}","plainCitation":"55","noteIndex":0},"citationItems":[{"id":858,"uris":["http://zotero.org/groups/2911188/items/KRBMAICU"],"itemData":{"id":858,"type":"article-journal","abstract":"This paper introduces a novel collection of word embeddings, numerical representations of lexical semantics, in 55 languages, trained on a large corpus of pseudo-conversational speech transcriptions from television shows and movies. The embeddings were trained on the OpenSubtitles corpus using the fastText implementation of the skipgram algorithm. Performance comparable with (and in some cases exceeding) embeddings trained on non-conversational (Wikipedia) text is reported on standard benchmark evaluation datasets. A novel evaluation method of particular relevance to psycholinguists is also introduced: prediction of experimental lexical norms in multiple languages. The models, as well as code for reproducing the models and all analyses reported in this paper (implemented as a user-friendly Python package), are freely available at: https://github.com/jvparidon/subs2vec.","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-020-01406-3","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"629-655","source":"Springer Link","title":"subs2vec: Word embeddings from subtitles in 55 languages","title-short":"subs2vec","volume":"53","author":[{"family":"Paridon","given":"Jeroen","non-dropping-particle":"van"},{"family":"Thompson","given":"Bill"}],"issued":{"date-parts":[["2021",4,1]]}},"locator":"2"}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) projects (length, frequency, orthographic neighborhood, bigram frequency, orthographic and phonographic Levenshtein distance). The descriptive statistics calculated from the trial level data will then be included: mean response latency, average standardized response latency, sample size, standard errors of response latencies, and accuracy rate. No data will be excluded for being a potential outlier; however, we will recommend a cut-off criterion for absolute value *z*-score outliers at 2.5 and 3.0, and we will calculate these same statistics with those subsets of trials excluded. For all real words, the age of acquisition, imageability, concreteness, valence, dominance, arousal, and familiarity values will be included. These values do not exist for non-words.

We will provide descriptive statistics on the average sample size, average *z*-scored response latencies, and average *SE* for the *z*-scored response latencies by each word type (word, nonword). These values will be calculated for the overall data set, separated by language, and without each level of *z*-score outlier criterion.

**Priming data**

In a separate file, we will also prepare information about priming results which includes the target word, average response latencies, averaged *Z*-scored response latencies, sample sizes, standard errors, and priming response latency. For each item, priming is 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 priming for the word CAT. The similarity scores calculated during stimuli selection will be provided in this file, as well as other popular measures of similarity if they are available in that language. For example, semantic feature overlap norms are also available in Italian[ITEM CSL\_CITATION {"citationID":"yJaNnPLv","properties":{"formattedCitation":"\\super 95\\nosupersub{}","plainCitation":"95","noteIndex":0},"citationItems":[{"id":814,"uris":["http://zotero.org/groups/2911188/items/P6ZA9HVE"],"itemData":{"id":814,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-012-0263-4","ISSN":"1554-3528","issue":"2","journalAbbreviation":"Behav Res","language":"en","page":"440-461","source":"DOI.org (Crossref)","title":"Semantic memory: A feature-based analysis and new norms for Italian","title-short":"Semantic memory","volume":"45","author":[{"family":"Montefinese","given":"Maria"},{"family":"Ambrosini","given":"Ettore"},{"family":"Fairfield","given":"Beth"},{"family":"Mammarella","given":"Nicola"}],"issued":{"date-parts":[["2013",6]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), German[ITEM CSL\_CITATION {"citationID":"WK1rU20S","properties":{"formattedCitation":"\\super 96\\nosupersub{}","plainCitation":"96","noteIndex":0},"citationItems":[{"id":806,"uris":["http://zotero.org/groups/2911188/items/6IL2J85V"],"itemData":{"id":806,"type":"article-journal","container-title":"Behavior Research Methods","DOI":"10.3758/s13428-010-0028-x","ISSN":"1554-3528","issue":"1","journalAbbreviation":"Behav Res","language":"en","page":"97-109","source":"DOI.org (Crossref)","title":"A set of semantic norms for German and Italian","volume":"43","author":[{"family":"Kremer","given":"Gerhard"},{"family":"Baroni","given":"Marco"}],"issued":{"date-parts":[["2011",3]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), Spanish[ITEM CSL\_CITATION {"citationID":"tSErTD1G","properties":{"formattedCitation":"\\super 26\\nosupersub{}","plainCitation":"26","noteIndex":0},"citationItems":[{"id":781,"uris":["http://zotero.org/groups/2911188/items/8QU8WI3W"],"itemData":{"id":781,"type":"article-journal","container-title":"Frontiers in Psychology","DOI":"10.3389/fpsyg.2018.02156","ISSN":"1664-1078","journalAbbreviation":"Front. Psychol.","page":"2156","source":"DOI.org (Crossref)","title":"SPALEX: A Spanish Lexical Decision Database From a Massive Online Data Collection","title-short":"SPALEX","volume":"9","author":[{"family":"Aguasvivas","given":"Jose Armando"},{"family":"Carreiras","given":"Manuel"},{"family":"Brysbaert","given":"Marc"},{"family":"Mandera","given":"Paweł"},{"family":"Keuleers","given":"Emmanuel"},{"family":"Duñabeitia","given":"Jon Andoni"}],"issued":{"date-parts":[["2018",11,12]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/), and Dutch[ITEM CSL\_CITATION {"citationID":"doadawAo","properties":{"formattedCitation":"\\super 97\\nosupersub{}","plainCitation":"97","noteIndex":0},"citationItems":[{"id":817,"uris":["http://zotero.org/groups/2911188/items/FRVH7FAM"],"itemData":{"id":817,"type":"article-journal","container-title":"Behavior Research Methods, Instruments, & Computers","DOI":"10.3758/BF03195597","ISSN":"0743-3808, 1532-5970","issue":"3","journalAbbreviation":"Behavior Research Methods, Instruments, & Computers","language":"en","page":"506-515","source":"DOI.org (Crossref)","title":"Dutch norm data for 13 semantic categories and 338 exemplars","volume":"36","author":[{"family":"Ruts","given":"Wim"},{"family":"De Deyne","given":"Simon"},{"family":"Ameel","given":"Eef"},{"family":"Vanpaemel","given":"Wolf"},{"family":"Verbeemen","given":"Timothy"},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2004",8]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/)**.**

We will provide the average statistics for *z*-score priming, *z-*score unrelated response latency, *z*-score related response latency, sample size for unrelated trials, and sample size for related trials. These values will be calculated overall, by language, and with/without *z*-score level exclusions. Last, we will calculate the participant level priming reliability[ITEM CSL\_CITATION {"citationID":"rv0tOmxJ","properties":{"formattedCitation":"\\super 98\\nosupersub{}","plainCitation":"98","noteIndex":0},"citationItems":[{"id":778,"uris":["http://zotero.org/groups/2911188/items/2T4KQ8RS"],"itemData":{"id":778,"type":"chapter","abstract":"The semantic/associative priming effect refers to the finding of faster recognition times for words preceded by related targets (e.g, cat—DOG), compared to words preceded by unrelated targets (e,g, hat—DOG), Over the past three decades, a voluminous literature has explored the influence of semantic primes on word recognition, and this work has been critical in shaping our understanding of lexical processing, semantic representations, and automatic versus attentional influences. That said, the bulk of the empirical work in the semantic priming literature has focused on group-level performance that averages across participants, despite compelling evidence that individual differences in reading skill and attentional control can moderate semantic priming performance in systematic and interesting ways. The present study takes advantage of the power of the semantic priming project (SPP; Hutchison et al., 2013) to answer two broad, related questions. First, how stable are semantic priming effects, as reflected by within-session reliability (assessed by split-half correlations) and between-session reliability (assessed by test-retest correlations)? Second, assuming that priming effects are reliable, how do they interact with theoretically important constructs such as reading ability and attentional control? Our analyses replicate and extend earlier work by Stolz, Besner, and Carr (2005) by demonstrating that the reliability of semantic priming effects strongly depends on prime-target association strength, and reveal that individuals with more attentional control and reading ability are associated with stronger priming. (PsycINFO Database Record (c) 2019 APA, all rights reserved)","collection-title":"Frontiers of cognitive psychology","container-title":"Big data in cognitive science","event-place":"New York, NY, US","ISBN":"978-1-138-79192-3","page":"203-226","publisher":"Routledge/Taylor & Francis Group","publisher-place":"New York, NY, US","source":"APA PsycNET","title":"Individual differences in semantic priming performance: Insights from the semantic priming project","title-short":"Individual differences in semantic priming performance","author":[{"family":"Yap","given":"Melvin J."},{"family":"Hutchison","given":"Keith A."},{"family":"Tan","given":"Luuan Chin"}],"issued":{"date-parts":[["2017"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) and item-level priming reliability[ITEM CSL\_CITATION {"citationID":"knSX3rUd","properties":{"formattedCitation":"\\super 44\\nosupersub{}","plainCitation":"44","noteIndex":0},"citationItems":[{"id":1896,"uris":["http://zotero.org/groups/2911188/items/LPPKP2N3"],"itemData":{"id":1896,"type":"article-journal","abstract":"The current study examines the underlying processes of semantic priming using the largest priming database available (i.e., Semantic Priming Project, Hutchison et al. Behavior Research Methods, 45(4), 1099–1114, 2013). Specifically, it compares priming effects in two tasks: lexical decision and pronunciation. Task similarities were assessed at two different stimulus onset asynchronies (SOAs) (i.e., 200 and 1,200 ms) and for both primary and other associates. To evaluate how consistent priming is across these two tasks, item-level priming effects obtained in each task were correlated for each condition separately. The results revealed significant correlations at the short SOA for both primary and other associates. The correlations at the long SOA were significantly smaller and only reached significance when z-transformed response times were used. Furthermore, this pattern remained essentially the same when only asymmetric forward associates (e.g., panda-bear) were considered, suggesting that the cross-task stability at the short SOA was not merely caused by retrospective processes such as semantic matching. Instead, these findings provide evidence for a rapidly operating, item-based, relational characteristic such as spreading activation.","container-title":"Psychonomic Bulletin & Review","DOI":"10.3758/s13423-015-0932-2","ISSN":"1531-5320","issue":"2","journalAbbreviation":"Psychon Bull Rev","language":"en","page":"540-547","source":"Springer Link","title":"Uncovering underlying processes of semantic priming by correlating item-level effects","volume":"23","author":[{"family":"Heyman","given":"Tom"},{"family":"Hutchison","given":"Keith A."},{"family":"Storms","given":"Gert"}],"issued":{"date-parts":[["2016",4,1]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/).

**Exclusion summary**

Data will be excluded for the following reasons in this order:

1. Participant level data: the entire participant’s data will be removed from the analyses.
   1. Participant did not indicate at least 18 years of age.
   2. Participant did not complete at least 100 trials.
   3. Participant did not achieve 80% correct.
2. Trial level data: only the individual trials will be removed from the analyses.
   1. Timeout trials (i.e., no response given in 3 s window).
   2. Incorrectly answered trials.
   3. Response latencies shorter than 160 ms.
3. Trial level exclusions dependent on test: trials marked for exclusion that are tested with and without these values in the hypotheses described below.
   1. Response latencies over the absolute value of *Z* = 2.5.
   2. Response latencies over the absolute value of *Z* = 3.0.

**Hypothesis 1**

Hypothesis information is presented in Table 1. Hypothesis 1 predicts semantic facilitation with reduced response latencies for related than unrelated words. Hypothesis 1 will be analyzed by calculating an intercept-only regression model using the *z*-scored priming response latency as the dependent variable. The intercept and its 95% confidence interval will represent the grand mean of the priming effect across all languages. The priming response latency is calculated by taking the average of the unrelated pair *z*-scored response latency minus the related pair response latency within each item. Therefore, values that are positive and greater than zero (e.g., > 0.0001) indicate priming because the related pair had a faster response latency than the unrelated pair. We will determine support for Hypothesis 1 if the lower limit of the confidence interval is greater than zero (i.e., a directional comparison). This process will be repeated for average priming scores calculated without trials that were marked as 2.50 *z*-score outliers and 3.00 *z*-score outliers separately. The decision criteria will remain the same, and we will identify any differences in decisions based on outlier statistics (e.g., priming only occurs when X trials are removed).

**Hypothesis 2**

Hypothesis 2 explores the extent to which these semantic priming effects vary across languages. Therefore, we will calculate a random effects model using the nlme[ITEM CSL\_CITATION {"citationID":"wsrufYAj","properties":{"formattedCitation":"\\super 99\\nosupersub{}","plainCitation":"99","noteIndex":0},"citationItems":[{"id":1716,"uris":["http://zotero.org/groups/2860599/items/XYFGU7TG"],"itemData":{"id":1716,"type":"document","title":"nlme: Linear and nonlinear mixed effects models","URL":"https://cran.r-project.org/package=nlme","author":[{"family":"Pinheiro","given":"J"},{"family":"Bates","given":"Douglas"},{"family":"Debroy","given":"S"},{"family":"Sarkar","given":"D"},{"family":"Team","given":"R Core"}],"issued":{"date-parts":[["2017"]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/) package in *R* wherein the random intercept of language will be added to the overall intercept only model for Hypothesis 1. We will report the standard deviation of the random effect, its 95% confidence interval, the AIC change between models, and the pseudo-*R2*values for the effect size of this parameter[ITEM CSL\_CITATION {"citationID":"fFOklKoS","properties":{"formattedCitation":"\\super 100\\nosupersub{}","plainCitation":"100","noteIndex":0},"citationItems":[{"id":1891,"uris":["http://zotero.org/groups/2911188/items/UBGKW5VE"],"itemData":{"id":1891,"type":"book","abstract":"Tools for performing model selection and model averaging. Automated model selection through subsetting the maximum model, with optional constraints for model inclusion. Model parameter and prediction averaging based on model weights derived from information criteria (AICc and alike) or custom model weighting schemes.","source":"R-Packages","title":"MuMIn: Multi-Model Inference","title-short":"MuMIn","URL":"https://CRAN.R-project.org/package=MuMIn","version":"1.43.17","author":[{"family":"Bartoń","given":"Kamil"}],"accessed":{"date-parts":[["2021",12,29]]},"issued":{"date-parts":[["2020",4,15]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). Results will support significant heterogeneity when the AIC for the random effects model is two points or more less than the AIC for the intercept-only model[ITEM CSL\_CITATION {"citationID":"J9cETfpq","properties":{"formattedCitation":"\\super 56\\nosupersub{}","plainCitation":"56","noteIndex":0},"citationItems":[{"id":1889,"uris":["http://zotero.org/groups/2911188/items/PKPWKL3M"],"itemData":{"id":1889,"type":"article-journal","abstract":"The model selection literature has been generally poor at reflecting the deep foundations of the Akaike information criterion (AIC) and at making appropriate comparisons to the Bayesian information criterion (BIC). There is a clear philosophy, a sound criterion based in information theory, and a rigorous statistical foundation for AIC. AIC can be justified as Bayesian using a “savvy” prior on models that is a function of sample size and the number of model parameters. Furthermore, BIC can be derived as a non-Bayesian result. Therefore, arguments about using AIC versus BIC for model selection cannot be from a Bayes versus frequentist perspective. The philosophical context of what is assumed about reality, approximating models, and the intent of model-based inference should determine whether AIC or BIC is used. Various facets of such multimodel inference are presented here, particularly methods of model averaging.","container-title":"Sociological Methods & Research","DOI":"10.1177/0049124104268644","ISSN":"0049-1241, 1552-8294","issue":"2","journalAbbreviation":"Sociological Methods & Research","language":"en","page":"261-304","source":"DOI.org (Crossref)","title":"Multimodel Inference: Understanding AIC and BIC in Model Selection","title-short":"Multimodel Inference","volume":"33","author":[{"family":"Burnham","given":"Kenneth P."},{"family":"Anderson","given":"David R."}],"issued":{"date-parts":[["2004",11]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}](https://www.zotero.org/). This analysis will be repeated with the 2.50 *z*-score outliers and 3.00 *z*-score outliers excluded. We will include a forest plot of the priming effect and their 95% confidence intervals to visualize the potential heterogeneity in the priming results. Simulations of models within and without variability in the priming effects can be found at <https://osf.io/fbhr8/>.

**Protocol Registration**

Our preregistration for this report can be found at <https://osf.io/u5bp6> (updated 5/31/2022).

**Data Availability**

All raw and processed data will be available for download from the website devoted to this project with backups provided on OSF and Zenodo.

**Code Availability**

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

References

[BIBL {"uncited":[],"omitted":[],"custom":[]} CSL\_BIBLIOGRAPHY](https://www.zotero.org/)

**Acknowledgements**

The authors received no specific funding for this work. Specific author funders had no role in study design, data collection and analysis, decision to publish or preperation of this manuscript. Krystian Barzykowski and Ewa Ilczuk were supported by a grant from the National Science Centre, Poland (2019/35/B/HS6/00528).

**Author contributions**

* 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
* Kelly M. Cuccolo: Data Curation, Investigation, Project Administration, Supervision, Writing - Review & Editing
* Nicholas Coles: Project Administration, Writing - Review & Editing
* Tom Heyman: Conceptualization, Methodology, Project Administration, Writing - Review & Editing
* Aishwarya Iyer: Project Administration, Writing - Review & Editing
* Neil Lewis Jr.: Project Administration, Writing - Review & Editing
* Kim Peters: Project Administration, Writing - Review & Editing
* Niels van Berkel: Project Administration, Software, Writing - Review & Editing
* Anna E. van ’t Veer: Project Administration, Writing - Review & Editing
* Jack E. Taylor: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing
* Maria Montefinese: Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing
* K. D. Valentine: Conceptualization, Writing - Original Draft, Writing - Review & Editing
* Nicholas P. Maxwell: Conceptualization, Writing - Review & Editing
* Belgüzar Nilay Türkan: Investigation, Resources, Writing - Review & Editing
* Glenn P. Williams: Investigation, Writing - Review & Editing
* Juan C. Oliveros-Chacana: Investigation, Resources, Writing - Review & Editing
* Jan Philipp Röer: Investigation, Writing - Review & Editing
* Chiara Fini: Investigation, Writing - Review & Editing
* Oguz A. Acar: Investigation, Writing - Review & Editing
* Joseph P. McFall: Investigation, Writing - Review & Editing
* Ekaterina Pronizius: Investigation, Writing - Review & Editing
* Jordan W. Suchow: Investigation, Writing - Review & Editing
* Luisa Batalha: Investigation, Writing - Review & Editing
* Asil Ali Özdoğru: Investigation, Resources, Writing - Review & Editing
* Hendrik Godbersen: Investigation, Resources, Writing - Review & Editing
* Muhammad Mussaffa Butt: Investigation, Resources, Writing - Review & Editing
* Jacek Buczny: Investigation, Resources, Writing - Review & Editing
* Bastian Jaeger: Investigation, Writing - Review & Editing
* Bradley J. Baker: Investigation, Writing - Review & Editing
* Philip A. Grim II: Investigation, Writing - Review & Editing
* Zainab A. Alsuhaibani: Investigation, Resources, Writing - Review & Editing
* Martín Martínez: Investigation, Resources, Writing - Review & Editing
* John Protzko: Investigation, Writing - Review & Editing
* Dermot Lynott: Investigation, Writing - Review & Editing
* Max Korbmacher: Investigation, Writing - Review & Editing
* Mehmet Peker: Investigation, Resources, Writing - Review & Editing
* Barnaby J.W. Dixson: Investigation, Writing - Review & Editing
* Mahmoud M. Elsherif: Investigation, Resources, Writing - Review & Editing
* Maital Neta: Investigation, Resources, Writing - Review & Editing
* Flavio Azevedo: Investigation, Writing - Review & Editing
* Paulo Roberto dos Santos Ferreira: Investigation, Writing - Review & Editing
* Fredrik Sigfrids: Investigation, Resources, Writing - Review & Editing
* Tiago J S Lima: Investigation, Resources, Writing - Review & Editing
* Sandra J. Geiger: Investigation, Writing - Review & Editing
* Anjali Thapar: Investigation, Writing - Review & Editing
* Manuel Perea: Investigation, Resources, Writing - Review & Editing
* Raluca D. Szekely-Copîndean: Investigation, Resources, Writing - Review & Editing
* Thomas Rhys Evans: NA, Investigation, Writing - Review & Editing
* Steven Verheyen: Investigation, Resources, Writing - Review & Editing
* David Moreau: Investigation, Resources, Writing - Review & Editing
* Ulrich S. Tran: Investigation, Writing - Review & Editing
* Dina Abdel Salam El-Dakhs: Investigation, Resources, Writing - Review & Editing
* Izuchukwu L. G. Ndukaihe: Investigation, Writing - Review & Editing
* Tijana Vesić Pavlović: Investigation, Resources, Writing - Review & Editing
* Debora I. Burin: Investigation, Resources, Writing - Review & Editing
* Patrí­cia Arriaga: Investigation, Resources, Writing - Review & Editing
* Dauren Kasanov: Investigation, Resources, Writing - Review & Editing
* Jacob J. Keech: Investigation, Writing - Review & Editing
* María Fernández-López: Investigation, Resources, Writing - Review & Editing
* Suzanne L. K. Stewart: Investigation, Writing - Review & Editing
* David C. Vaidis: Investigation, Resources, Writing - Review & Editing
* Théo Besson: Investigation, Resources, Writing - Review & Editing
* Carlota Batres: Investigation, Writing - Review & Editing
* Leigh Ann Vaughn: Investigation, Writing - Review & Editing
* Magdalena Senderecka: Investigation, Resources, Writing - Review & Editing
* Claudia Mazzuca: Investigation, Writing - Review & Editing
* Leticia Micheli: Investigation, Resources, Writing - Review & Editing
* Martin R. Vasilev: Investigation, Resources, Writing - Review & Editing
* Kathleen Schmidt: Investigation, Writing - Review & Editing
* Cameron Brick: Investigation, Writing - Review & Editing
* Bruno Schivinski: Investigation, Writing - Review & Editing
* Susana Ruiz-Fernandez: Investigation, Resources, Writing - Review & Editing
* Ewa Ilczuk: Investigation, Writing - Review & Editing
* Carmel A Levitan: Investigation, Writing - Review & Editing
* Emily Higgins: Investigation, Writing - Review & Editing
* Gerit Pfuhl: Investigation, Resources, Writing - Review & Editing
* Jackson G. Lu: Investigation, Writing - Review & Editing
* Miroslav Sirota: Investigation, Writing - Review & Editing
* Zoran Pavlović: Investigation, Resources, Writing - Review & Editing
* Ettore Ambrosini: Investigation, Resources, Writing - Review & Editing
* Nienke Böhm: Investigation, Resources, Writing - Review & Editing
* Aslan Karaaslan: Investigation, Resources, Writing - Review & Editing
* Marietta Papadatou-Pastou: Investigation, Resources, Writing - Review & Editing
* Sezin Öner: Investigation, Resources, Writing - Review & Editing
* Ernest Baskin: Investigation, Writing - Review & Editing
* Kate E. Mulgrew: Investigation, Writing - Review & Editing
* José Luis Ulloa: Investigation, Resources, Writing - Review & Editing
* Ewa Szumowska: Investigation, Resources, Writing - Review & Editing
* Patricia Garrido-Vásquez: Investigation, Writing - Review & Editing
* Krystian Barzykowski: Investigation, Resources, Writing - Review & Editing
* Alexandra I. Kosachenko: Investigation, Resources, Writing - Review & Editing
* Chin Wen Cong: Investigation, Resources, Writing - Review & Editing
* Claus Lamm: Investigation, Writing - Review & Editing
* Andrei Dumbravă: Investigation, Resources, Writing - Review & Editing
* Vanessa Era: Investigation, Writing - Review & Editing
* Luis Carlos Pereira Monteiro: Investigation, Writing - Review & Editing
* Peter R. Mallik: Investigation, Writing - Review & Editing
* Chris Isloi: Investigation, Writing - Review & Editing
* Ali H. Al-Hoorie: Investigation, Resources, Writing - Review & Editing
* Natalia Irrazabal: Investigation, Resources, Writing - Review & Editing
* Yuri G. Pavlov: Investigation, Resources, Writing - Review & Editing
* Anna O. Kuzminska: Investigation, Resources, Writing - Review & Editing
* William E. Davis: Investigation, Writing - Review & EditingSarah E. Fisher: Investigation, Writing - Review & Editing
* Mai Helmy: Investigation, Writing - Review & Editing
* Julia Valeiro Paterlini: Investigation, Resources, Writing - Review & Editing
* Guanxiong Huang: Investigation, Resources, Writing - Review & Editing
* Anna M. Borghi: Investigation, Writing - Review & Editing
* Balazs Aczel: Investigation, Resources, Writing - Review & Editing
* Stefan Stieger: Investigation, Writing - Review & Editing
* S. C. Chen: Investigation, Resources, Writing - Review & Editing
* Laura M. Stevens: Investigation, Writing - Review & Editing
* Christophe Blaison: Investigation, Writing - Review & Editing
* Abigail G. Sanders: Investigation, Writing - Review & Editing
* Robert M. Ross: Investigation, Writing - Review & Editing
* Madeleine P. Ingham: Investigation, Writing - Review & Editing
* Tia C. Bennett: Investigation, Writing - Review & Editing
* Jason Geller: Formal Analysis, Validation, Writing - Review & Editing
* Ogeday Çoker: Investigation, Writing - Review & Editing
* Erin Sievers: Investigation, Writing - Review & Editing
* Christopher R. Chartier: Investigation, Writing - Review & Editing
* Heather D. Flowe: Investigation, Resources, Writing - Review & Editing
* Melissa F. Collof: Investigation, Writing - Review & Editing
* Francesco Foroni: Investigation, Writing - Review & Editing
* Tess M. Atkinson: Investigation, Writing - Review & Editing
* Amanda Kaser: Investigation, Writing - Review & Editing
* Zdenek Meier: Investigation, Writing - Review & Editing
* Nwadiogo Chisom ARINZE: Investigation, Writing - Review & Editing
* Marton Aron Varga: Investigation, Resources, Writing - Review & Editing
* David Willinger: Investigation, Resources, Writing - Review & Editing  
  Rumandeep K. Hayre: Investigation, Writing - Review & Editing
* Miguel A. Vadillo: Investigation, Resources, Writing - Review & Editing
* Otto Loberg: Investigation, Writing - Review & Editing
* Aspasia Eleni Paltoglou: Investigation, Resources, Writing - Review & Editing
* Gianni Ribeiro: Investigation, Writing - Review & Editing
* Roxana-Elena Morariu: Investigation, Writing - Review & Editing
* Timo B. Roettger: Investigation, Resources, Writing - Review & Editing
* Tolga Ergiyen: Investigation, Resources, Writing - Review & Editing
* Maja Becker: Investigation, Resources, Writing - Review & Editing
* Yoann Julliard: Investigation, Writing - Review & Editing
* Fatima Zakra Sahli: Investigation, Resources, Writing - Review & Editing
* Kelly Wolfe: Investigation, Writing - Review & Editing
* Klara Malinakova: Investigation, Writing - Review & Editing
* Michal Parzuchowski: Investigation, Resources, Writing - Review & Editing
* Radka Zidkova: Investigation, Writing - Review & Editing
* Lukas Novak: Investigation, Writing - Review & Editing
* Sarah E MacPherson: Investigation, Writing - Review & Editing
* Christopher L Aberson: Investigation, Writing - Review & Editing
* Wolf Vanpaemel: Investigation, Resources, Writing - Review & Editing
* Bernhard Angele: Investigation, Writing - Review & Editing
* Dominique Muller: Investigation, Writing - Review & Editing
* Elif Gizem Demirag Burak: Investigation, Resources, Writing - Review & Editing
* Peter Tavel: Investigation, Writing - Review & Editing
* Günce Yavuz-Ergiyen: Investigation, Resources, Writing - Review & Editing
* Savannah C. Lewis: Project Administration, Resources, Writing - Review & Editing

**Competing interests**

The authors declare no competing interests.

**Figure 1.** Stimuli selection method flow chart. 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.

Diagram

Description automatically generated

**Figure 2.** Flow chart of the procedure for the study. Within the lexical decision task, participants are given short breaks after 100 trials (i.e., each answer given). The answer choices for that language will always be displayed on the bottom of the screen during the lexical decision task.

**Diagram

Description automatically generated**

**Table 1.**

**Design Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Sampling plan (e.g., power analysis)** | **Analysis Plan** | **Interpretation given to different outcomes** |
| Is semantic priming a non-zero effect? | HA: Response latencies will be faster for related word-pairs in comparison to unrelated word pairs.  H0: 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 HA 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? | **HA: Priming response latencies will be variable between languages (i.e., heterogeneous).**  **H0: 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 HA when the Δ**AIC (intercept-only minus random-intercept) is ≥ 2 points.**  The results will be inconclusive when the Δ**AIC (intercept-only minus random-intercept) is < 2 points.** |

[DOCUMENT\_PREFERENCES {"style":{"styleID":"http://www.zotero.org/styles/nature","hasBibliography":true,"bibliographyStyleHasBeenSet":false},"prefs":{"fieldType":"Field","automaticJournalAbbreviations":false,"delayCitationUpdates":true,"noteType":0,"dontAskDelayCitationUpdates":true},"sessionID":"mQrsWu7Y","zoteroVersion":"6.0.35","dataVersion":4}](https://www.zotero.org/)