**PSA007 (SPAM-L) ETHICS INFORMATION PACK**

This information pack contains what you will need to apply for ethics approval from your review board. If you have any questions, please contact the PSA007 project lead (at 007spaml@gmail.com) or ethics monitor ([k.peters@uq.edu.au](mailto:k.peters@uq.edu.au)). The remainder of this document provides the following supporting documentation to help you prepare your submission to your review board:

**TABLE OF CONTENTS:**

[OSF Page:](#_xexogw9e17ml)

[Project Lead’s IRB:](#_bbc1e0hty2j6)

[Participant Information:](#_jls3jw1aa9gw)

[Review Type:](#_p8kybvqrmpva)

[Project Description:](#_i482x5vpf2ut)

[Project Methodology:](#_nwes7o408ask)

[Consent Form:](#_kyefh2cy8jb5)

[Recruitment Materials:](#_7y343bb3y9g9)

### OSF Page:

<https://osf.io/wrpj4/>

### Project Lead’s IRB:

<https://osf.io/tcxk2/>

* The modified IRB is provided above to ease confusion on what to submit. We updated the instructions between V1 and V2 files. Both approvals are provided if you need it.
* We suggest you use the V2\_IRB\_HU\_modified.docx file (this information is what is provided below) to prepare your information.

### Participant Information:

Number of participants proposed/anticipated: 50,000

Type(s) of participants:

☐ Children (17 or younger) X Adults (18 years of age or older)

☐ Patients in institutions X HU students (18 years of age or older)

☐ Prisoners X Faculty or external collaborators

☐ Pregnant women X Other:

This project is in collaboration with the Psychological Science Accelerator. Each team that contributes to data collection will obtain their own exemptions (i.e., some countries do not have IRB regulations), reliance on the HU IRB, or their own IRB review. Here we are indicating that we will have external collaborators who will select participants in many different ways (their classrooms, paid samples, social networks, etc.).

### Review Type:

**Exempt** **Review** (based on the following categories):

☐ Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

X Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, **unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.**

☐ Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available **or** if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Less likely types in the exempt category include the following:

☐ Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) Public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.

☐ Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture.

### Project Description:

Semantic priming has been studied for nearly fifty years across various experimental manipulations and theoretical frameworks. Critically, the understanding of semantic priming relies on reliable, well-studied stimuli with defined similarity values. In the last twenty years, the publication rates of normed stimuli databases and corpora (i.e., large bodies of text) has exponentially increased. Further, newer computational models of concept representation have been detailed using these databases. Using these newer models, we can define similarity between concepts to create reliable stimuli for study in semantic priming. This research project will meet the need for a database of semantic priming values, particularly in non-English languages that are cross linguistically complete. Large-scale data in this area is sparse, unlike the other published databases found in Buchanan et al. [(2019a)](https://paperpile.com/c/d09zVa/NVFy/?noauthor=1). Therefore, this study aims to provide data that complements and extends the published data, which would encourage research on methodology, item characteristics, models, cross-language consistency in priming, and other theoretical areas that semantic priming has been applied to previously. The global aims of this project include:

1) Create an online framework to collect semantic priming data, modeled after the success of the Small World of Words project [(De Deyne et al., 2019)](https://paperpile.com/c/d09zVa/Sp2f). The online framework would allow data collection from any internet capable computer, thus lowering the burden on research labs to collect data of this nature. The online framework can then be used to deliver updates to the data, even after the conclusion of the initial data collection.

2) Provide a large dataset of response latencies and priming scores for prime and target words in up to 55 languages, as available on the Open Subtitles Project. Further, these prime and target words will be supplemented with variables that are theoretically important for research in cognitive architectures to provide a dataset with less missing data. The dataset provided allows researchers to continue to use these datasets to select carefully controlled stimuli, as well as investigate questions about items, participants, reliability, and language.

### Project Methodology:

**Participants**

Data from the English Lexicon Project [(Balota et al., 2007)](https://paperpile.com/c/d09zVa/o46g) and the Semantic Priming Project [(Hutchison et al., 2013)](https://paperpile.com/c/d09zVa/mfwo) were used to estimate the minimum sample size necessary for the study. The aim of this study is to provide a large dataset, rather than test a hypothesis, so traditional ways to estimate sample size via power and effect size were not applicable. Therefore, an accuracy in parameter estimation approach was employed using the previous data as a metric. In this approach, one focuses on finding a confidence interval around a parameter that would be “sufficiently narrow” [(Kelley, 2007; Kelley, Darku, & Chattopadhyay, 2018; Maxwell, Kelley, & Rausch, 2008)](https://paperpile.com/c/d09zVa/J8iv+vG6z+ZIrT)**.** Both the English Lexicon Project and Semantic Priming Project used a lexical decision task, which will be employed in this study. These data were used to estimate the likely standard errors of lexical decision data for individual words. These values were used as the rubric of accurately measured lexical decision response latencies.

Given proposed standard error value, 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 selected, and samples starting at *n* = 5 to *n* = 400 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. From this calculation, *n* for each target concept was estimated at 100-320 participants. The design of the study, the number of words per session, expected data loss due to incorrect answers, number of target words desired, and number of required conditions were all taken into account and the final estimate for sample size per language is 741 to 4741. The complete code and description of this process is detailed at: [here](https://github.com/doomlab/SPAML/tree/master/parameter_estimation).

This sample size estimation represents a major improvement from previous database collection studies, as many have used the traditional *n* >= 30 as a way to guess at minimum sample size. As indicated, it’s often unclear how to exactly estimate a sample size for these types of studies, and this study will detail that procedure to provide guidance for future work. The upper range of estimated participants is high because of the uncertainty in estimating an “accurate” parameter. 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 minimized. The minimum sample size will be 50 participants per concept or 741 total participants, and the maximum will be 320. After 50 participants, each concept will be examined for standard error, and data collection for that concept will be stopped when the standard error reaches an average of the two metrics found in this exploration (0.06, 0.012; see supplemental material) or 0.09. This process will be automated online and checked in a daily subroutine. From the current simulations, this approximates to 100-150 participants per word, and 1482-2223 participants per language total.

**Materials**

Semantic priming focuses on word-pair relatedness or similarity, and therefore, prime-target pairs are often chosen for their similarity in the related condition. The unrelated condition pairs are then created by shuffling the prime-target pairs so that the prime word is combined with a target word it has no relationship to. Non-words are created by changing one to two letters in a prime or target word to create a nonsense word (*nurse* → *lurse*), with the stipulation that they must be pronounceable and not pseudo-homophones (i.e., wherein the pronunciation sounds like a real word, *keep* → *keap*). Consequently, the choice of related words is key for the study. There are multiple measures of semantic similarity including the cosine between overlapping features [(Buchanan et al., 2019b)](https://paperpile.com/c/d09zVa/eXKH), free association probabilities [(De Deyne, Navarro, & Storms, 2013)](https://paperpile.com/c/d09zVa/C1nC), and local/global coherence values from network models [(Siew & Vitevitch, 2016; Vitevitch et al., 2014)](https://paperpile.com/c/d09zVa/jKpT+6ExH). However, the underlying data for these calculations is spotty across languages. Therefore, one solution is to use the Open Subtitles and subs2vec projects to calculate lexical co-occurrence as a measure of semantic similarity (Lison & Tiedemann, 2016; van Paridon & Thompson, 2020). With the subtitle data, we will take the first 10000 most frequent nouns, adjectives, adverbs, and verbs from each language, and these will be cross-referenced using the *translateR* package [(Lucas & Tingley, 2014)](https://paperpile.com/c/d09zVa/5RQM). Next, a distributional space model for each language will be used from the subs2vec project to identify concepts related to the 10000 most frequent words and to calculate their respective similarity values [(Mandera et al., 2017)](https://paperpile.com/c/d09zVa/hhqi). The top five most related words will be selected, and these will be cross-referenced across languages. Native speakers will be recruited to ensure the accurate translation of word pairs. The related word pairs (*n* = 1000) will be selected from the list using each concept only once, favoring pairs with translations in most languages. If a selected pair does not exist in a language, translation from a Native speaker will be used to create that pair. Words will also be cross referenced for polysemy (i.e., multiple meanings) and these will be restricted when possible. Lastly, concepts will be examined for their relative statistics on lexical measures (length, part of speech, neighborhood, phonemes/morphemes) and subjective measures (age of acquisition, imageability, concreteness, valence, dominance, arousal, and familiarity) because of their known associations with concept representation. Psuedowords will be created by replacing a random letter in the selected words while ensuring this letter matches potential bigrams found in the language. The code for this selection procedure and current wordlist can be found: [here](https://github.com/SemanticPriming/SPAML/tree/master/stimuli).

**Procedure**

A small demonstration of the experiment can be found at: [here](https://open-lab.online/code/PSA%20LDT%20Example/?generate=true). The study will be programmed using lab.js [(Henninger, Shevchenko, Mertens, Kieslich, & Hilbig, 2019)](https://paperpile.com/c/d09zVa/6aux), which is an online, open-source study creation project. Precise timing measurement is required for this study, and the lab.js team has documented the accuracy of measurement within their framework [(Henninger, Shevchenko, Mertens, Kieslich, & Hilbig, 2018)](https://paperpile.com/c/d09zVa/3a8Q), and previous work has shown no differences between lab and web-based data collection for response latencies [(Hilbig, 2016)](https://paperpile.com/c/d09zVa/1857). In addition, SPALEX, a large lexical decision database in Spanish was collected completely online [(Aguasvivas et al., 2018)](https://paperpile.com/c/d09zVa/JEZ6). We will recommend that research labs use Chrome as their browser, however, meta-information about the browser and operating system are saved when participants take the experiment to control for implementation differences. Participants will be directed to an online web portal to take the study, and all data will be retained in the online platform with nightly backups to GitHub. They will be asked to indicate their gender (male, female, other, prefer not to say), year of birth for age, and education level (none, elementary school, high school, bachelors, masters, doctorate) for demographic variables. To continue in the study, they will select their primary language, which will direct them to the appropriate stimuli set. The research lab ID will be collected (to track external collaborators) but no identifying information will be collected about participants (i.e., IP address or other information than listed here for demographics).

Participants will be required to complete the study on a computer, rather than a mobile or tablet device. 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. 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 concept will then be displayed in the middle of the screen in uppercase San-Serif font (i.e., NURSE). On the bottom of the screen the answer choices will be shown as the traditional keys next to the *shift* key depending on the common keyboard layout for that language (i.e., Z and / on a QWERTY keyboard or > and - on a QWERTZ keyboard). These choices will be reversed in half of the subjects, which will be randomly selected at the start of the study to counterbalance word/nonword selection. Participants will enter their choice for each concept, and then the next word will appear with an intertrial interval of 500 ms (i.e., the time between the offset of the first concept and onset of the next concept, when the fixation cross is showing). Responses will time out after 5 seconds and move on to the next trial. After ten 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 six blocks of 100 trials (600 words), 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 (course credit, payment), and they will be given instructions on how to indicate they have completed the study to the appropriate lab. Participants will be allowed to take the study multiple times (see below). These values will be customized based on data collection type (i.e., Mechanical Turk, participant pool, etc.). An estimate for the amount of time required for the study is approximately twenty to thirty minutes including practice trials, instructions, and breaks. We will pilot test the number of stimuli to keep the study under 30 minutes and will lower the number of trials accordingly.

A primary goal of this project is to provide a complete dataset of priming and other important related linguistic variables. 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, such as age of acquisition (approximate age you learned a concept), imageability (how easy the concept is to imagine), concreteness (how concrete is the concept), valence (emotion), arousal, dominance (controlled versus dominated), and familiarity. For concepts that are missing these values in a target language, participants will be asked to provide ratings on a single metric (i.e., they would only see instructions for familiarity or arousal). Each participant will be asked to provide 25-50 ratings of concepts, given the need for a particular language, while also controlling for the length of the task to prevent fatigue in the experiment. These will only be presented at the end of the experiment to prevent interactions with priming effects. We will use the available large databases of these variables to estimate sample size necessary for these ratings using the same simulation procedure detailed above.

**Specifiek Studiemateriaal:**

**Demografische gegevens:**

Wat is je moedertaal?

Kan je ons een klein beetje over jezelf vertellen?

Wat is je gender? Man vrouw anders zeg ik liever niet

In welk jaar ben je geboren? Graag 4 cijfers voor het jaartal invoeren:

Wat is je onderwijsniveau? Middelbare school niet afgerond, middelbare school diploma, MBO diploma, HBO diploma, WO Bachelor diploma, WO Master diploma, Doctoraat (Ph.D), Anders, namelijk:

**Semantisch Primen Taak:**

Dit experiment gaat over hoe mensen woorden verwerken. Je zal gevraagd woorden om woorden te bekijken en deze vervolgens te beoordelen. De woorden zullen in het midden van het scherm getoond worden. De taak is om te bepalen of het woord op het scherm een echt woord of een nep woord is. Als je bijvoorbeeld het echte woord **KOUD** herkent dan druk je op de **KEY** toets voor het echte woord. Als het om een nep woord gaat, bijvoorbeeld **WERM**, dan druk je op de **KEY** toets voor het nep woord. Er is een oefensessie waarin je feedback op je antwoorden zal ontvangen. Het doel is om zo snel mogelijk te reageren terwijl je zo accuraat mogelijk blijft. Je hebt vijf seconden om op ieder woord te reageren. Druk op de **SPATIE BALK** om te beginnen met de oefensessie.

Druk op **KEY** als je een echt woord ziet en **KEY** wanneer je een nep woord ziet.

Goed gedaan! Onthoud, druk op **KEY** als je een echt woord ziet en **KEY** wanneer je een nep woord ziet. Je gaat nu door naar de echte taak. Onthoud dat je zo snel mogelijk moet reageren terwijl je zo accuraat mogelijk blijft. Tijdens de echte taak zal je geen feedback op je antwoorden meer ontvangen. Druk op de **SPATIE BALK** om te beginnen.

Neem alsjeblieft een korte pauze voordat je verder gaat. Je kan op de **SPATIE BALK** drukken wanneer je klaar bent om verder te gaan.

Bedankt voor het afronden van het eerste gedeelte van het experiment. Op het volgende scherm zal je de woorden op bepaalde eigenschappen beoordelen om ons helpen te meten hoe deze woorden in hun moedertaal begrepen worden.

**Woord betekenis taak: (McRae et al., 2005)**

Voor dit onderzoek willen we beter begrijpen hoe mensen de betekenis in woorden lezen. Vul alsjeblieft de eigenschappen van het woord in die je kan bedenken. Voorbeelden van verschillende eigenschappen zijn: hoe het eruitziet, klinkt, ruikt, voelt, smaakt; waarvan het gemaakt is; waarvoor het gebruikt wordt; en waar het vandaan komt. Hier volgt een voorbeeld:

eend: is een vogel, is een dier, waggelt, vliegt, migreert, legt eieren, kwaakt, zwemt, heeft vleugels, heeft een snavel, heeft zwemvliezen, heeft veren, leeft in vijvers, leeft in water, gejaagd door mensen, is eetbaar

Vul deze vragenlijst in op een redelijk snel tempo, maar probeer wel op zijn minst een paar eigenschappen per woord in te vullen. Heel erg bedankt voor het invullen van deze vragenlijst.

**Leeftijd van verkrijgen taak: (Kuperman et al., 2012)**

Geef alsjeblieft aan (in jaren) op welke leeftijd je elk woord op de lijst hebt geleerd. Een schatting van de leeftijd is goed genoeg als je het niet precies weet. Als je de betekenis van een woord niet kent, druk dan op de **X** toets. Met het ‘leren van een woord’ bedoelen we de leeftijd waarop je het woord had begrepen als iemand dit gebruikte, ZELFS ALS JE ZELF NIET het woord gebruikte of kon lezen of schrijven op dit punt.

**Concreetheid taak: (Brysbaert et al., 2014)**

Sommige woorden verwijzen naar dingen of handelingen in de realiteit die je direct kan ervaren door middel van je zintuigen. We noemen deze woorden ‘concrete woorden’. Andere woorden verwijzen naar betekenissen die niet direct ervaren kunnen worden maar die we kennen omdat ze in andere woorden uitgedrukt kunnen worden. Deze woorden noemen we ‘abstracte woorden’. Weer andere woorden vallen ergens tussen deze twee extremen omdat we ze tot op zekere hoogte kunnen ervaren, maar tevens van taal afhankelijk zijn om ze volledig te begrijpen. We willen je vragen om aan te geven hoe concreet de betekenis van ieder woord is door gebruik te maken van een 5-punt schaal van abstract tot concreet. Een concreet woord krijgt dus een hogere beoordeling en verwijst naar iets dat bestaat in de realiteit; je kan er een directe ervaring van hebben door middel van je zintuigen (ruiken, proeven, tast, horen, zien) en de handelingen die je verricht. De makkelijkste manier om een concreet woord uit te leggen is door er naar te wijzen of het te demonstreren (e.g., om ‘zoet’ uit te leggen kan je iemand suiker laten eten; om ‘springen’ uit te leggen kan je simpelweg op en neer springen of iemand een videoclip laten zien waarin iemand op en neer springt; om ‘bank’ uit te leggen kan je naar een bank wijzen of een afbeelding van een bank laten zien). Een abstract woord komt met een lage beoordeling op de 5-punt schaal en verwijst naar iets wat je niet direct door je zintuigen of handelingen kan ervaren. De betekenis van een abstract woord wordt door de taal uitgedrukt. De makkelijkste manier om een abstract woord uit te leggen is door andere woorden te gebruiken (e.g., er is geen makkelijke manier om ‘gerechtigheid’ te demonstreren; maar we kunnen de betekenis uitleggen door andere woorden te gebruiken die een deel van de betekenis vangen). Omdat we de metingen van alle woorden in het woordenboek aan het verzamelen zijn (60.000 woorden in totaal) zal je zien dat er verschillende typen woorden zijn, soms zelfs enkel letters. Probeer altijd te bedenken hoe concreet (op basis van ervaring) de betekenis van een woord is voor jou. De kans is groot dat je woorden zal tegenkomen die je niet goed genoeg kent om een nuttige beoordeling te geven. Dit willen wij tevens graag weten omdat we in ons onderzoek alleen woorden willen gebruiken die mensen kennen. Het is tevens mogelijk dat er één of twee nep woorden tussen staan die je niet kan kennen. Wanneer je een woord niet kent geef dit dan alsjeblieft aan door de letter N (of n) te gebruiken.

We willen je vragen om een 5-punt schaal te gebruiken van abstract tot concreet en de letter N wanneer je het woord niet goed genoeg kent om een antwoord te geven.

Abstract (op basis van taal) Concreet (op basis van ervaring)

1 2 3 4 5

N = Ik ken het woord niet goed genoeg om een beoordeling te geven.

**Emotionele Staat, Opgewondenheid, en Dominantie Taak: (Bradley & Lang, 1999)**

**Diagram

Description automatically generated with medium confidence Emotionele Staat**

**Diagram

Description automatically generated Opgewondenheid**

**Diagram

Description automatically generated with medium confidence Dominantie**

De studie die vandaag wordt uitgevoerd heeft tot doel om gevoelens te onderzoeken en hoe mensen reageren op verschillende type woorden.

We noemen deze set figuren SAM en je zal deze figuren gebruiken om te beoordelen hoe je je voelde tijdens het lezen van ieder woord. SAM toont drie verschillende gevoelens: Vrolijk vs. Ongelukkig, Opgewonden vs. Kalm, Overheerst vs. Overheersend. Je zal deze schalen gebruiken om 3 beoordelingen voor ieder woord te geven dat je leest. Merk op dat ieder van de drie gevoelens op een andere schaal gerangschikt zijn. Het linker paneel toont de vrolijk-ongelukkig schaal die loopt van een glimlach tot een frons. Aan het ene uiteinde van de schaal ben je vrolijk, voldaan, tevreden, vergenoegd, of hoopvol. Wanneer je je volledig vrolijk voelt bij het lezen van een woord dan geef je dit aan door een 1 te selecteren aan de linkerkant. De andere kant van de schaal is voor wanneer je je volledig ongelukkig, geërgerd, ontevreden, melancholisch, wanhopig, of verveeld voelt. Je kan aangeven dat je je volledig ongelukkig voelt bij het lezen van een woord door een 9 te kiezen aan de rechterkant. Je kan ook gevoelens tussen deze twee extremen kiezen door een waarde in het midden te selecteren. Als je bijvoorbeeld een compleet neutraal gevoel hebt bij het lezen van een woord, noch vrolijk of ongelukkig, dan kies je een 4 in het midden.

De opgewondenheid of kalm schaal is het 2de type gevoel dat hier afgebeeld is. Aan het ene uiteinde van deze schaal voel je je gestimuleerd, opgewonden, waanzinnig, gejaagd, klaar wakker, of opgewekt. Wanneer je je volledig opgewekt voelt bij het lezen van een woord, selecteer dan de 1 aan de linkerkant. De andere kant van de opgewondenheid-kalm schaal vertegenwoordigd het compleet tegenovergestelde gevoel. Aan deze kant van de schaal voel je je volledig ontspannen, kalm, sloom, dof, slaperig, of niet opgewekt. Dit gevoel geef je aan door een 9 te selecteren aan de rechterkant. Net als de vrolijk-ongelukkig schaal kan je ook tussenliggende gevoelens van opgewondenheid en kalmte aangeven door een waarde in het midden te kiezen. Als je je bijvoorbeeld noch opgewonden of kalm voelt, selecteer dan een 4 in het midden.

De laatste schaal gaat over of je je overheerst of overheersend voelt. Aan de linkerkant van de schaal voel je je volledig overheerst, beïnvloed, verzorgd, onder de indruk, onderdanig, of gestuurd. Als je je volledig overheerst voelt bij het lezen van een woord, kies dan een 1 aan de linkerkant. Aan de andere kant van de schaal kies je een 9 wanneer je je volledig overheersend, invloedrijk, belangrijk, dominant, autonoom, of beheersend voelt. Je geeft aan dat je je overheersend voelt door een 9 aan de rechterkant te selecteren. Merk op dat wanneer de afbeelding groot is je je overheersend voelt en klein is wanneer je je overheerst voelt. Wanneer je je noch overheerst of overheersend voelt, kies dan een 4.

Het is de bedoeling dat je op een snel tempo werkt en niet te veel over ieder woord nadenkt. Probeer een beoordeling te geven op basis van je eerste en directe indruk wanneer je het woord leest.

**Vertrouwdheidstaak: (Gilhooly & Logie, 1980)**

Dit experiment is bedoeld om te onderzoeken hoe vaak je wordt blootgesteld aan bepaalde woorden. Je krijgt een lijst met woorden die je zal beoordelen op basis van hoe vaak je aan dat woord bent blootgesteld door een cijfer op een schaal van 1 tot 7 te geven. Op deze schaal staat een 1 voor “NOOIT,” ofwel, je hebt het woord nooit gezien, gehoord, of gebruikt in je leven; een 2 staat voor “ZELDEN,” ofwel, je hebt het woord op z’n minst een keer eerder gezien, gehoord, of gebruikt, maar slechts zelden; enzovoort tot een 7 die staat voor “ZEER VAAK,” ofwel, je ziet, hoort, of gebruikt het woord bijna dagelijks.

Het is niet erg als je bepaalde woorden niet goed genoeg kent om een definitie te geven. Geef simpelweg een beoordeling op basis van hoe vaak je bent blootgesteld aan het woord, ongeacht de betekenis van het woord. Er zitten wellicht bepaalde woorden tussen die je vaker hebt gebruikt of gehoord dan dat je ze hebt gezien. Of er zijn wellicht andere woorden die je vaker hebt gezien dan dat je ze hebt gebruikt of gehoord. In zulke gevallen gebruik je de hoogste beoordeling. Als je bijvoorbeeld het woord “proost” vaak gebruikt of hoort maar zelden in een tekst hebt gezien, dan beoordeel je het woord “proost” met “VAAK” door een 6 neer te schrijven.

Ga naar de lijst met woorden en beoordeel deze op je eigen tempo. Snelheid staat **niet** centraal in dit experiment, elke participant zal voldoende tijd krijgen om het af te ronden. Daarentegen is het niet nodig om te veel tijd aan ieder woord te besteden. Het belangrijkste is dat je zo accuraat mogelijk bent in deze taak. Wees dus zo eerlijk mogelijk in je beoordelingen. Veel van de woorden in het experiment zijn zeer zeldzaam, dus het is niet vreemd als je niet bekend bent met alle woorden. Probeer de woorden te beoordelen naar je beste vermogen.

**Iconiciteit Taak:**

In deze taak willen we onderzoeken hoe iconisch je denkt dat bepaalde woorden zijn. Woorden worden beschouwd als iconisch wanneer het woord ‘klinkt zoals de betekenis’. Bijvoorbeeld, het woord ‘haha’ klinkt als het geluid dat je maakt wanneer je lacht. Je zou misschien zelfs de betekenis van het woord kunnen raden wanneer je de taal van het woord niet kende. Andere woorden zijn helemaal niet iconisch; bijvoorbeeld, er is niks ‘zelden’ of ‘frequent’ klinkend aan deze woorden. Als je de taal niet kende zou je de betekenis niet kunnen raden. Voor elk woord dat je ziet moet je op een schaal van 1 tot 7 aangeven hoe iconisch dat woord is, met een 1 voor helemaal niet iconisch en een 7 voor zeer iconisch. Spreek het woord hardop uit voordat je de beoordeling maakt. Bijvoorbeeld, het woord ‘haha’ is zeer iconisch en zou daarom een 7 ontvangen; het woord ‘boom’ is helemaal niet iconisch en ontvangt een 1. Woorden met een tussenliggende mate van iconiciteit krijgen een beoordeling tussen deze twee extremen, bijvoorbeeld een 3 of een 4.. Als je de betekenis van een woord niet kent, vul dan een X in. Werk op een redelijk snel tempo terwijl je zo accuraat mogelijk blijft in je beoordelingen. Voel je vrij om de volledige reikwijdte aan beoordelingen van 1 tot 7 te gebruiken en maak je geen zorgen over hoe vaak je een bepaalde beoordeling gebruikt zolang je maar zo eerlijk mogelijk bent.

**Study Stimuli:**

<https://osf.io/m56z7/>

Note that the stimuli chart includes many blanks. These will be filled in with translators for the language listed in the column name (language code is the first two letters of each column). We want to ensure the stimuli selection procedure is approved before moving to this step because the process is slow to run if we need to do it again.

**Toestemmingsformulier:**

**Toestemmingsformulier**

Het begrijpen van de verwerking en betekenis van woorden

Je bent uitgenodigd om deel te nemen aan een studie over hoe je leest en woorden en hun betekenis verwerkt. We willen je vragen dit formulier te lezen en eventuele vragen te stellen die je hebt voordat je je toestemming geeft aan deze studie deel te nemen.

Deze studie wordt uitgevoerd door Dr. Erin M. Buchanan, Professor in Cognitive Analytics aan Harrisburg University of Science and Technology.

**Achtergrondinformatie:**

In deze studie willen we je vragen om verschillende vragen te beantwoorden over woord concepten. Je zal bijvoorbeeld gevraagd worden om de eigenschappen van een woord te bepalen, te beoordelen hoe bekend je bent met een woord, of simpelweg beoordelen of een reeks letters een echt woord is of niet.

**Procedure:**

Deze studie zal volledig online plaatsvinden vanaf een desktop of laptop computer met een toetsenbord. Je zal instructies ontvangen voor het onderdeel van het experiment dat per persoon willekeurig geselecteerd zal worden. Zodra je klaar bent met het experiment kan je meer leren over deze studie en het doel van dit onderzoek. Het gehele experiment zou minder dan 30 minuten in beslag moeten nemen om af te ronden.

**Risico’s en Voordelen van het deelnemen aan deze Studie:**

Er zal geen informatie over je verzameld worden die jou persoonlijk kan identificeren en je antwoorden in dit experiment zijn daarom anoniem. De huidige studie is vergelijkbaar met een online spel en kan enige vermoeidheid of verveling veroorzaken op basis van de taak die je krijgt.

Er is geen direct voordeel voor het deelnemen aan deze studie. Echter, je antwoorden zullen helpen bij het beter begrijpen van ons begrip van de taal en de betrokken cognitieve geheugenprocessen.

**Compensatie:**

Het is mogelijk dat je compensatie ontvangt voor deelname aan deze studie van de onderzoeker die dit experiment bij jou afneemt.

**Vertrouwelijkheid en het Delen van Data:**

Er worden maatregelen genomen om er zeker van te zijn dat alle informatie die je verstrekt anoniem zal blijven. De data die uit dit project voortkomt zal openbaar gepubliceerd worden voor andere onderzoekers om te gebruiken; echter, de data zal niet persoonlijk aan jou gekoppeld kunnen worden. Je naam of andere identificerende informatie zal niet in de dataset geplaatst worden en er zullen geen gesproken of geschreven verwijzingen gemaakt worden die jou kunnen koppelen aan deze studie. In alle publicaties zal de informatie gepresenteerd worden op een manier die jou niet kan identificeren.

Voordat de data gedeeld wordt buiten het onderzoeksteam zal alle mogelijke identificerende informatie verwijderd worden. De anonieme data kan gebruikt worden door het onderzoeksteam of gedeeld worden met andere onderzoekers voor zowel gerelateerde als ongerelateerde onderzoeksdoeleinden in de toekomst. Je anonieme data kan ook beschikbaar worden gesteld in online opslagplaatsen zoals de Open Science Framework (dit is een gratis data opslagplaats die registratie vereist voor toegang). Dit stelt andere onderzoekers of geïnteresseerde partijen in staat om toegang te krijgen tot de data voor verdere analyse.

**Merk op dat we je data na afloop van de studie niet meer kunnen verwijderen omdat de gegevens geanonimiseerd zijn.**

**Vrijwillige aard van de Studie:**

**Deelname aan deze studie is vrijwillig:**

Je besluit om wel of niet deel te nemen aan deze studie zal geen invloed hebben op je huidige of toekomstige relaties met de Harrisburg University of Science and Technology of je lokale instituut of universiteit. Als je besluit deel te nemen ben je alsnog vrij om antwoord op bepaalde vragen te weigeren of je terug te trekken uit de studie zonder dat dit invloed heeft op je relaties met de genoemde instituten.

**Contactgegevens en Vragen:**

De onderzoekers die deze studie uitvoeren zijn Dr. Erin M. Buchanan samen met de Psychological Science Accelerator. Je kan op dit punt iedere vraag stellen die je hebt. Als je later een vraag hebt **moedigen we je aan** om contact op te nemen met Dr. Erin M. Buchanan via ebuchanan@harrisburgu.edu.

**Vragen of Bezorgdheden:**

Deze studie is beoordeeld en goedgekeurd door Harrisburg University of Science and Technology’s Institutional Review Board (IRB). De IRB heeft bepaald dat deze studie voldoet aan de eisen voor de bescherming van proefpersonen zoals vastgelegd door staats- en wetgeving en het universitaire beleid.

***Een kopie van deze informatie is op verzoek beschikbaar.***

**Wervingsmaterialen:**

<https://osf.io/gp8nv/>