Online data collection to address language sampling bias:

Lessons from the COVID-19 pandemic

Rowena Garcia[1], Jens Roeser[2], & Evan Kidd[1, 3, 4]

[1]Max Planck Institute for Psycholinguistics

[2]Nottingham Trent University

[3]Research School of Psychology, The Australian National University

[4]ARC Centre of Excellence for the Dynamics of Language

Address for correspondence:

Rowena Garcia
Language Development Department
Max Planck Institute for Psycholinguistics
P.O. Box 310 6500XD
Nijmegen, Netherlands
Email: rowena.garcia@mpi.nl
https://orcid.org/0000-0003-1363-542X

Online data collection to address language sampling bias:

Lessons from the COVID-19 pandemic

The COVID-19 pandemic has massively limited how linguists can collect data, and out of necessity, researchers across several disciplines have moved data collection online. Here we argue that this rising popularity of remote web-based experiments also provides an opportunity for widening the context of linguistic research by facilitating data collection from understudied populations. We discuss collecting production data from adult native speakers of Tagalog using an unsupervised web-based experiment. Compared to equivalent lab experiments, data collection went quicker, and the sample was more diverse, without compromising data quality. However, there were also technical and human issues that come with this method. We discuss these challenges and provide suggestions on how to overcome them.

Keywords: web-based experiment, language production, Gorilla, structural priming, Tagalog

1 Introduction

The collection of linguistic data requires access to speakers of the language of interest. For researchers who study the languages of their local communities, COVID-19 has no doubt made research difficult. However, for those whose research takes them into the field beyond their

home city or country, the pandemic has tested the limits of resourcefulness. In this paper, we describe how we overcame the challenge of being unable to go to the field during the pandemic, and how we remotely obtained experimental production data in an understudied language from the other side of the world. We show that, given the right local conditions (e.g., internet connection, access to technology), collection of high-quality data is possible, and make practical recommendations for researchers considering web-based data collection.

1.1 Language coverage across subdisciplines

The primary goal of linguistics, broadly construed, is to build explanatory theories of the capacity for language in all its instantiations. Thus, given that there are around 7,000 languages currently spoken across the world (Eberhard et al. 2020), two important sets of questions are: (i) how and why are languages so diverse?, and (ii) how does the human brain, which contains the same language-supporting neural structures across linguistic groups, acquire and process such a diverse set of systems?

Research in fields such as language documentation and linguistic typology address question (i), with the last two decades heralding important steps in increasing data coverage against the backdrop of rapid language endangerment and death (see Seifart et al. 2018). Question (ii) is the primary concern of psycholinguistics, whose relationship to linguistic diversity is comparatively poor. For instance, Anand et al. (2011) found that 85% of experimental studies were based on only 10 languages (with English comprising 30%). Similarly, Jaeger and Norcliffe (2009) estimated that studies of language production have been conducted on only 0.6% of the world's languages.

One reason for this focus on languages from WEIRD (Weird, Education, Industrialized, Rich, Democratic; Henrich et al. 2010) societies has been the technological challenge of creating laboratory conditions outside of university campuses. However, in the past decade many of these technological challenges have been overcome, and it is now possible to collect reliable experimental data online (see Stewart et al. 2017). COVID-19 ratcheted up that effort, with labs rapidly moving their data collection online (Sauter et al. 2020). Here we focus on collecting language production data. Although collecting spoken data through the internet is not new (Vogt et al. 2021; Ziegler et al. 2018), past studies typically rely on testing fixed participant pools and not on collecting community samples that are likely the target of research on understudied languages.

1.2 The field context

The research we describe here was on Tagalog, a Western Austronesian language spoken in the Philippines. Notably, we are interested in a typologically unique feature of the language: symmetrical voice (Foley 2008; Riesberg et al. 2019), which means that the language has more than one basic transitive structure, which are equally marked with voice and noun morphology, and no argument is demoted to a lower syntactic position. The voice inflection on the verb marks the thematic role of the argument that is marked by *ang* (see [1] and [2]; Himmelmann 2005).

(1) K<**um**>ain   **ang** bata   ng   mangga
      <AV>PFV.[1]eat   child   mango
      'The child ate a mango / mangoes.'

---

[1]   *AV* refers to agent voice, *PV* to patient voice, *PFV* to perfective aspect

(2) K<**in**>ain   ng   bata   **ang** mangga
     <PV>PFV.eat   child   mango
     'The/A child ate the mango.'

   There is a considerable debate on how to theoretically explain this voice system. For

example, Carrier-Duncan (1985) claims that in both (1) and (2), the *child* is the subject, which

also means that the *ng*-phrase in these two sentences have different grammatical functions,

similar to De Guzman's (2000) claim. Other linguists claim that the *ang*-phrase is the subject

(Himmelmann 2005; Kroeger 1993); thus the subject in (1) is the agent, but the patient in (2),

while the propositional meaning remains the same (*child* is the agent and *mango* is the patient of

the action *eat*). As experimental linguists, we investigated the representational relationship

across the voices using a *structural priming* task.

   Structural priming—the tendency to repeat a previously encountered structure (Pickering

& Ferreira 2008)—has been argued to be a suitable method to test for shared linguistic

representations (Branigan & Pickering 2017). Accordingly, if processing of utterance A (prime)

affects the subsequent processing of utterance B (target), then A and B are assumed to share

some representational features. Structural priming experiments have shown, for example, that

speakers produce more passive sentences after repeating or hearing a passive structure compared

to an active structure (Bock 1986), even without lexical overlap between prime and target

(Branigan & Messenger 2016). In comparison to acceptability judgment and contingency tests,

Branigan and Pickering (2017) consider structural priming more advantageous because it is

implicit and does not require metalinguistic judgments. While priming studies have been

conducted across various participant types (e.g., adults, children, brain-damaged individuals), we

concentrate here on its use with adults.

We conducted two structural priming experiments to investigate how voice-marking, syntactic role order, and thematic role order are mentally represented. Prior to the onset of the pandemic, we had conducted similar research that focused on children acquiring Tagalog (Garcia & Kidd 2020). Some unclear results from an adult comparison group led us to design a follow-up study, with the pandemic forcing us to conduct this study online. Thankfully, several features of the Philippines and its Tagalog-speaking population are conducive for web-based testing: 1) the population is highly literate (UNESCO Institute of Statistics 2015), and 2) the majority of the Metro Manila population, where Tagalog is acquired as a native language, has internet access (Department of Information and Communications Technology 2019). For statistics for the Philippines and other countries, please refer to a Shiny app we built as a resource for researchers wishing to conduct similar research on non-WEIRD populations: https://bit.ly/3cePdx9.

2 Running an unsupervised web-based experiment

2.1 Preparing the experiment

Instead of conducting an experiment via video calling platforms, which would require careful scheduling of individual participant sessions, we programmed an unsupervised experiment which could be done simultaneously by many participants. We used Gorilla (Anwyl-Irvine et al. 2020a), a relatively new web-based environment for both building and deploying experiments. Gorilla is fully compliant with the European Union's General Data Protection Regulation (GDPR) rules, and allows the construction of different experiment designs (ex. randomised, between-subjects) without the need for coding. Similar to other web-based experiment builders (e.g., PsychoJS, jsPsych, Lab.js), Gorilla supports different web-enabled devices (e.g., computers, tablets, smartphones), operating systems (e.g., Windows, Mac OS, iOS, Android),

and browser types (e.g., Chrome, Safari, Firefox). Additionally, Gorilla, like other browser-based systems, has been shown to provide reasonable precision for onset and duration of stimulus presentation as well as reaction time recordings, especially on visual stimuli (Anwyl-Irvine et al. 2020b; Bridges et al. 2020).

The use of experiment builders is usually free, but companies usually charge for access to the collected data (for a comparison of pricing and features of different tools for online studies, see Sauter et al. 2020). Gorilla, for example, offers a lab subscription (1,440€ annually)[2] which is inclusive of access to data from 2000 experimental sessions. Other subscriptions, as well as a pay-per-participant (1€) option, are also available.

2.2 Recruitment

We set the target number of participants on Gorilla, and recruited participants through social media posts which contained the link to the experiment. Interested individuals with internet connection only had to open the link in a browser, and could immediately start the experiment. In order to prevent participants from re-doing the experiments, we explicitly mentioned in the advertisement and in the consent form that participation is limited to one session, and compensation will be awarded only once per person.

2.3 Experiment flow

Since there was no experimenter present, we had to take measures to ensure that participants follow the instructions correctly and finish the experiment on their own. Participants were asked

[2]   Pricing information was obtained on September 22, 2021 from https://app.gorilla.sc/pricing

to find a quiet space to complete the experiment. Figure 1 shows the participants' progression through the experiment.
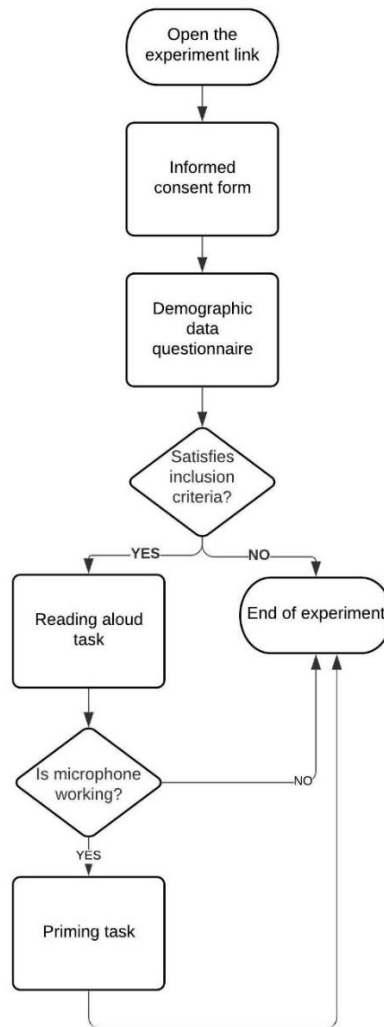

Figure 1. Flowchart showing the participants' progression through the experiment.

Similar to lab-based experiments, participants first had to give consent and fill out a demographic data questionnaire. However, exclusion of participants was done automatically by Gorilla based on the participants' answers to the questionnaire. For example, if a participant reported that he/she had been previously diagnosed with any speech/language impairment, the succeeding tasks would no longer be displayed, and the experiment would end.

An important inclusion criterion was that participants were native and dominant speakers of Tagalog. To assess this, we asked participants to read aloud a short passage (the first half of the revised *Halo-Halo Espesyal,* Ligot et al. 2004). This was recorded and later used to screen participants (a native speaker listened to the recordings and judged the speaker as native versus non-native). The task also served an additional technical function as it included a microphone check. Participants who did not have a working microphone were automatically excluded.

Gorilla's programming interface allows randomized allocation of participants to experimental lists. Once assigned to a list, participants saw the instructions for the priming task. Unlike in lab-based experiments, misunderstandings cannot be resolved by the experimenter. To overcome this potential problem, we presented a short video clip detailing the procedure, without displaying our target structures so as to not bias responses. The participants then completed four practice trials, followed by the priming task, which followed standard procedure (Pickering & Branigan 1999; Vernice et al. 2012; see Figure 2 for the trial progression). Participants were required to judge whether or not a presented sentence (prime) and picture matched. This judgment served as a cover task for participants to read the prime sentence. Then, participants were provided with a target prompt and a new picture, and were asked to describe this target picture using the prompt. Audio-recording started automatically once the target picture was shown, so participants did not have to manually click "Start". This was done in order to prevent the first part of the recording from being cut off, based on Gorilla's report of a delay in starting the recording. Participants were asked to click a "Stop" button once they had finished recording, which also triggered the presentation of the next trial.
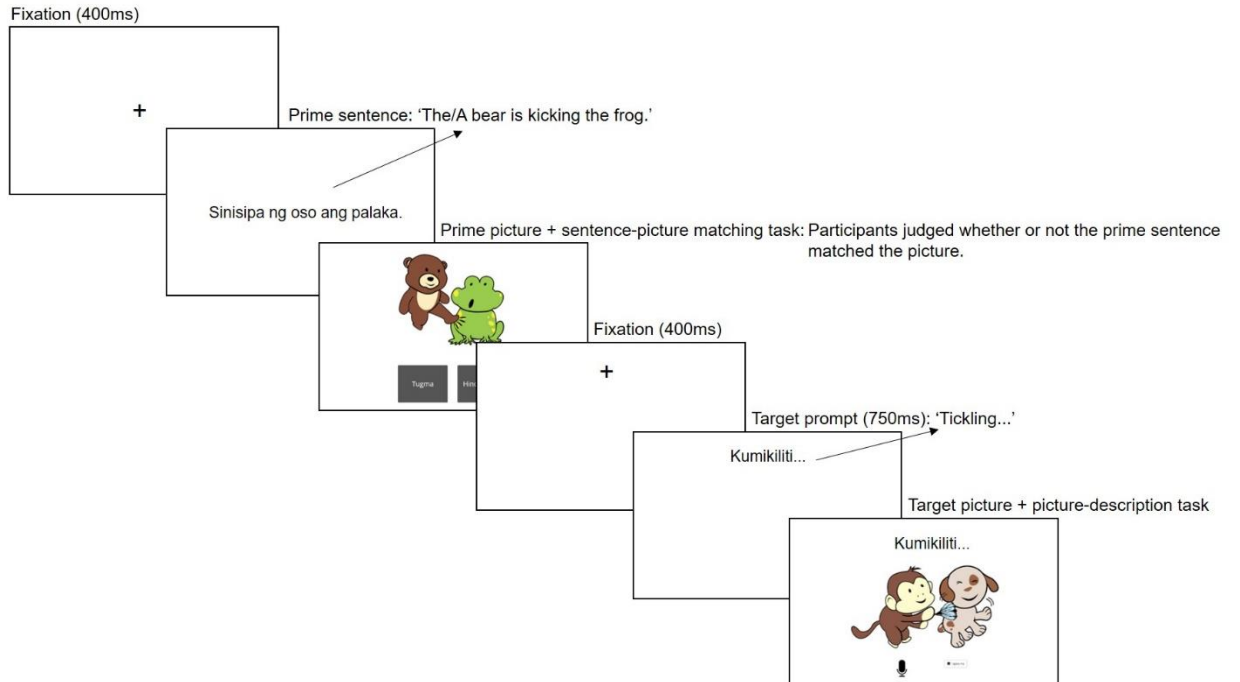
Figure 2. Time course of an experimental trial in the priming task.

After the last trial, participants were encouraged to type what they thought was the aim of the experiment, and to indicate if they encountered issues such as slow loading of pictures. The whole experiment session took 29 minutes on average (SD = 7.59, range: 13 to 49 minutes). As compensation, participants were manually sent a convenience store voucher via email.

The web-based experiment including the stimuli sentences and pictures can be found and reused at https://app.gorilla.sc/openmaterials/283351.

3 Advantages

The main advantages of collecting our data online were (i) the speed of collection, and (ii) the access to a more diverse participant pool. With respect to (i), we completed collection for two experiments, each with a sample size of 64 participants, in 10 days. This underscores the ability

of online testing to both ramp up the scale of participant numbers, leveraging off the fact that multiple participants can complete the task at any time at their own convenience. As regards (ii), the most important advantage was that we were able to collect data from a non-WEIRD population. Additionally, instead of testing predominantly 18- to 21-year-old undergraduate students, as is typical of traditional laboratory-based studies in the Cognitive Sciences (Henrich et al. 2010), our sample was more diverse. Participants had a mean age of 27 years (SD = 4.71, range: 18–36 years), most of whom were college graduates (68%) followed by college students (26%), and the rest were high school graduates (4%) and elementary graduates (2%). Sixty-six percent of participants were females.

More importantly, Gorilla delivers interpretable data. After each experiment session, spreadsheets for each experiment task (in our case: consent form, demographic data, reading aloud task, and priming task) and the audio-recordings become available for downloading. Although the audio-recording quality is dependent on the participant's device and the amount of background noise, the files were clear enough for transcriptions: based on a speech intelligibility analysis of 20% of the data from one experiment, 99% of produced words could be transcribed.

While our focus was on participants' productions, Gorilla records, by default, accuracy and reaction time data (Anwyl-Irvine et al. 2020b; Bridges et al. 2020), as might be used in studies of sentence comprehension or lexical access. Accordingly, we obtained accuracy data for the picture verification component of our task. Gorilla also reports information that can be taken into account statistically, such as notifications of loading delays longer than 10s, and the type of device and screen size that the participants used. Additionally, Gorilla provides excellent and prompt support, enabling us to quickly troubleshoot any problems.

4 Challenges

The use of web-based data collection methods also presents a range of challenges that lead to variability in data quality across participants (Vogt et al. 2021; Woods et al. 2015). Most crucially, our sample included a large amount of unusable data due to technological issues and human factors. Across our two experiments, we needed to conduct 195 experimental sessions to obtain the target sample of 128 participants; thus we needed to test an additional 52.3% (see Table 1).[3] In comparison, in an experimenter-led study with largely similar methods, only one adult participant was excluded (Garcia & Kidd 2020).

Table 1. Reasons for participant exclusion in the web-based experiments.

Reason for exclusion	Number of participants (% in brackets)
Contributed less than 20 experimental items	26 (13.33%)
Had previously done the experiment	21 (10.77%)
Did not satisfy the inclusion criteria (i.e., age, fluency, hometown)	8 (4.10%)
Did not follow instructions / use the target prompt	7 (3.60%)
Accuracy below 60% in the sentence-picture matching task	5 (2.56%)

[3]   This number does not include participants who only started the experiment but who stopped the process before reaching the priming task, which was considerably higher. We expected that they were the same participants who eventually reached the priming task in the other experiment sessions, as Gorilla recognizes each click of the experiment link as a different participant. It must be noted that data from unfinished experiments (e.g., name of participant) could only be seen once they have been paid for.

4.1 Technological issues

In Gorilla, stimuli are preloaded to facilitate smooth trial presentation and accurate timing. However, the time to load the files varies with file size and importantly, internet speed. Given the slow internet connection in the Philippines, half of the pilot participants reported loading delays of visual stimuli. This problem was reduced by rescaling the quality of the pictures (without visibly changing the appearance) after the pilot, leaving us with loading delays reported for only 0.1% of trials. However, out of the 195 experimental sessions, 13 participants reported that one or a few pictures did not load, while 12 others reported that picture presentation was slow. The participants' varying internet connection speed was also most likely a contributing factor for the large variability in completion durations (SD = 7.59 minutes).

Gorilla also does not recognize the microphone when the experiment link is opened from a social media platform on a smartphone. As our experiment was programmed to proceed to an early exit if no microphone was detected, pilot participants who clicked the link got automatically excluded. We addressed this issue by removing a clickable experiment link from our advertisement, so participants would have to type or copy the shortened link on their browser. Unfortunately, well-meaning participants started inviting others through a clickable link.

Consistent with Gorilla's report of a delay in starting audio-recording from the participants' microphones, 14% of audio files were cut off in the beginning (based on 20% of data from one experiment). Truncations of the end of utterances was observed in 4% of the files, possibly due to participants clicking the "Stop" recording button too soon. Fortunately, in our experiments, the first word was the given prompt, so it was not the most crucial constituent and did not result in data loss. Additionally, because the first argument was sufficient to code the

dependent variable (i.e., thematic role order), early truncations of the recordings were not a problem either.

As participants' data and audio-recordings have to be uploaded to a server, internet bandwidth affects the collection of the data. For 24 participants, either there was data for only a few items (unfinished), or the data showed that the participant finished the experiment but there were no audio recordings found. One participant reported that upon finishing the experiment, she received a notification that her recorded files were still being uploaded even after two hours. Gorilla marked this as an unfinished session, and only half of her trials were recorded. The same may have happened in other unfinished sessions.

Finally, Gorilla's counterbalance node is not sensitive to eventual attrition of participants. It does not automatically rebalance assignment to lists after drop-outs, which can lead to an uneven number of participants per experiment list. Therefore, the set of available experiment lists (in the counterbalance node) have to be manually changed. In order to minimize manual work, we edited the lists near the end of the data collection when it was clear which lists missed participants.

4.2 Human factors

Given our recruitment method of simple link sharing, and since Gorilla does not store the participants' IP addresses following GDPR, it was easier for individuals to participate repeatedly (e.g. to receive extra vouchers). We recognized 21 data sets as coming from participants who had already previously completed the experiment. Suspicious data were identified through email addresses which (1) were previously used by another participant, or (2) contained a name that was different from the reported name. For other data sets, the voice in the recording did not

match the reported gender, and upon review of other recordings, this voice matched that of a previous participant (sometimes with the same family name).

The lack of experiment supervision also resulted in exclusions. The programmed automatic exclusion based on demographic data could not exclude those who grew up outside Greater Manila Area, or those who were not fluent in Tagalog, and did not exclude those who were way above the age range of the other participants (until 36 years of age). Additionally, there were a few participants who did not follow the instruction, and changed the voice-marking on the target prompt.

The other human issues we encountered were not specific to web-based methods but were probably worsened by the lack of an experimenter. There were a few who performed poorly in the simple picture verification task, which also served as an indicator of their attention during the task, and language proficiency. Others skipped items to finish the experiment quickly.

5 Recommendations for future use

Researchers who plan to conduct web-based experiments should consider several things that are not particularly relevant for in-person experiments.

4.1 Participant recruitment

For researchers targeting populations in Europe, Australia, or North America, it might be worth paying for a recruitment platform which ensures that participants can only do the experiment once (e.g., Prolific, Amazon Mechanical Turk). This might be a better choice because screening for participants who had completed the experiment multiple times is time-consuming. Additionally, such platforms directly send compensation to the participants. However,

participants registered on these platforms come primarily from the Global North. This limits their utility for research targeting understudied languages spoken in non-WEIRD societies. Currently, Prolific accepts participants who reside in member-countries of the Organisation for Economic Co-operation and Development (OECD) except for Turkey, Lithuania, Colombia, and Costa Rica; but also accepts participants from South Africa (Moodie 2021). For Amazon Mechanical Turk, the majority of the workers reside in the United States (75%), followed by India (16%), and the remaining 9% from the rest of the world (Difallah et al. 2018). For our study, there was a limited number of registered participants in recruitment platforms who would fit our inclusion criteria. This is the case for most non-WEIRD societies, and in these cases, Gorilla or other platforms that allow flexible recruitment are the best option.

To prevent participants from completing the experiment more than once, Gorilla offers a stricter recruitment option where participants will be sent an email with a personalized link to access the experiment. This link can only be used once, but it does not prevent an email recipient from forwarding this link to someone who has already taken part in the experiment. Nevertheless, it makes it more difficult for a person to re-do the experiment. For our study, we simply shared a link because the additional step of first asking potential participants to send a message to express their interest in participating (so they could be sent a personalized link) might prevent many people from joining. The stricter Gorilla recruitment option is a good choice but is also time-consuming.

5.2 User requirements

In Gorilla, researchers have the option to restrict which types of devices can be used to access the experiment. This might be worth considering if the experiment involves complex visual stimuli,

which would not be properly shown on a small screen. Because many Filipinos mostly rely on smartphones to connect to the internet (as is likely the case in Global South countries), we decided not to require a particular device for doing the experiment. Moreover, our visual stimuli were not complex: short sentences/words and simple action pictures. Additionally, pilot tests on different devices showed that the layout we used was suitable for most devices.

Another option is to require a minimum internet connection speed, in order to prevent download and upload issues. Due to the low average bandwidth in the Philippines, we did not have such a requirement. Instead, we reduced the size of our stimuli to speed up loading. Fortunately, it was relatively simple to locate trials in which the stimulus did not load properly, as participants would not be able to describe a picture they had not seen.

5.3 Supervision of data collection

Even if web-based data can be collected without supervision, we still advise researchers to regularly check the ongoing recruitment. Given that each click of the link starts a new session in Gorilla, the recruitment slots can quickly fill without actual data being collected (because participants click on a link but do not complete the experiment). We thus recommend that researchers manually reject inactive sessions (e.g., an experimental slot that has remained open but unfinished for over an hour). However, the time limit for experiment completion should not be too short, as it is indeed possible that participants with lower bandwidth require more time to finish. Any time limit will depend on the goals of the task and the nature of the effect under investigation.

Finally, we advise researchers to check the quality of the data as they are collected, and to have clear a priori ideas about data inclusion. As the experiment runs in the absence of an

experimenter, researchers do not have a rough idea which participant would have to be excluded until they look at the data. More importantly, since Gorilla's counterbalance node for assigning participants to specific stimuli lists does not account for eventual drop-outs, one should be ready to edit the counterbalance node to add slots in lists that still have incomplete data.

5.4 Experimental tasks

In consideration of the higher exclusion rates due to non-laboratory conditions, researchers should opt for simpler and shorter experiments. In order to reduce distractions, it would help to ask participants to do the task in a quiet environment and to use head phones, if possible. Some tasks may require simple attention checks (e.g., simple addition or subtraction). We also highly recommend doing a pilot study to test if the experimental procedure works and to solve possible technological issues, before the experiment becomes widely available to participants.

6 Conclusion

While the Covid-19 pandemic has limited our ability to access the population of speakers that is required for our research, the development of web-based experimental platforms coupled with greater accessibility of technological advances to the world's population, mean that it is possible to conduct research remotely. Accordingly, the pandemic has in fact presented the field with an opportunity to start widening the context of linguistic research by making data collection from non-WEIRD populations easier.

We end by saying that, while we have described the use of web-based platforms with reference to our production priming experiment, we see much value in web-based data collection for many types of linguistic data that are sorely needed for understudied languages. These

platforms more easily allow the collection of large comparable cross-linguistic data, which opens up important new ways of addressing the sampling bias in the discipline.

References

Anand, Pranav, Sandra Chung & Matthew W. Wagers. 2011. Widening the net: Challenges for gathering linguistic data in the digital age. *NSF SBE 2020. Rebuilding the mosaic: Future research in the social, behavioral and economic sciences at the National Science Foundation in the next decade.*

Anwyl-Irvine, Alexander, Edwin S. Dalmaijer, Nick Hodges & Jo K. Evershed. 2020a. Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52(1), 388-407.

Anwyl-Irvine, Alexander, Edwin S. Dalmaijer, Nick Hodges & Jo K. Evershed. 2020b. Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods,* 1-19.

Branigan, Holly P. & Katherine Messenger. 2016. Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition* 157, 250-256.

Branigan, Holly P. & Martin J. Pickering. 2017. An experimental approach to linguistic representation. *Behavioral and Brain Sciences* 40, 1–61.

Bridges, David, Alain Pitiot, Michael R. MacAskill & Jonathan W. Peirce. 2020. The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ* 8:e9414.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3), 355–87.

Carrier-Duncan, Jill. 1985. Linking of thematic roles in derivational word formation. *Linguistic Inquiry* 16(1), 1–34.

Difallah, Djellel, Elena Filatova & Panos Ipeirotis. 2018. Demographics and Dynamics of Mechanical Turk Workers. *WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 135–143).

De Guzman, Videa P. 2000. Some remarks on the grammatical functions of the nonabsolutive agent in Tagalog. *Oceanic Linguistics Special Publications* 29, 224-240.

Department of Information and Communications Technology. 2019. *National ICT Household Survey 2019*. https://dict.gov.ph/ictstatistics/nicths2019/ (10 February, 2021.)

Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2020. *Ethnologue: Languages of the World*. Twenty-third edition. Dallas, Texas: SIL International.

Foley, William A. 2008. The place of Philippine languages in a typology of voice systems. In Peter K. Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 22–44. Stanford, CA: CSLI Publications.

Garcia, Rowena & Evan Kidd. 2020. The acquisition of the Tagalog symmetrical voice system: evidence from structural priming. *Language Learning and Development* 16(4), 399-425.

Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world?. *Behavioral and Brain Sciences* 33(2-3), 61-83.

Himmelmann, Nikolaus P. 2005. Tagalog. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 350–376. Oxford: Routledge.

Jaeger, T. Florian & Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence
production. *Language and Linguistics Compass* 3(4), 866-887.

Kroeger, Paul R. 1993. Another look at subjecthood in Tagalog. *Philippine Journal of
Linguistics* 24(2), 1-16.

Ligot, Fernando Alejandro C., Glenda B. Gacer, Maria Tedie Rose D. Mateo & Juan Paolo D.
Santuele. 2004. *Revision and pilot testing of the "Halo-Halo Espesyal" reading passage
for Filipino cleft lip and/or palate speakers*. Manila: University of the Philippines Manila
Undergraduate thesis.

Moodie, Jim. 2021. *Who are the people in your participant pool?*. https://researcher-
help.prolific.co/hc/en-gb/articles/360009220833-Who-are-the-people-in-your-participant-
pool- (7 September, 2021.)

Pickering, Martin J. & Holly P. Branigan. 1999. Syntactic priming in language production.
Trends in Cognitive Sciences 3(4), 136-141.

Pickering, Martin J. & Ferreira, Victor S. 2008. Structural priming: a critical review.
Psychological bulletin 134(3), 427.

Riesberg, Sonja, Kurt Malcher & Nikolaus P. Himmelmann, N. P. 2019. How universal is agent-
first? Evidence from symmetrical voice languages. *Language* 95(3), 523–561.

Sauter, Marian, Dejan Draschkow & Wolfgang Mack. 2020. Building, hosting and recruiting: A
brief introduction to running behavioral experiments online. *Brain Sciences* 10(4), 1–11.

Seifart, Frank, Nicholas Evans, Harald Hammarström & Stephen C. Levinson. 2018. Language
documentation twenty-five years on. *Language* 94(4), e324-e345.

Stewart, Neil, Jesse Chandler & Gabriele Paolacci. 2017. Crowdsourcing samples in cognitive
science. *Trends in Cognitive Sciences* 21(10), 736-748.

UNESCO Institute of Statistics. 2015. *Literacy rate in the Philippines*.

    http://uis.unesco.org/en/country/ph (10 February, 2021.)

Vernice, Mirta, Martin J. Pickering & Robert J. Hartsuiker. 2012. Thematic emphasis in

    language production. *Language and Cognitive Processes* 27(5), 631-664.

Vogt, Anne, Roger Christoph Hauber, Anna K. Kuhlen & Rasha Abdel Rahman. 2021. Internet

    based language production research with overt articulation: Proof of concept, challenges,

    and practical advice. https://doi.org/10.31234/osf.io/cyvwf

Woods, Andy T., Carlos Velasco, Carmel A. Levitan, Xiaoang Wan & Charles Spence. 2015.

    Conducting perception research over the internet: A tutorial review. *PeerJ* 3, e1058.

Ziegler, Jayden, Jesse Snedeker & Eva Wittenberg. 2018. Event structures drive semantic

    structural priming, not thematic roles: Evidence from idioms and light verbs. *Cognitive*

    Science 42(8), 2918-2949.