VIDUL: A Tool for Elevating Inclusivity in Psycholinguistics

Introduction. Psycholinguistics has an inclusivity problem. Foundational theorems are often derived from studies that focus exclusively on the English language, and research in the field generally tends to skew towards western societies. Recently published literature highlights that "the particular characteristics of English and the linguistic habits of English speakers bias the field by both warping research programs and overgeneralizing observations from English speakers' behaviors, brains, and cognition to our entire species" [1]. Presently, when one wants to identify work that is being done on understudied languages, there is a tendency to rely on availability - that is, defaulting to a common group of researchers who often come to mind for the given domain. With this in mind, a question was posed on Fernanda Ferreira's social media account calling for all linguists and psycholinguists who were producing research on historically understudied languages. The post accrued a significant amount of engagement across multiple metrics, indicating a keen interest among the academic community in sharing their work on underrepresented topics in psycholinguistics. For this reason, we decided to build a Visual Interactive Database of Understudied Languages (VIDUL) to facilitate collaboration and contribute towards greater inclusivity in psycholinguistics.

Method. This study uses social media as a modern method of collecting global, self-reported data. Over a period of six months, we collected relevant information on researchers and their work through recurring Google Form posts, ultimately building towards a spreadsheet database.

A key pain point was uncovered throughout the process of arranging the collected data into a spreadsheet. Due to the use of self-report, social media surveys for data collection, we struggled to maintain the consistency of responses. As such, additional research into department websites and published literature was required to fill out incomplete entries and edit inadequate responses in the database. As for cleaning the data, we utilized the fuzzywuzzy library in Python to correct any misspelled entries that were inconsistent with language family data cross-referenced from the WALS database.

The final phase involved the development of a data visualization model using Python. This process made use of several libraries in Python to effectively work with the pre-existing database of researchers studying understudied languages. As for the data visualization itself, we used the Pyvis library to create an interactive network visualization of the database. We opted for an interactive network as opposed to a static spreadsheet in an effort to increase usability and leave the door open for future iterations. Each vertex in the network represents a researcher and contains the languages they study and their affiliated institutions. The vertices are grouped together by language family in an effort to create a correlation between the length of the edges and the similarity of the studied language. Ultimately, we aim to deploy this network visualization on a website to increase both visibility and accessibility for the global psycholinguistics community.

Discussion. The key motivator that drove this project was providing an interactive resource to facilitate collaboration on research concerning historically understudied languages. To date, we have collected over 250 entries and are actively working on translating this spreadsheet into the outlined data visualization tool. We aim for this network visualization to serve as a catalyst in creating greater visibility for researchers and their corresponding body of work. As we strive towards a more inclusive view of psychology and cognitive science, we are curious to see how our new age of digital interconnectedness can drive studies to highlight underrepresented topics in science.

	Researcher Last Name	Researcher First Name	Institution	Families	Language
1	**Researcher Last Name**	**Researcher First Name**	**Institution**	**Families**	**Language**
2	Sarvasy	Hannah	Western Sydney University	Trans-New Guinea	Nungon
3	Fascinetto-Zago	Karina	Benemérita Universidad Autónoma de Puebla / Leiden University	Indo-European	Venetian
4	Park	Kihyo	Cornell University		Korean
5	Zhu	Rebecca	UC Berkeley	Niger-Congo	Swahili
6	Zhu	Rebecca	UC Berkeley	Eastern Sudanic	Luo
7	Chanchaochai	Nattanun	Department of Linguistics, Chulalongkorn University	Tai-Kadai	Thai
8	Culbertson	Jennifer	University of Edinburgh	Bantu	Kîîtharaka
9	Kirby	Simon	University of Edinburgh	Sign Language	Nicaraguan Sign Language
10	Ernstreits	Valts	UL Livonian Institute	Uralic	Livonian
11	Parafita Couto	M Carmen	Leiden University / Universidade de Vigo	Indo-European	Galician
12	Parafita Couto	M Carmen	Leiden University / Universidade de Vigo	Creole-Based	Papiamento
13	Parafita Couto	M Carmen	Leiden University / Universidade de Vigo	Indo-European	Welsh
14	de-Dios-Flores	Iria	Universidade de Santiago de Compostela	Indo-European	Galician
15	Rajeg	Gede Primahadi Wijaya	Universitas Udayana	Austronesian	Indonesian
16	Rajeg	Gede Primahadi Wijaya	Universitas Udayana	Austronesian	Balinese
17	Lopukhina	Anastasiya	HSE University	Indo-European	Russian
18	Faghiri	Pegah	University of Amsterdam	Indo-European	Persian
19	Garcia	Azler	University of the Basque Country	Basque	Basque
20	Ribu	Ingeborg Sophie	Oslomet - Oslo Metropolitan University	Indo-European	Norwegian
21	Sehyr	Zed	San Diego State University	Sign Language	American Sign Language
22	Arvindam	Vishal	UC Santa Cruz	Dravidian	Telugu
23	Khanova	Aigul	HSE University (Higher School of Economics)	Altaic	Tatar
24	Dal Ben	Rodrigo	Ambrose University	Indo-European	Portuguese
25	Botezatu	Roxana	University of Missouri	Indo-European	Spanish
26	Botezatu	Roxana	University of Missouri	Sino-Tibetan	Mandarin
27	Botezatu	Roxana	University of Missouri	Indo-European	Dutch
28	Lahiri	Aditi	University of Oxford	Indo-European	Bengali
29	Wetterlin	Allison	University of Agder, Norway	Indo-European	Norwegian

Figure 1. Spreadsheet database created through recurring Google Form posts on social media.
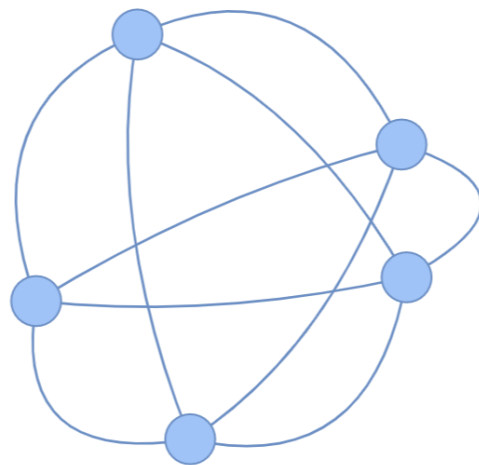

Figure 2. A low-level mockup of a network visualization. VIDUL will utilize this structure to offer an interactive database of researchers studying understudied languages.

[1] Blasi et al., 1988. *Trends in Cognitive Science.*