Investigating the Use of Native and Secondary Language with Data Visualization in Madagascar

Henintsoa S. Andrianarivony[*1], Taratra D. Raharison[†1], I. Tamby Rakotoniaina[‡1],
Mirindra T. N. Ramarokoto[§1], Noëlle Rakotondravony[¶2], and Lane Harrison[‖2]

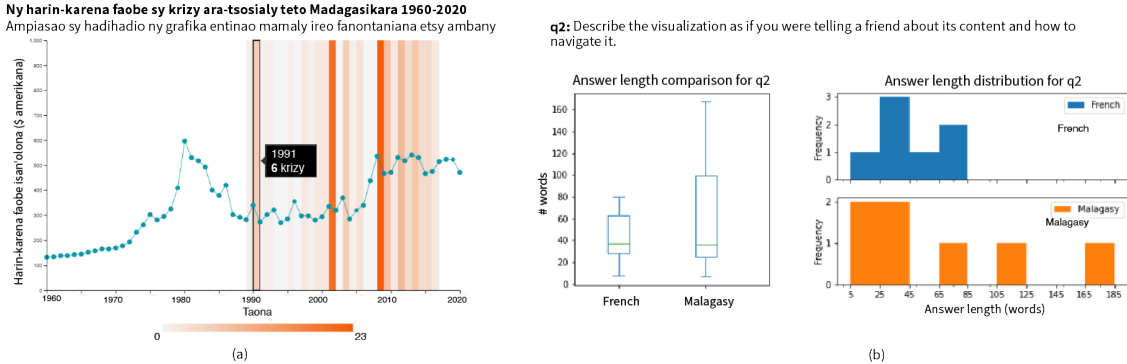[1]Université d'Antananarivo
[2]Worcester Polytechnic Institute

Figure 1: (a) The visualization experiment in Malagasy showing the GDP/capita and social conflicts between 1960 and 2020. (b) A comparison of answer lengths for **(q2)** and the distributions of word count used by participants in French and in Malagasy.

ABSTRACT

In this preliminary work, we investigate the impact of native versus secondary language in the design, exploration, and interpretation of data visualization. We focus on bilingual speakers in Madagascar, a predominantly Malagasy-speaking country with French as a common secondary language. In a between-subjects online study, $n = 14$ participants answered open-ended questions on how they interact with and communicate about a data visualization delivered in either their native (Malagasy) or secondary (French) language. Results suggest that a lack of expressive terms for visualization, data, and technical terminology in Malagasy impacted participant responses, with several using French terms even when prompted to answer in Malagasy, along with differences in answer length, fluency, and conciseness.

1 INTRODUCTION

With the democratization of interactive visualizations and their authoring tools, more interactive visualizations are created with the purpose of helping people make sense of or communicate about data. Visualization creators often align the methodologies to specific attributes of the target audience, such as their cultural values and languages which help identify paradigms and metaphors that can shape an interface design [2, 7]. Additionally, the use of and communication about data and visualization can reflect audiences' familiarity with the visualizations, statistical and numeracy skills [4, 5], or the availability of linguistic tools for verbalizing visualization [3].

For many countries, such as several African countries, concepts and terminologies that are essential part of data visualization practices are often in languages most used for science, technology and education. Such languages are in most cases secondary, and access to them can be limited and is not granted to all populations. Exploring this interplay of language therefore can surface interesting aspects in the study and practice of data visualization in native language vs. secondary language.

In this preliminary study, we investigate the different characteristics related to the design and exploration of data visualization in Madagascar. In Madagascar, the native language is Malagasy, and French is used in education and for most technology-related topics[1]. While only about 23% of Malagasy speak French, access to data and public online resources such as educational websites and national statistics remain in French [1]. The same goes for the practice and use of interactive technologies and the education of science.

We design an online study in two languages (Malagasy and French) with which we investigate: (i) how people scan elements on a graph and visualization, (ii) how people talk about visualization components and its interaction features, (iii) how people reason about the data behind a visualization delivered in Malagasy vs. in French. Initial result of a between-group analysis shows differences between the way people use and talk about data visualization. It may stem from the availability of data and visualization-related terminology in the respective language and culture.

2 METHODOLOGY

We design an online study in both Malagasy and French. After an introduction to the study and the content of the data visualization,

[*]e-mail: henintsoa.s.andrianarivony@gmail.com
[†]e-mail: taratra.d.raharison@gmail.com
[‡]e-mail: i.tambyrakotoniaina@gmail.com
[§]e-mail: ramarokoto.mirindra@yahoo.com
[¶]e-mail: ntrakotondravony@wpi.edu
[‖]e-mail: ltharrison@wpi.edu

[1]Le Français langue d'enseignement, accessed: 2022-06-26

participants access and interact with the designed visualization and answer a set of questions.

2.1 Visualization and study design

Our goal for the visualization is to be reasonably complex and to help elicit interpretations and insights. We visualize two datasets: the GDP per capita in Madagascar between 1960 to 2020[2] and the recorded social conflicts between 1990 and 2017 (source: [6]). The definitions of both topics are provided to participants at the beginning of the experiment. The GDP per capita is visualized using a linechart with dots marking the values in USD at each year. The count of social conflicts per year is visualized using colored stripes, in which a darker color corresponds to a higher number of conflicts (see Figure 1 (a)).

The visualization is interactive. Hovering over a dot in the line chart reveals the GDP per capita and year pair, and hovering over a stripe shows the count of social conflicts in a year. The study was designed and translated by native Malagasy speakers who are fluent in French, and the instructions are kept as aligned as possible in both languages.

2.2 Experiment procedure and participants

To investigate the interplay between language and data visualizations, we transform the investigation goals in Section 1 into the following study questions. (Questions are translated in English)

- (i) → (**q1a**) Which year has the second highest number of social conflicts? (**q1b**) Which year has the second highest GDP? explain how you find your answers.
- (ii) → (**q2**) Describe the visualization as if you were telling a friend about its content and how to navigate it.
- (iii) → (**q3**) Describe the fluctuation of GDP per capita in line with the recorded social conflicts in Madagascar between 1960 and 2020.

Participants in the study are recruited through the authors' professional contacts. Each participant is given only one version, in Malagasy or in French. Participants are encouraged to use the language and terms they are most comfortable with in their answers. The study took between 10 minutes to 20 minutes depending on the subject's familiarity with data visualizations.

3 ANALYSIS AND DISCUSSION

We recruited $n = 7$ participants for the Malagasy version (2 female, 5 male, avg age = 31, 4 use Malagasy mostly in their everyday life, 6 reported to have some experience and exposure to data visualization from classes, or professional experiences such as statisticians, survey analysts, and visual designers) and $n = 7$ for French (3 female, 4 male, avg age = 26.85, 4 use Malagasy mostly in their everyday life, 6 users reported to have received a graduate level education, 3/4 participants have beginner-level knowledge of data visualization through professional experience in marketing and storytelling). The dataset and analysis files are provided as supplemental materials.

When preparing the study, a lack of precise terminology to describe elements of visualization and interaction tasks has challenged the design of the questionnaire, especially in Malagasy. When answering **q1a** and **q1b**, 4/7 participants could identify the correct answers in the Malagasy experiment, against 6/7 participants in French. When reporting how they arrived to their answers, 1/7 participants in each language actually referred to the use of the available interaction in the visualization. A participant in Malagasy reports:

> **P2-mg:** "[...] Nalefako teo ambonin'ilay <u>barre</u> faharoa farany marevaka ny <u>pointeur</u> dia nampiseho ilay isa sy

ny daty mifandraika aminy" ("I moved the <u>cursor</u> over the second <u>bar</u> with darkest color to show the number and date corresponding to it")

In the collected answers to (**q2**), participants in French appear to be more concise in their descriptions and use of data-related terms (Avg. 44 words/participants). Participants in Malagasy tend to elaborate more (Avg. 65 words/participants), often with verbose descriptions of visualization-related terms (see Figure1 (b)). They reported struggling to find adequate Malagasy expressions and terminologies to talk about the visualization. In most cases, this posed a challenge in reporting about the data visualization, and we observed an extensive use of French terminologies in the Malagasy dataset, especially when referring to concepts related to computers, interactive interfaces, the data visualization, and the data.

Interestingly, with (**q3**), participants also refer to the data and the relationship between the GDP fluctuation and the presence of social conflicts using mostly French terminologies, even if in the visualization itself all references to the data exist only in Malagasy. For both experiments, we suspect that the difference in visualization literacy and familiarity is related to the availability of linguistic tools specific to data visualizations in general, and impacts participants' preference in the format in which the data is presented. For instance, while participants in the French experiments did not comment the practicality of data visualization, one participant in Malagasy **P5-mg**, expert in survey design and data collection, suggested to use tables with numerical values instead of our visualization. He also pointed out that while initial documentation exists at the National Institute of Statistics' archive, Malagasy terminologies for designating graphs are rarely employed.

4 CONCLUSION

Our study suggests that in Madagascar, a country with two major languages, communicating and designing data visualizations are easier for participants in French (language of education) than in Malagasy (native language). The main challenges emerge from the lack of native terminologies and linguistic tools to use, causing distraction and challenges that can impact both the design and the use of a given data visualization. This disfluency and bias towards French in important data and communication may be amplified when considering that a sizable majority (over 80%) of people in Madagascar speak only Malagasy. Future studies should further investigate data and visualization-focused reasoning strategies in monolingual local and native populations, and uncover other possible barriers to access to the state-of-the-art that visualization has to offer.

REFERENCES

[1] Institut National de la Statistique de Madagascar (National Institue for Statistics Madagascar). `https://www.instat.mg/`. Online; accessed: 2022-06-23.

[2] V. Evers. Cross-Cultural Understanding of Metaphors in Interface Design. *Proceedings CATAC'98, Cultural Attitudes towards Technology and Communication*, pp. 1–11, 1998.

[3] R. Henkin and C. Turkay. Words of Estimative Correlation: Studying Verbalizations of Scatterplots. *CoRR*, abs/1911.12793, 2019.

[4] H.-K. Kong, Z. Liu, and K. Karahalios. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018.

[5] C. Kreuzmair, M. Siegrist, and C. Keller. High Numerates Count Icons and Low Numerates Process Large Areas in Pictographs: Results of an Eye-Tracking Study. *Risk Analysis*, 36(8):1599–1614, 2016. doi: 10.1111/risa.12531

[6] I. Salehyan, C. S. Hendrix, J. Hamner, C. Case, C. Linebarger, E. Stull, and J. Williams. Social conflict in africa: A new database. *International Interactions*, 38(4):503–511, 2012.

[7] H. Winschiers-Theophilus and N. J. Bidwell. Toward an Afro-Centric Indigenous HCI Paradigm. *International Journal of Human-Computer Interaction*, 29(4):243–255, 2013.

---

[2]Data source: The World Bank database: GDP per capita (current US$) - Madagascar