

Situated Practices in the Construction of Language Technologies

Melissa Gasparotto^a

^aRutgers University, United States of America (USA)

gasparom@rutgers.edu

ABSTRACT

A number of recent studies have pointed to the lack of both standard documentation and discipline-based codes of ethics to explain the ways in which the practices of data science have resulted in allocational and representational harms to user communities of language technologies. Yet relatively few researchers have undertaken the in-depth study of these practices in context. The need to empirically document the practices of data scientists remains pressing, as digital language technologies will only continue to grow in importance for the expression of rights and identity online and offline. This poster presents work in progress toward an ethnographic study of data scientists working on indigenous language technologies in Mexico. A background literature review covers data science for language technologies, interdisciplinarity in data science, ethical codes in computing, and documentation practices of data scientists. This work is viewed through the lens of Huvila's (2009) ecology of information model, which supports the exploration of the situated and contextual affordances and constraints of information infrastructure and information work, both of which influence the possibilities for making knowledge claims. The results of the literature review suggest both structure and content for an interview protocol, which will lay the groundwork for further in-depth ethnographic work.

References

Huvila, I. (2009). Ecological framework of information interactions and information infrastructures. *Journal of Information Science*, 35(6), 695–708.
<https://doi.org/10.1177/0165551509336705>

ALISE RESEARCH TAXONOMY TOPICS

Natural language processing; Data science; Information ethics; Information system design; Sociology of information.

AUTHOR KEYWORDS

Computational linguistics; Ecology of information; Indigenous languages; Language technologies; Mexico.