

LREC 2014 Workshop

CCURL 2014

**Collaboration and Computing for Under-Resourced Languages
in the Linked Open Data Era**

Reykjavík (Iceland), 26th May 2014

Poster Presentation

Title

Collaborative Language Documentation: the Construction of the Huastec Corpus

Authors

Anuschka van 't Hooft, José Luis González Compeán

Abstract

In this paper, we describe the design and functioning of a web-based platform called Nenek, which aims to be an on-going language documentation project for the Huastec language. In Nenek, speakers, linguistic associations, government instances and researchers work together to construct a centralized repository of materials about the Huastec language. Nenek not only organizes different types of contents in repositories, it also uses this information to create online tools such as a searchable database with documents on Huastec language and culture, E-dictionaries and spell checkers. Nenek is also a monolingual social network in which users discuss contents on the platform. Until now, the speakers have created a monolingual E-dictionary and we have initiated an on-going process of the construction of a repository of written texts in the Huastec language. In this context, we have been able to localize and digitally archive documents in other formats (audios, videos, images), yet the retrieval, creation, storage, and documentation of this type of materials is still in a preliminary phase. In this presentation, we want to present the general methodology of the project.