

Creating an annotated corpus for Puno Quechua language

Elwin Huaman¹ University of Innsbruck, 6020, Innsbruck, Austria

1. Resource: Runasimi Vocabulary

I identified several sources, like Quechua dictionaries, vocabularies, and so on. One of the sources that I could use for the annotation task is the Quechua Wiktionary², it contains a list of words, phrases, senses, and forms. Furthermore, I have found a rich resource, the Runasimi Vocabulary³, which contains about 22866 Quechua words described in different quechua variants. I aim to create consistent subsets that can be annotated for the Quechua language.

2. Why the Training School

I have been creating a structure data source for the Quechua language, however I have been focusing on modelling, processing the data, and on storing the data. However, I neglected to explore treebank capabilities for the Quechua language, therefore, I am very interested on the talk: Annotation of Universal Dependencies treebank for a new language. In addition, I plan to annotate multiword expressions, but I do not have much experience on that aspect, so the talk: Annotation of multiword expressions in a new language, can provide me with a background to tackle this planned task. Finally, I believe that the talk: Corpus annotation infrastructure, can provide me with the background knowledge to explore corpus infrastructure from a linguistic perspective.

3. Open Questions

I propose to work on the Quechua language, however I would be open to collaborate with other participants in order to succeed during the entire Sumer School and the Hackathon.

4. Project Status

I have been working on Knowledge Graphs, most specifically, in the Quechua language Knowledge Graph, which contains 553636 triples. Furthermore, I am planning to add speech pronunciation, for the verbs of one specific variant of the quechua lexemes, during the following months.

References

¹[orcid=0000-0002-2410-4977](https://orcid.org/0000-0002-2410-4977), email=elwin.huaman@hotmail.com, url=<https://elwin.huamanquispe.com>

²<https://qu.wiktionary.org/>

³<https://www.runasimi.de/>