



**TECHNOLOGY
FOR
LANGUAGE REVITALIZATION**



THE ANNOTATION PROJECT



VOCES DEL VALLE

- A digital literacy project taking place in Tlacolula, Oaxaca
- Partnership between the CETis #124 high school and Haverford College
- High school students committed to tweeting several times a week for six weeks
- Students also participated in poetry and comic-making workshops
- External readers supported the students by retweeting, translating their tweets

20A LOCAL NOTICIAS NET.MX Domingo 21 de Febrero de 2016
NOTICIAS VOZ E IMAGEN DE OAXACA

DÍA INTERNACIONAL DE LA LENGUA MATERNA, 21 DE FEBRERO



El CETIS 124 desarrolla diversos proyectos culturales en favor de los estudiantes migrantes.
"Yo escribo en zapoteco", es el mensaje dirigido prácticamente a todo el mundo.
La estudiante Eunice Olivera señaló que difundir las costumbres y tradiciones de su comunidad es muy importante.

TLACOLULA DE MATAMOROS, Oaxaca.- Con el propósito de rescatar y difundir el zapoteco de esta región, el Haverford College, de la comunidad de Pensilvania, Estados Unidos, y el Centro de Estudios Tecnológicos Industrial y de Servicios (CETIS) Número 124 pasaron en marcha el proyecto Voces del Valle, a través del cual los jóvenes de ambas instituciones educativas intercambian tweets en zapoteco.

El director del CETIS 124, Marco Antonio Precya Rito, explica que en esencia el proyecto consiste en twitar en zapoteco por lo menos 10 veces a la semana durante dos meses, a estudiantes del Haverford College. En el mismo participan 8 estudiantes de la institución educativa oaxaqueña a quienes la doctora, Brook Danielle Lillehaugen, entregó en comodato igual número de iPads, aunque se busca que participen todos los jóvenes que hablan el zapoteco de los Valles.

Añade que otra particularidad es que los mensajes cuentan con un contenido literario, a fin de que en el corto plazo los emisores se conviertan en líderes culturales en sus propias comunidades o comunidades de inmigrantes en los Estados Unidos, por esta razón los mensajes deben estar relacionados con poemas, leyendas y cuentos zapotecos.

El trilingüe
Noel Gómez Sánchez nació en California. Estados Uni-

VÍA TUTTS, DIFUNDEN DESDE TLACOLULA SU LENGUA EN EU

Zapoteco, "tendencia" en las redes sociales

Ponen tecnología al servicio del zapoteco del valle; "cuelgan" poemas, leyendas, fotos y cultura en la web

Luis Ignacio Velázquez / Fotos: Mario Jiménez Leyva - Enviados



El grupo Voces del Valle tiene por propósito utilizar las nuevas tecnologías para preservar y difundir el zapoteco del Valle.

EN OAXACA
32.1% habla una lengua indígena
85.2% habla español
13.42% no habla español
66% no habla lengua indígena

Cifra hablantes de alguna lengua indígena en Oaxaca

FIRST VOCES DEL VALLE WORKSHOP







Moisés García Guzmán @BnZunni · Mar 8
#Usatuvos #ZapotecodeTlacochohuaya

Bel gú'uld Gugh rsilly,
Zed niss guia o naldaa guni chiy.
Zé'ey gwendilaaz de Bën Zunni.

If Zebra Doves calls are heard in the morning,
Either It will rain, or It will be a very hot day.
That's what Zapotecs of Tlacochohuaya believed.

Si la tortola canta temprano,
lloverá, o hará mucho calor ese día.
creencias de los Zapotecos de Tlacochohuaya.



ANNOTATING FOR DIGITAL HUMANITIES PROJECTS

We are going to get some hands-on experience doing annotation work. Annotation is the foundation of all digital humanities projects. Annotation is, at core, about getting data into a format that is useful.

Data is messy!

Data is complex!

Data often has implicit structure that we'd like to capture.

ANNOTATION PRACTICALITIES

- Data annotation often includes **metadata**: data about the data
- Our first step is to annotate our tweets with metadata:
 - Who were they tweeted by?
 - When where they tweeted?
 - Where where they tweeted?
 - Are there images included?
 - Who replied?
 - How many likes did the tweet get?
 - Was it retweeted? Link to retweets
 - Was it a retweet of another tweet? Link to original

PROPERTIES OF THE LANGUAGE

- Lots of tones
- Lots of vowel quality / phonation contrasts
- Verb - Subject - Object word ordering, but also lots of front-focusing, so OVS and SVO are also fairly common
- Tense and aspect: marked by prefixes on verbs; aspect is more common than tense
- A fair number of borrowings from Spanish that happened so long ago that they are not noticeable to native speakers
- Rich clitic system: pronouns often appear in clitic form on the end of verbs; many adverbials are also clitics
- Interesting possession system

LINGUISTIC ANNOTATION



LINGUISTIC ANNOTATION

Linguists annotate their data so that other linguists who are less familiar with the language can understand it. This is called **glossing** the data.

You saw examples of linguistic glossing last week:

OBJ	VERB	SUBJ	
gyia	rtyug	Lia	Petr
gyia	r-tyug	Lia	Petr
flower	HAB-cut	HON	Petra
'Petra cuts flowers .'			

LINGUISTIC GLOSSING

Glossing provides a **morpheme-by-morpheme** breakdown of each example.

A **morpheme** is the smallest morphological unit that has meaning.

Glossing means annotating each morpheme in a way that other linguists can understand.

OBJ	VERB	SUBJ	
gyia	rtyug	Lia	Petr
gyia	r-tyug	Lia	Petr
flower	HAB-cut	HON	Petra
'Petra cuts flowers .'			

LINGUISTIC GLOSSING CONVENTIONS

Glossing conventions are standardized in the Leipzig rules. These define standard abbreviations for different kinds of functional morphology.

For instance, the Leipzig abbreviation for **habitual aspect** is HAB, as in the example below.

OBJ	VERB	SUBJ	
gyia	rtyug	Lia	Petr
gyia	r-tyug	Lia	Petr
flower	HAB-cut	HON	Petra
'Petra cuts flowers .'			

LINGUISTIC GLOSSING CONVENTIONS

Glossing uses **small caps** for functional morphology (like the habitual aspect marker).

It uses **lower case** for open-class morphemes (like *flower*).

OBJ	VERB	SUBJ	
gyia	rtyug	Lia	Petr
gyia	r-tyug	Lia	Petr
flower	HAB-cut	HON	Petra
'Petra cuts flowers .'			

HOW DO YOU GLOSS SOMETHING?

So you have some data. How do you figure out what it says? This is a creative process, but these are roughly my steps:

1. Try to **identify the verb**. In Valley Zapotec, this is often the longest word. And it usually (but not always) has several consonants at the beginning.
2. Try to **figure out the nouns**. One is the subject. This may be a clitic on the verb. If there are more, one is probably an object of the verb.
3. **Look up the meaning of those words**. Nouns are easiest to look up, because they don't have any additional morphological inflection.

HOW DO YOU LOOK UP WORDS?

There are two resources for looking up San Lucas Quiaviní Zapotec words:

1) The Talking Dictionary:

<http://talkingdictionary.swarthmore.edu/sanlucasquiavini/>

2) The Cali Chiu textbook series.

I've uploaded a PDF of the fourth volume (a dictionary) on Moodle.

Please do not share this with anyone.

HOW TO GLOSS AN EXAMPLE

```
{  
  "name": "DizhSa",  
  "url": "https://twitter.com/DizhSa",  
  "date": "Apr 5",  
  "permalink": "https://twitter.com/DizhSa/status/1114214883661352962",  
  "likes": "Like\n5",  
  "retweets": "Retweet\n3",  
  "text": "Ra nde na tyop chon ra dishsa ni rgwe ra buny Ldua. Maru  
zyeiny buny ni rgwe Dizhsa. Here are some indigenous languages spoken  
in Oaxaca. Zapotec is the most spoken. Aquí están algunas lenguas  
originarias que se hablan en Oaxaca. El Zapoteco es el más hablado  
#usatuvoz #IYIL2019",  
  "id": 32  
},
```

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra dishsa ni rgwe ra buny Ldua.
Maru zyeiny buny ni rgwe Dizhsa.

Here are some indigenous languages spoken in
Oaxaca. Zapotec is the most spoken.

Aquí están algunas lenguas originarias que se hablan
en Oaxaca. El Zapoteco es el más hablado

#usatuvovoz #IYIL2019

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

dishsa ni rgwe ra buny Ldua.

dishsa ni rgwe ra buny Ldua.

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni rgwe Dizhsa.

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

? ? ? ? ? ?

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

? ? HAB-speak ? ? ?

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

? ? ? ? ? HAB-speak ?

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

? ? ? ? ? ?

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

? ? HAB-speak ? ? ?

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

? ? ? ? ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

? ? ? ? ? ?

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

? ? HAB-speak ? ? Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

? ? ? ? ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

? ? ? ? ? ?

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

language ? HAB-speak ? ? Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

? ? ? ? ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

? ? ? ? ? ?

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

language ? HAB-speak ? people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

? ? people ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra

Ra nde na tyop chon ra

? ? ? ? ? PL

dishsa ni rgwe ra buny Ldua.

dishsa ni r-gwe ra buny Ldua.

language ? HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.

Maru zyeiny buny ni r-gwe Dizhsa.

? ? ? people ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra
Ra nde na tyop chon ra
? ? ? two three PL

dishsa ni rgwe ra buny Ldua.
dishsa ni r-gwe ra buny Ldua.
language ? HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.
Maru zyeiny buny ni r-gwe Dizhsa.
? ? ? people ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra
Ra nde na tyop chon ra
? ? ? two three PL

dishsa ni rgwe ra buny Ldua.
dishsa ni r-gwe ra buny Ldua.
language ? HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.
Maru zyeiny buny ni r-gwe Dizhsa.
more/most ? people ? HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra
Ra nde na tyop chon ra
? ? ? two three PL

dishsa ni rgwe ra buny Ldua.
dishsa ni r-gwe ra buny Ldua.
language REL HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.
Maru zyeiny buny ni r-gwe Dizhsa.
more/most ? people REL HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra
Ra nde na tyop chon ra
? ? ? two three PL

dishsa ni rgwe ra buny Ldua.
dishsa ni r-gwe ra buny Ldua.
language REL HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.
Maru zyeiny buny ni r-gwe Dizhsa.
more/most many people REL HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra
Ra nde na tyop chon ra
PL this.one ? two three PL

dishsa ni rgwe ra buny Ldua.
dishsa ni r-gwe ra buny Ldua.
language REL HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.
Maru zyeiny buny ni r-gwe Dizhsa.
more/most many people REL HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

HOW TO GLOSS AN EXAMPLE

Ra nde na tyop chon ra
Ra nde na tyop chon ra
PL this.one NEUT-is two three PL

dishsa ni rgwe ra buny Ldua.
dishsa ni r-gwe ra buny Ldua.
language REL HAB-speak PL people Oaxaca

Maru zyeiny buny ni rgwe Dizhsa.
Maru zyeiny buny ni r-gwe Dizhsa.
more/most many people REL HAB-speak Zapotec

Here are some indigenous languages spoken in Oaxaca. Zapotec is the most spoken.

NEXT WEEK

- No class
- Optional reading: Lopez & Lillehaugen (2017)