1. CorpusXml : contiene oraciones con la anotación de lema, postag y afijos que se ha realizado.

1.1. Formato:

<corpus>

<sentence>

<word>

<affix/>

</word>

</sentence>

</corpus>

1.2. Detalle:

-Corpus: formado por un conjunto de oraciones (sentence)

-Sentence: formado por un conjunto de tokens (word)

id: orden de la oración en el corpus

text: oración en Shipibo

translation: traducción en Español

-Word: formado por un conjunto de afijos (affix)

id: orden del token en la oración.

lemma: lema del token.

posTag: categoría gramatical del token.

subPosTag: subcategoría gramatical del token.

token: token en Shipibo

-Affix:

order: < -oo , 0 ] prefijos, [ 1, oo > sufijos

text: afijo en Shipibo

type: tipo de afijo

2. Corpus Morfología: contiene oraciones procesadas del xml ( lemas + afijos )

3. procesarMorfología.py : script para procesar los xml de la carpeta "corpusXml"