Documento de estructura de corpus bilingüe Iskonawa-Español

1. Introducción

El presente documento presenta la estructura que se ha definido para un corpus bilingüe Iskonawa-Español a partir de transcripciones de conversaciones con hablantes nativos. Cada sección consiste en la descripción de los campos seleccionados para el corpus. Adicionalmente, se incluyen ejemplos de los datos con los que se cuenta.

2. Id

Se trata de un identificador único para cada oración del corpus. Consiste en el nombre del documento fuente anexado junto con un número que corresponde al orden de la oración. Cabe aclarar que no siempre se cuenta con oraciones enumeradas desde el 1, dado que no todas las oraciones de los documentos eran en Iskonawa.

Ejemplo: *EC-cancion.aku-2013_012*

3. Transcription

Es el campo principal del corpus pues contiene la oración de acuerdo con la transcripción obtenida del documento. Es importante mencionar que, debido a la naturaleza del origen de los textos, algunas oraciones elementos ininteligibles. Por otro lado, en este corpus no se cuenta con oraciones repetidas debido a que su fin es ofrecer ejemplos de traducciones.

Ejemplo: *kapa meken beneme*

4. Morpheme break

Esta sección contiene la segmentación morfológica de cada palabra de la transcripción correspondiente. El proceso de segmentación fue realizado por un lingüista experto en Iskonawa; sin embargo, no se realizó en todas las oraciones, por lo que existen elementos en los que este campo está vacío.

Ejemplo: *kapa meken bene =me*

5. **Pos**

En este campo se incluyen las *pos tags* correspondientes a cada morfema de acuerdo con la segmentación realizada en “morpheme break”, por lo tanto, se utilizan los mismos caracteres de separación como ‘=’ o ‘-‘. De igual manera, existen oraciones en las que no se realizaron las anotaciones correspondientes por parte del experto en Iskonawa.

Ejemplo: *n. n. n. =clit*

6. **Gloss es**

Este campo incluye la glosa de cada morfema, que proporciona una traducción al español o función del morfema respectivo en el contexto del enunciado. También se utilizan los mismos caracteres de separación que los 2 campos anteriores, y existen filas vacías ante la falta de anotaciones.

Ejemplo: *ardilla mano macho =?*

7. **Free translation**

Traducción libre de la oración correspondiente que captura el significado general de la misma en lugar de una traducción palabra por palabra como el campo “gloss es”.

Ejemplo: *mano de ardilla macho (hay metatesis)*

8. **File**

Se trata del archivo del cual se extrajo la oración y sus datos respectivos.

Ejemplo: *EC-cancion.aku-2013.txt*