

Búsqueda de Ejemplos Lingüísticos con Grew

June 6, 2025

El Desafío de la Minería de Ejemplos Lingüísticos

- Probar las predicciones de la teoría lingüística en corpus de lenguas con pocos recursos suele ser difícil.
- Las predicciones suelen depender en gran medida de la estructura.
- Esto hace que la búsqueda de ejemplos relevantes mediante una simple búsqueda por cadenas de texto sea difícil o imposible.
- Consideremos el caso de los *sufijos de estatus* en las lenguas mayas de la rama k'icheana.

¿Qué son los Sufijos de Estatus?

- Los sufijos de estatus son una característica particular de las lenguas mayas.
- Son sufijos que aparecen en verbos y en direccionales (que históricamente vienen de verbos).
- En k'iche', hay cuatro sufijos de estatus principales:
 - ik
 - oq
 - u ~ -o
 - u' ~ -a' ~ -o'
- Los dos últimos muestran armonía vocálica.

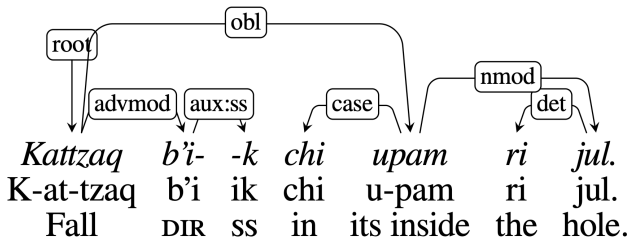
¿Qué son los Sufijos de Estatus?

El sufijo de estatus que toma un verbo está condicionado por una combinación compleja de factores morfosintácticos dentro de la cláusula. Estos incluyen:

- La **transitividad** del verbo.
- Si el verbo es un **verbo raíz** (forma CVC) o ha recibido derivación.
- El **tiempo-aspecto-modo** de la cláusula.
- Si la cláusula es una cláusula **independiente o dependiente**.

Un Ejemplo en Contexto

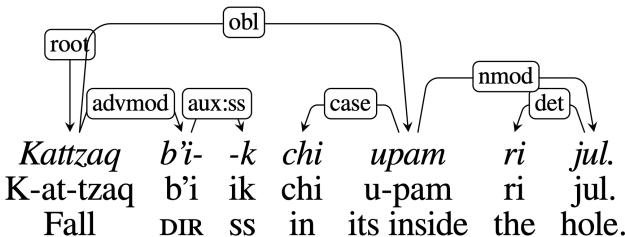
- En k'iche', los sufijos de estatus solo aparecen en los bordes de ciertas frases prosódicas.
- En el ejemplo de abajo, no hay pausa frasal entre el verbo *tzaq* ('caer') y el direccional *bi* ('hacia adentro DIR').
- Por lo tanto, solo el direccional lleva el sufijo de estatus, no el verbo principal, lo que en este caso indica que el direccional es intransitivo y la cláusula no es dependiente.



Estrategia de Anotación en UD

¿Cómo Representar los Sufijos de Estatus?

- Elegimos vincular los sufijos de estatus a sus verbos usando una variante de la relación aux: **aux:ss**.
- ¿Por qué? Son palabras funcionales que acompañan al verbo y expresan información aspectual y de modo, de manera similar a como funcionan los auxiliares verbales en otras lenguas.
- p. ej., para verbos intransitivos, cambiar el sufijo de estatus de *-ik* a *-oq* es suficiente para cambiar el modo de declarativo a imperativo.
 - Katwaroq! ¡Duerme!
 - Katwarik. Duermes.



Caso de Estudio: Sufijos de Estatus en Final de Frase en K'iche'

- **(Henderson, 2012):** Los sufijos de estatus que aparecen solo al final de la frase son sensibles a los límites de la frase intonacional, que se corresponden aproximadamente con los límites de la cláusula.
- **Generalización:** Un sufijo de estatus en final de frase solo debería aparecer si el verbo/direccional que lo lleva está:
 - 1 al final del enunciado,
 - 2 directamente antes de una cláusula incrustada, o
 - 3 directamente antes de un núcleo funcional que a su vez incrusta una cláusula.
- **Problema:** Encontrar contraejemplos requiere buscar oraciones que *no* satisfacen una descripción estructural, lo cual es imposible solo con la búsqueda por cadenas de texto.

Metodología: Análisis Sintáctico Automático y Consulta de Corpus

1 Creación de un Corpus Extendido:

- Se combinaron textos de varias fuentes, incluido el corpus Crúbadán de k'iche' (Scannel 2007).
- Esto resultó en un corpus de 16,196 oraciones (352,509 tokens).

2 Anotación Automática:

- Se procesó todo el corpus con el modelo UDPipe 2.0.
- El resultado, aunque no fue anotado a mano y lleva algunos errores, vamos a ver que es suficiente para nuestra tarea.

3 Consulta de Corpus:

- Se utilizó el lenguaje de consulta de corpus Grew para extraer estructuras específicas.

Esto muestra cuán poderoso es tener incluso un corpus UD relativamente pequeño. Podemos analizar texto nuevo que no ha sido anotado a mano, e incluso si no es perfecto, sigue siendo útil para la documentación y descripción del lenguaje.

La Consulta en Grew

Para probar la generalización, se diseñó una búsqueda para encontrar oraciones donde un verbo tenía un auxiliar dependiente con la relación `aux:ss` y un sustantivo con la relación `obj`.

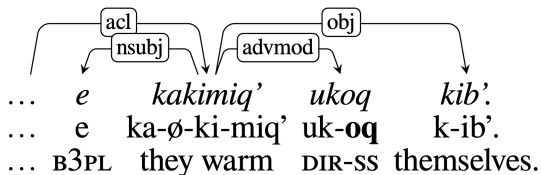
Búsqueda Grew Esquemática (27)

```
pattern {  
  V [upos=VERB] ;  
  A [upos=AUX] ;  
  N [upos=NOUN] ;  
  V -[aux:ss]-> A ;  
  V -[obj]-> N ;  
}
```

Esta estructura desafía directamente la hipótesis, ya que coloca un objeto *después* del sufijo de estatus, una posición donde el sufijo no es "final de frase".

Resultados: Un Contraejemplo

- La búsqueda descubrió una serie de ejemplos con estructuras aún no consideradas en la literatura sobre los sufijos de estatus.
- Esto incluyó contraejemplos directos a la generalización propuesta por Henderson (2012), ay mierda!



Contraejemplo

"Se quedaron allá con aquellos [que se estaban calentando]."

- El direccional lleva el sufijo de estatus *-oq*.
- Sin embargo, no está en un límite de cláusula, sino que ocurre antes de un pronombre reflexivo (*kib'*), que es una construcción de sustantivo relacional.

- El proceso de buscar encontró rápidamente un contraejemplo teóricamente interesante a una generalización prominente en la literatura lingüística.
- Esto demuestra la utilidad del treebank y el análisis sintáctico automático para la minería de ejemplos lingüísticos, incluso con una anotación imperfecta.
- Este enfoque abre nuevas vías para probar hipótesis lingüísticas complejas y dependientes de la estructura a una escala mayor de lo que permitiría el análisis manual.

Recuerden la generalización:

- **(Henderson, 2012):** Un sufijo de estatus en final de frase solo debería aparecer si el verbo/direccional que lo lleva está:
 - 1 al final del enunciado,
 - 2 directamente antes de una cláusula incrustada, o
 - 3 directamente antes de un núcleo funcional que a su vez incrusta una cláusula.
- Tarea: Piensen en algunas construcciones que deberían seguir la hipótesis o ir en contra de ella.
- Tarea: Dibujen una imagen esquemática de cómo se verían estas construcciones en UD.
- Tarea: Escriban una expresión en GREW para buscar esos (sub)árboles.
- Tarea: Busquen en el corpus de k'iche' por ellos.

El usanteko no tiene canónicamente sufijos de estatus que aparecen/desaparecen, pero los he oído en el habla (probablemente porque la mayoría de los hablantes de usanteko también hablan k'iche').

- Tarea: Identifiquen patrones que deberían producir sufijos de estatus (posiblemente reutilizando nuestros patrones del k'iche').
- Tarea: Modifiquen nuestras expresiones de GREW para ejecutarlas en usanteko (no tenemos aux:ss en el corpus de usanteko, p.ej.).
- Tarea: Busquen sufijos de estatus en el corpus de usanteko usando estos patrones.
- Tarea: Asimilen lo que han encontrado.