

Manual para la elaboración del Corpus de Paráfrasis

El Corpus de Paráfrasis estará constituido por una serie de textos que serán reescritos a partir de un texto base.

En esta ocasión, se realizará el segundo tipo de paráfrasis, la cual se denomina paráfrasis alta y se darán más detalles a continuación.

Es importante que los textos finales tengan coherencia y sean legibles para todo lector humano.

TEXTO ORIGINAL

El texto principal es un artículo que trata sobre un tema específico; éste documento consta de 25 partes que llamaremos oraciones, cada una está separada por un salto de línea. En este nivel de paráfrasis, se alterará la estructura de las oraciones para lograr el objetivo. Las técnicas que se mencionarán están basadas en los artículos *A German Corpus for Similarity Detection Tasks (2014)* y *Detección automática de plagio: de la copia exacta a la paráfrasis (2010)*.

PARÁFRASIS BAJA

En este nivel, no habrá cambios en la longitud del texto, es decir, el resultado deberá tener las 25 oraciones que se presentan en el texto original. Todos los cambios se harán dentro de la misma oración. A continuación, se muestran algunos recursos que se pueden usar para realizar la paráfrasis; el círculo negro muestra la oración original y el círculo blanco, las variaciones:

- **Sinónimos / antónimos:** El cambio de palabras es factible mientras no se modifique el significado de la oración.

- El **proceso** que **se empleaba** en China era **muy novedoso**.
- La **técnica** que **se usaba** en China era en **extremo nueva**.
- El **proceso** **usado** en China **no** era **muy antiguo**.

Una sola palabra también puede ser reemplazada por más de un elemento.

- Este método de conservación de pescados enteros...
- Esta forma de hacer duraderos los pescados...

- **Cambios sintácticos:** Mover de lugar los elementos de la oración.

- Hoy en día es conocida **en Japón** con el nombre de shiokara...
- Actualmente se le conoce como shiokara **en Japón**...
- **En Japón**, es conocida como shiokara actualmente...

- **Elisión, adición y reordenamiento:** Algunos fragmentos o palabras dentro de la oración se pueden omitir y también se puede agregar nuevos elementos, siempre que la información extraída o agregada no altere, en gran medida, el significado original.

- una especie de salsa de pescado (similar al garum romano).
- un tipo de salsa de pescado.

La elisión es recomendable, principalmente, cuando se trata de una enumeración; en la cual, también sus componentes pueden cambiar de lugar.

- La fermentación proporciona diversos compuestos químicos que aromatizan, dan textura y proporcionan sabor.
- ... se fermenta y da varios productos químicos que dan sabor y aroma.

También, cuando un verbo tiene como complemento otro verbo, se puede omitir si el resultado es coherente. De igual forma, se puede hacer lo contrario y añadir un verbo al complemento.

- Las poblaciones de la costa no necesitaban aplicar estos métodos de conservación...
- Las poblaciones de la costa no necesitaban estos métodos de conservación...

- **Sustitución de un elemento por un actante:** Sustituir algún elemento que representa una acción por otro que represente al actor de la misma:

- Uno de los países que acepta japoneses es EEUU para trabajar en la **construcción** de vías ferroviarias.
- Uno de los países que acepta japoneses es EEUU para trabajar como **constructores** de vías ferroviarias.

- **Sustitución palabra-definición:** Sustituir una palabra por su definición o descripción:

- El pescado es un alimento **perecedero**.
- El pescado es un alimento **que se echa a perder rápidamente**.

- **Números:** Su escritura puede ser por medio de cifras (números arábigos o números romanos) o de texto, cuando su uso sea posible.

- casi cerca de cuarenta mil japoneses inmigran a Estados Unidos.
- cerca de 40 mil japoneses se van a EEUU.
- En la época del **s. XVII** se inventaba el vinagre de arroz.
- En el **siglo diecisiete**, a partir de la receta antes dicha, surgió el vinagre de arroz.

Como se puede notar del último ejemplo, las **abreviaturas** también son otro recurso a tomar en cuenta.

Cuando se habla de años o épocas, puede haber variantes.

- en 1910
- a principios del siglo XX.
- en la primera década del s. XX

- **Hiperónimos e hipónimos:** Esto se refiere a emplear un término que englobe al significado de la palabra en cuestión, es decir, que sea más general.

- uno de los países que acepta japoneses es EEUU para trabajar en la construcción de **trenes**.
- en E.U. aceptaron a varias personas de Japón para que construyeran **medios de transporte**.

O, por el contrario, uno más específico.

- Después de ponerlo fuera de los **recipientes**...
- tras sacarlo de las **ánforas**...

- **Transformación y segmentación:** Se pueden usar cambios estructurales o sintagmáticos dentro de las oraciones, tales como el uso de voz pasiva:

- **El moho** se dedica a consumir los hidratos de carbono existentes en el arroz...
- Los hidratos de carbono existentes en el arroz son consumidos **por el moho**.
- Los hidratos de carbono existentes en el arroz son consumidos.

Cuando se usa la voz pasiva, el **actor** puede ser omitido, como se observa en el ejemplo anterior.

Asimismo, la puntuación puede ser alterada:

- Esta preparación rápida del sushi es la que se conoce mayoritariamente en la actualidad **y** su preparación rápida la ha convertido en un elemento de fast food.
- Esta preparación rápida del sushi es la que se conoce mayoritariamente en la actualidad. **Su** preparación rápida la ha convertido en un elemento de fast food.

Hay que tomar en cuenta que el trabajo no está limitado a estos recursos, se puede usar cualquier otro que no se encuentre en este manual, siempre y cuando no haya cambios en la longitud del texto, como ya se mencionó. La longitud de las oraciones es independiente y sí pueden ser más largas o más cortas que la original. Tampoco es necesario usar absolutamente cada uno de estos recursos.

Por último, todas las oraciones deben tener algún tipo de cambio con respecto a la original, esto es, ninguna puede quedar inalterada.

PARÁFRASIS ALTA

Este tipo de paráfrasis es más amplio debido a que el texto resultante debe tener cambios relevantes que lo distingan del original, pero sin afectar el tema o el significado.

Todas las técnicas usadas en la paráfrasis baja pueden ser empleadas en este nivel. También hay otras adicionales:

- **Juntar / separar oraciones:** A diferencia de la paráfrasis baja, aquí sí es factible cambiar el orden de las oraciones, juntar dos de ellas en una sola o, incluso, separar una oración en dos.

Supongamos que tenemos la primera y la tercera oración del texto original:

- [1] Durante el período del Imperio Mongol que se expandió por el territorio Chino, esta forma de conservar el pescado quedó parcialmente olvidada, quizás por ser las costumbres mongolas más carnívoras.
- [3] Las poblaciones de la costa no necesitaban aplicar estos métodos de conservación ya que disponían de pescado durante todo el año.

La referencia puede yacer en más de una oración del texto final. Después de un trabajo de paráfrasis alta, el texto resultante queda con el siguiente orden en las oraciones:

- [1] Por su parte, las poblaciones de la costa no precisaban de estas técnicas debido a la abundancia de pescado de la que gozaban a lo largo de todo el ciclo orbital.
- [2] Por otra parte, es muy probable que las carnívoras costumbres gastronómicas de los mongoles, hayan orillado al olvido esta técnica de

conservación del pescado en China durante la invasión sufrida durante el Imperio Mongol.

- [3] Pese a que durante el Imperio Mongol – que llegó hasta China - esta técnica de conservación quedó un poco marginada (por los diferentes tipos de dieta en ambas culturas), en el Japón este alimento pudo sobrevivir como una técnica de conservación del pescado.

En este caso ocurren varios fenómenos. Primero se hizo un trabajo de paráfrasis baja. Posteriormente, se separaron las oraciones y se repartieron para generar una nueva estructura, por lo tanto, varios segmentos están repartidos en distintas oraciones. De esta manera, la estructura del texto parafraseado puede variar, la única restricción en este trabajo es que el texto parafraseado puede llegar a contener hasta 5 oraciones de más o de menos, esto es, quedar entre 20 y 30 oraciones. Pueden omitir algunas frases o incluir nuevas, pero se deben parafrasear todas las oraciones del original.

No hay limitaciones en cuanto a las oraciones que pueden juntarse o separarse, ni en el orden que estas pueden aparecer.

Es de **suma importancia anotar la referencia de cómo quedan las oraciones parafraseadas con respecto al texto original, es decir, si se juntaron, se separaron o quedaron igual**. Estas referencias serán escritas en un archivo de texto aparte. La notación es la siguiente: deben hacer un listado con el número de las 25 oraciones originales, y luego de dos puntos anotar el número de la oración en la que quedó parafraseada.

El número que se encuentra antes de los dos puntos se refiere a la oración del texto original y los que se encuentran después, a las oraciones del texto final. Por ejemplo, si la oración 1 del original corresponde a la 3 de las oraciones parafraseadas, se representaría como "1: 3".

Si la oración 2 del texto original se divide en dos y corresponden a las oraciones 1 y 5 del texto final, esto se debe representar como "2: 1, 5".

Se debe anotar aún cuando no hubo cambios en la posición de los elementos, por ejemplo "1: 1"

Por consiguiente, la notación que se haría con el ejemplo que se dio para la paráfrasis alta quedaría de esta forma:

1: 2, 3

3: 1, 3

...

Y así en adelante.

La notación con los números entre corchetes sólo es con el fin de ejemplificar, no es necesario numerar las oraciones en la tarea.

El molde para anotar las referencias será como el siguiente:

1:
2:
3:
4:
5:
6:
7:
8:
9:
10:
11:
12:
13:
14:
15:
16:
17:
18:
19:
20:
21:
22:
23:
24:
25:

No es necesario poner título ni que el archivo de texto tenga un nombre en específico.

La tarea se puede realizar y entregar en el formato en el que más se acomoden.