

6. Morfología

II. Teoría

6.1 Conceptos básicos

6.1.1 Morfología y palabra

6.1.2 Tipos de palabras

6.1.3 Tipos de morfemas

6.1 Conceptos básicos

- **Morfología:** estudio de la estructura interna de las palabras. Haspelmath & Sims (2010)
- **Palabra:** es una noción difícil de definir. Si se define bajo criterios ortográficos, es aquello que se encuentra entre dos espacios en blanco; sin embargo, no puede aplicarse el mismo criterio para la lengua oral, puesto que, en la cadena fónica, las pausas no aparecen entre palabras.

Otra problemática en su definición es que cada rama de la lingüística tiene su propio concepto, pues se enfocan en características diferentes.

6.1 Palabra

Bloomfield (1926 en Pena, 1999) definió la palabra como “forma libre mínima”, es decir que es una unidad que puede aparecer aislada en el habla. Pena considera que esta definición es insuficiente y agrega ciertas características que permiten identificar una palabra.

- La posibilidad de cambiar de posición en el enunciado
- La posibilidad de insertar otra unidad entre una palabra y otra
- La pausa potencial antes o después de ella y la imposibilidad de interrumpirla con una pausa en una conversación normal
- El orden fijo de los morfemas que la integran y la inseparabilidad de estos (La noción de morfema se abordará más adelante.)

6.1 Conceptos básicos

Ya que el uso de palabra puede resultar sumamente problemático, tanto en el estudio de la morfología como en el Procesamiento de Lenguaje Natural, existen dos nociones básicas más acotadas y que por ello resultan más convenientes:

- **Palabra-forma o forma flexiva:** coincide con la noción intuitiva de cada hablante de lo que es una palabra. Por ejemplo:

cantar, cantábamos, cantaríamos

- **Lema**

6.1 Conceptos básicos

- **Lema o forma citada de una palabra:** Es una entidad abstracta que se usa para referirse a todas las formas concretas de una palabra, es decir, en todas su formas flexivas. Se puede decir que es la forma de la entrada de un diccionario.

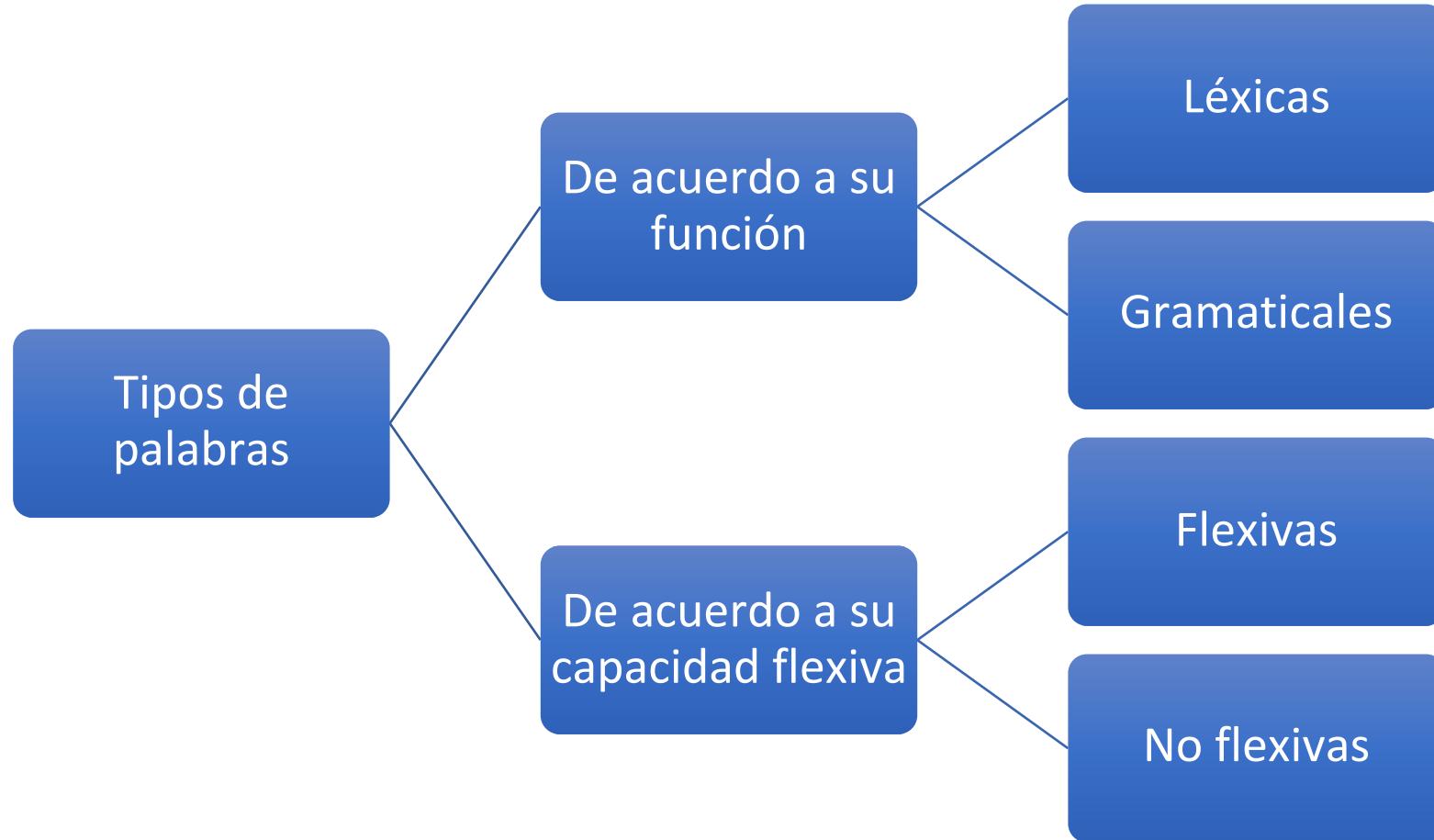
Formas flexivas	Lema
<ul style="list-style-type: none">• <i>cantas</i>• <i>cantábamos</i>• <i>cantarían</i>	CANTAR
<ul style="list-style-type: none">• <i>suaves</i>• <i>suavecita</i>• <i>suavecitos</i>	SUAVE

6.1 Tipos de palabras

Antes de abordar las clases de palabras y las categorías gramaticales es importante reconocer dos nociones: las clases abiertas y clases cerradas.

- **Clases abiertas:** conjuntos grandes de palabras a los que se puede, además, agregar nuevas palabras.
- **Clases cerradas:** conjuntos limitados de palabras a los que no se puede agregar nuevas palabras.

6.1 Tipos de palabras



6.1 Tipos de palabras

- **Palabras léxica o de contenido:** clases abiertas de palabras que aportan información semántica central.
 - Sustantivos
 - Verbos no auxiliares
 - Adjetivos
 - Adverbios
- **Palabras gramaticales o funcionales:** clases cerradas de palabras que aportan significados gramaticales o abstractos que resultan relevantes para la sintaxis como referencia, pluralidad, etc.
 - Artículos
 - Pronombres
 - Preposiciones
 - Verbos auxiliares
 - Conjunciones

6.1 Tipos de palabras

- **Palabras flexivas o variables:** aquellas palabras que permiten algún tipo de proceso morfológico como flexión o derivación
 - Artículos (*la, las*)
 - Adjetivos (*grandes, grandísimas*)
 - Sustantivos (*doctor, doctora, doctoras*)
 - Verbos (*bailo, bailaron, bailaríamos*)
- **Palabras no flexivas:** aquellas que no permiten ningún cambio en su forma
 - Pronombres (*yo, tú, él*)
 - Preposiciones (*a, contra, para*)
 - Conjunciones (*y, o, pero, sino*)
 - Interjecciones (*ay, adiós, olé*)
 - Adverbios (*lento, rápido, furiosamente*)

6.1 Categorías gramaticales (palabras léxicas)

- **Sustantivos:** expresan la pertenencia de las cosas a una clase si son comunes; si son propios, identifica una cosa entre los elementos de su clase. (Bosque, 1999)

• <i>libro</i>	• <i>arena</i>	• <i>Juan</i>
• <i>calidad</i>	• <i>ejército</i>	• <i>parvada</i>

- **Verbos:** son palabras que refieren a acciones, cambios de estado, estados mentales, etc. (RAE, 2010)

• <i>bailar</i>	• <i>hervir</i>	• <i>asegurar</i>
• <i>cocinar</i>	• <i>creer</i>	• <i>caer</i>

6.1 Categorías gramaticales (palabras léxicas)

- **Adjetivos:** describen propiedades o cualidades y tienen como función predicar o ser atributos de los sustantivos. (Di Tullio, 2005)
 - *La casa es roja.*
 - *La casa roja*
- **Adverbios:** modifican un verbo, un adjetivo u otro adverbio. (Van der Auwera, 1999)
 - *Llamó desesperadamente a la puerta.*
 - *El trabajo es enteramente suyo.*
 - *Llegamos sorpresivamente rápido.*

6.1 Categorías gramaticales (palabras funcionales)

- **Artículos:** restringen y definen la referencia de los nominales. (Leonetti, 1999)

• <i>la</i>	• <i>una</i>	• <i>esto</i>
• <i>los</i>	• <i>unas</i>	• <i>estas</i>

- **Pronombres:** se les considera una subclase de sustantivos con características semánticas y funciones sintácticas distintas. Pueden ser personales, posesivos, demostrativos, relativos, interrogativos, indefinidos y cuantitativos. (Di Tullio, 2005)

• <i>usted</i>	• <i>esta</i>	• <i>cuál</i>
• <i>mío</i>	• <i>cuyo</i>	• <i>algo</i>

6.1 Categorías gramaticales (palabras funcionales)

- **Preposiciones:** palabras que introducen un complemento, típicamente un nominal (RAE, 2010)

• <i>a</i>	• <i>desde</i>	• <i>para</i>
• <i>con</i>	• <i>hacia</i>	• <i>sin</i>

- **Verbos auxiliares:** ocurren con otras formas verbales plenas y no pueden aparecer independientemente. (Steele, 1999)

- *Hemos construido quinientas escuelas.*
- *Tres personas fueron rescatadas*

- **Conjunciones:** expresan relaciones de diversos entre enunciados: temporales, causales, lógicas, etc

• <i>cuando</i>	• <i>porque</i>	• <i>si</i>
• <i>una vez que</i>	• <i>puesto que</i>	• <i>y</i>

Ahora que tenemos claro qué es una palabra y los tipos de palabras que existen, podemos estudiar la estructura interna de estas.

A continuación se presentan las unidades del análisis morfológico.

6.1 Conceptos básicos

- **Morfema:** unidad significativa y gramatical mínima.
- **Raíz (*stem*):** la forma de la palabra sin afijos que resulta irreductible.
cant en *cantamos, cantaríamos, cantaste*
- **Base:** constituyente de una palabra sobre el que puede realizarse un proceso morfológico.

En los siguientes ejemplos *blanc* es la raíz de todas las palabras y la base se marca con negritas.

- *blanc-a* > ***blanca-s***
- *emblanquec-e-r* > ***emblanquece-mos***

(Pena,

1999)

6.1 Afijos

- **Afijos:** morfemas ligados, es decir que se añaden a una base. Estos se clasifican de acuerdo a su posición con respecto de la base.

Tipos de afijos

- Prefijos
- Sufijos
- Circunfijos
- Infijos
- Interfijos

6.1 Tipos de afijos

- **Prefijos:** se colocan antes a la base.
des- en *despeinar*, *-ante* en *anteponer*
- **Sufijos:** se colocan después de la base.
-oide en *humanoide*, *-dizo* en *escurridizo*, *-a* en *doctora*, *-s* en *sapos*
- **Circunfijos:** un prefijo y un sufijo que necesariamente coocurren con un solo significado.
en ... er en *enaltecer*
a ... ar en *alivianar*

(Bauer, 2004)

6.1 Tipos de afijos

- **Infijos:** ocurren en medio del morfema al que se adhieren, el cual es usualmente la raíz. Algunos consideran *-it-* en *azuquitar* un infijo y no una ocurrencia del sufijo diminutivo *-it*, puesto que no produce la forma *azucarita*; sin embargo, éste es el único caso en el que podría hablarse de un infijo en el español.
- **Interfijos:** se colocan entre dos morfemas para unirlos y por lo general carecen de significado propio. Éstos no ocurren en el español.
 - s- en la palabra alemana *Liebesgeschichte* (*historia de amor*), que une los morfemas *Liebe* (*amor*) y *Geschichte* (*historia*).

6.2 Procesos morfológicos

6.2.1 Morfología flexiva

6.2.2 Morfología léxica

6.2.3 Morfología no concatenativa

6.2.4 Clitización

6.2.5 Productividad morfológica

6.2 Tipos de procesos morfológicos

6.2.1 Morfología flexiva

6.2.2 Morfología léxica

- Derivación
- Composición

6.2.3 Morfología no concatenativa

6.2.4 Clitización

6.2 Morfología flexiva

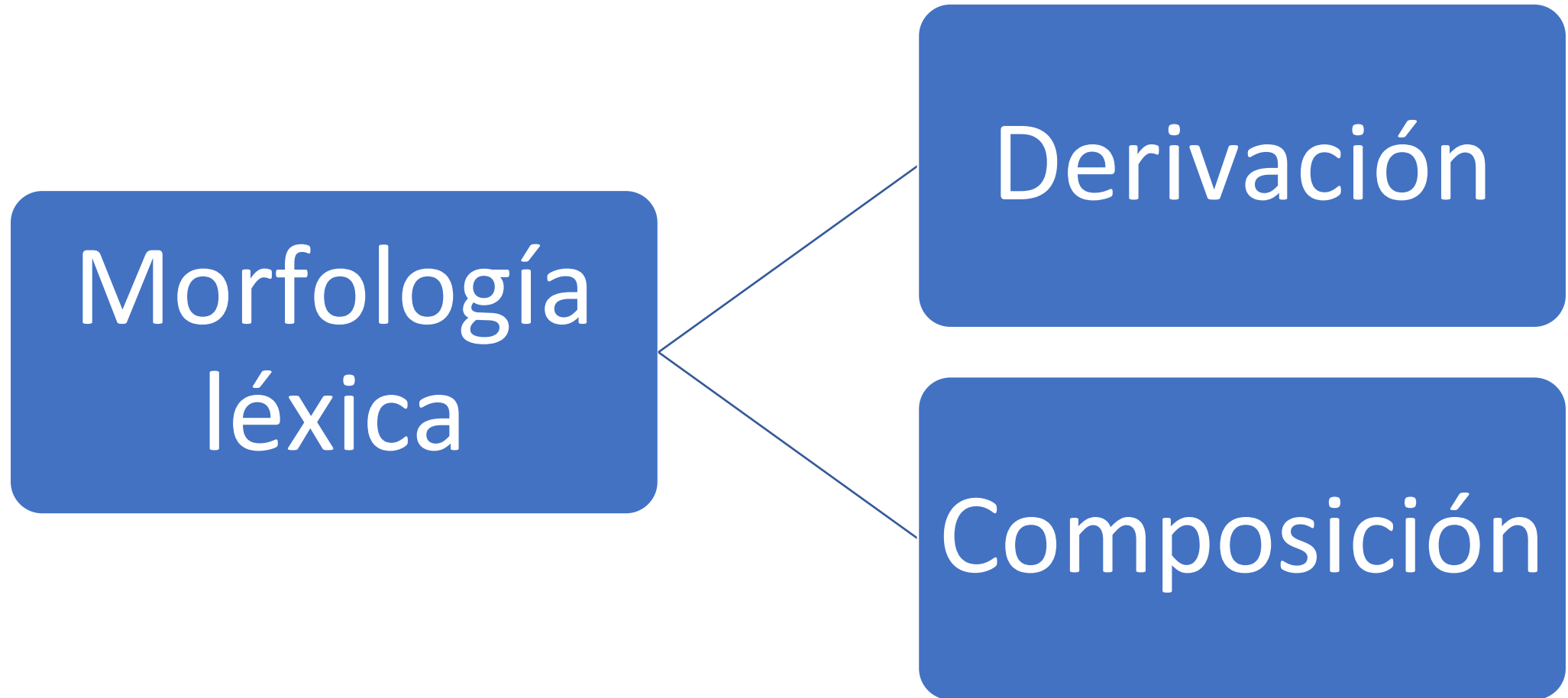
- **Flexión:** cambia la forma de la palabra con el fin de expresar un significado gramatical, como puede ser la concordancia de género o número en el caso de los sustantivos, o la concordancia de persona y número en el caso de los verbos, sin que cambie la categoría gramatical de la palabra es decir, si se modifica un adjetivo, el resultado será igualmente un adjetivo. (Pena, 1999)

Tenemos así los morfemas de género (-a, -o, -e) y los de plural (-s, -es) que se sufijan a la raíz del adjetivo para concordar con el género y el número del nominal:

Libro roj-o

casas blanc-a-s

6.2 Morfología léxica



6.2 Morfología Léxica

- **Derivación:** proceso que permite crear nuevas palabras agregando afijos con significado léxico a una raíz. Este proceso puede cambiar la categoría gramatical de la base; por ejemplo, se puede crear un sustantivo a partir de un verbo.

Verbo > Sustantivo	<i>jurar > juramento</i>
Adjetivo > Sustantivo	<i>desnudo > desnudez</i>
Sustantivo > Adjetivo	<i>cristal > cristalino</i>
Verbo > Adjetivo	<i>comer > comestible</i>



Actividad formativa

- Piensa en algunos ejemplos de sustantivos que deriven de adjetivos.
- Piensa en algunos ejemplos de verbos que deriven de sustantivos.
- Piensa en algunos ejemplos de palabras cuya derivación no cambie la categoría gramatical.

Respuesta

Algunos ejemplos:

- *suave > suavidad, transparente > transparencia, grande > grandeza*
- *almacén > almacenar, cristal > cristalizar, favor > favorecer*
- *líder > liderazgo, dormir > dormitar, guapo > guapísimo*

6.2 Morfología Léxica

- **Composición:** proceso morfológico mediante el cual se obtienen lexemas que contienen dos o más raíces sin ningún afijo derivacional. (Bauer, 2004)

tela

araña

hispano

hablante

abre

cartas

6.2 Morfología no concatenativa

Se le llama a todos aquellos cambios en la forma de una palabra que no implican afijación o composición, sino en cambios en la forma de la raíz como palatalización o sonorización de alguna consonante, alargamiento de una vocal, etc. (Haspelmath y Sims, 2010; Bauer, 2004)

Un ejemplo de morfología no concatenativa que puede resultar familiar es el proceso de ***ablaut*** o **apofonía** mediante el cual se forman los pasados y algunos participios irregulares en inglés. Como se observa, el cambio ocurre dentro de la raíz y no se adhiere ningún otro elemento.

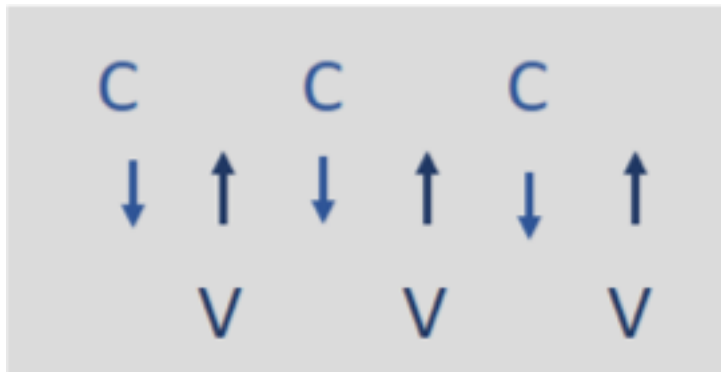
sink > sank > sunk

fall > fell

drink > drank > drunk

6.2 Morfología no concatenativa

Otro caso de morfología no concatenativa es la **morfología templética** de las lenguas semíticas como el hebreo o el árabe. En estas lenguas la raíz está formada por tres consonantes que se intercalan con vocales que forman transfijos.



(2.16)	ACTIVE PERFECT	PASSIVE PERFECT
	<i>kataba</i> 'wrote'	<i>kutiba</i> 'was written'
	<i>ḥalaqa</i> 'shaved'	<i>ḥuliqa</i> 'was shaved'
	<i>farada</i> 'decided'	<i>furiḏa</i> 'was decided'

Tomado de Haspelmath y Sims (2010)



Actividad formativa

- Descubre cuál es la raíz de *escribir*, *rasurar* y *decidir*, y los transfijos del activo y pasivo perfectos con los ejemplos de Haspelmath & Sims (2010).

(2.16) ACTIVE PERFECT

<i>kataba</i>	'wrote'
<i>ħalaqa</i>	'shaved'
<i>farada</i>	'decided'

PASSIVE PERFECT

<i>kutiba</i>	'was written'
<i>ħuliqa</i>	'was shaved'
<i>furiḏa</i>	'was decided'

Respuesta

- *escribir*: k-t-b
- *rasurar*: h-l-q
- *decidir*: f-r-d
- activo perfecto: -a-a-a
- pasivo perfecto: -u-i-a

Recursividad morfológica

La morfología del español permite realizar diferentes procesos sobre una raíz; así se producen palabras como:

- *sals-er-o*
- *sub-acuá-tic-o-s*
- *arma-ment-íst-ic-o*



Actividad formativa

- Analiza morfológicamente las siguientes palabras e identifica su raíz:
 - *trigales*
 - *globalización*
 - *desabotonar*
 - *salsacumbiendo*
 - *anticonstitucionalmente*

Respuesta

- *trig-al-es*
- *glob-al-iza-ción*
- *des-a-boton-ar*
- *salsa+cumbia-ando*
- *anti-constitu-cion-al-mente*

6.2 Clitización

- **Clíticos:** son elementos con características tanto de los afijos como de las palabras. Son semejantes a los afijos, ya que no son formas independientes (palabras plenas) y contienen información gramatical (Bauer, 2004), sin embargo tienen cierta libertad de movimiento y usualmente no provocan ni sufren un cambio fonológico (Haspelmath & Sims, 2010).

Un ejemplo de clítico es *'ll*, que a pesar de tener el mismo significado que *will*, siempre tiene que colocarse junto al sujeto de la oración y no en los mismos contextos en los que aparecería el auxiliar:

She'll call you back

**'ll she come to the party?*

6.2 Clitización

Como se mencionó anteriormente los clíticos se distinguen de los afijos por su libertad de movimiento. Un claro ejemplo de esto son los clíticos de objeto en el español:

Me lo voy a comer en la mañana.

Voy a comérmelo en la mañana.

6.2 Clitización

Aunque la clitización no es en sí un proceso morfológico, sino más bien ortográfico, resulta un problema de interés en el procesamiento de lenguaje natural al momento de la segmentación morfológica y la resolución de anáforas que se abordará en la sesión 7.

6.3. Métodos de segmentación morfológica

6.3.1 Importancia de la segmentación morfológica en PLN

6.3.2 *Stemming* y lematización

6.3.3 Algoritmo de Porter

6.3.4 Herramientas para *Stemming* y Lematización

6.3.5 Distancia de Levenshtein

6.3 Importancia de la segmentación morfológica en PLN

- Identificación de la lengua en la que está una palabra o un texto completo
- Checadores y correctores automáticos de ortografía y gramática
- Síntesis de habla (producción artificial)

6.3 Métodos en segmentación morfológica

Stemming

Lematización

¿Recuerdas cuál es la diferencia
entre raíz (*stem*) y lema?

Stemming es el método que permite reducir una palabra a su raíz (*stem*), removiendo todos los afijos y, en algunos casos, realizando alguna transformación a la base.



juntábamos > junt

pequeñitas > pequeñ

juntábamos > juntar

pequeñitas > pequeño



Lematización es el método por el cual se obtiene el lema de una palabra, es decir su forma no flexionada.

6.3 Algoritmo de Porter

- Es un algoritmo de *stemming* basado en reglas presentado por Martin Porter en 1980.
- Tiene como input una sola palabra, que es transformada en su raíz tras una serie de pasos.
- No siempre es preciso pero es el más usado en inglés.
- Ha sido adaptado a otras lenguas como el español.

Una noción importante en el algoritmo de Porter es la **medida** de una palabra (m), que indica cuántas sílabas tiene esta.

- Una secuencia de consonantes = **C**
- Una secuencia de vocales = **V**
- La **C** y la **V** iniciales son opcionales (no cuentan)

Cada palabra se representaría así:

[C] VCVC... [V] o

[C] (VC){ m }[V] (donde m es medida)

Ejemplos de medidas:

$m=0$: *o, la, el, WWE, AAA*

$m=1$: *casa, ave*

$m=2$: *alado, sombrilla*

$m=3$: *librería, caminero*



Actividad formativa

- ¿Cuál sería la medida de la palabra *elefantiasis*?



Respuesta

- *elefantiasis*: $m = 5$

	<i>el</i>	<i>ef</i>	<i>ant</i>	<i>ias</i>	<i>is</i>	
[C]	VC	VC	VC	VC	VC	[V]

6.3 Algoritmo de Porter

- La palabra inicial se checa con una secuencia de patrones de transformación en un orden específico. Un patrón sería:

$(m > 0)$ ATION \longrightarrow ATE *dedication* \longrightarrow *dedicate*

- Cuando una palabra coincide con un patrón, esta se transforma y el algoritmo vuelve a empezar desde el principio de la lista de patrones. El algoritmo para cuando la palabra deja de coincidir con un patrón, así, esa última palabra se convierte en el *output*.

Ejemplos de reglas en el algoritmo de Porter

1a

SSSES → S

SS → S

presses → *press*

press → *press*

1b

(m>0) EED → EE

refereed → *referee*

2

(m>0) ATIONAL → ATE

inflational → *inflate*

(m>0) BILITI → BLE

abiliti → *able*

3

(m>0) CERTIFICATE → IC

certificate → *certific*

(m>0) NESS → ∅

sadness → *sad*

4

(m>1) AL → ∅

arrival → *arrive*

(m>1) IC → ∅

academic → *academ*

IES → S

ponies → *poni*

S → ∅

lots → *lot*

(m>0) IZATION → IZE

realization → *realize*

(m>0) IVENESS → IVE

inventiveness → *inventive*

(m>0) ALIZE → AL

realize → *real*

(m>0) FUL → ∅

cheerful → *cheer*

(m>1) ANCE → ∅

abundance → *abund*

(m>1) IBLE → ∅

admissible → *admiss*

(snowball.tartarus.org, 2018a)

6.3 Algoritmo de Porter adaptado para el español

Definiciones

Vocales del español: a, e, i, o, u, á, é, í, ó, ú

- **R1:** después de la primera consonante precedida por una vocal o región nula al final de la palabra si dicha vocal no existe.
- **R2:** después de la primera consonante seguida por una vocal en R1 o la región nula al final de la palabra si no se presenta dicha consonante.
- **RV:** región después de la tercera letra o de la primera vocal que no es inicial o final. Si la segunda letra es consonante, RV es la región que sigue a la vocal siguiente. Si las dos letras iniciales son vocales, RV es la región después de la consonante siguiente.

(snowball.tartarus.org, 2018b)

0 Clítico (ME, SE, SELA, SELO, LA, LE, NOS) tico (ME, SE, SELA, SELO, LA, LE, NOS)

→ ∅

ÍENDO, ÁNDO, ÁR, ÉR, ÍR, ANDO, IENDO, AR, ER, IR, YENDO

→ ∅

1

• ANZA, ANZAS, ICO, ICA, ICOS, ICAS, ISMO, ISMOS, ABLE, ABLES, IBLE, IBLES, ISTA, ISTAS, OSO, OSA, OSOS, OSAS, AMIENTO, AMIENTOS, IMIENTO, IMIENTOS

→ ∅ si en R2

• ADORA, ADOR, ACIÓN, ADORAS, ADORES, ACIONES, ANTE, ANTES, ANCIA, ANCIAS

→ ∅ si aparece en R2 o precedidos por IC en R2

• LOGÍA, LOGÍAS

→ LOG si en R2

• UCIÓN, UCIONES

→ U si en R2

• ENCIA, ENCIAS

→ ENTE

• MENTE

→ ∅

• IDAD, IDADES

→ ∅ si en R2

• IVA, IVO, IVAS, IVOS

→ ∅ si en R2 precedidas por AT

Realizar si no se ha removido nada en ①

②^a

YA, YE, YAN, YEN, YERON, YENDO, YO, YÓ, YAS, YES, YAIS, YAMOS

→ ∅ si son precedidos por U en RV

Realizar si no se ha removido nada en ②^a

• EN, ES, ÉS, EMOS

→ ∅ si son precedidos por GU.

Además, U → ∅

• ARÍA, ARÍAS, ARÁN, ARÁS, ARÍAIS, ARÍA, ARÉIS, ARÍAMOS, AREMOS, ARÁ, ARÉ, ERÍAN, ERÍAS, ERÁN, ERÁS, ERÍAIS, ERÍA, ERÉIS, ERÍAMOS, EREMOS, ERÁ, ERÉ, IRÍAN, IRÍAS, IRÁN, IRÁS, IRÍAIS, IRÉIS, IRÍAMOS, IREMOS, IRÁ, IRÉ, ABA, ADA, IDA, ÍA ARA, IERA, AD, ED, ID, ASE, IESE, ASTE, ISTE, AN, ABAN, ÍAN, ARAN, IERAN, ASEN, IESEN, ARON, IERON, ADO, IDO, ANDO, IEDNO, IÓ, AR, ER, IR, AS, ABAS, ADAS, IDAS, ÍAS, ARAS, IERAS, ASES, IESES, ÍS, ÁIS, ABAIS, ÍAIS, ARAIS, IERAIS, ASEIS, IESEIS, ASTEIS, ISTEIES, ADOS, IDOS, AMOS, ÁBAMOS ÍAMOS, IMOS, ÁRAMOS, IÉRAMOS, IÉSEMOS, ÁSEMOS

→ ∅ si son precedidos por U en RV

Siempre realizar ③

• OS, A, O, Á, Í, Ó

• E, É

→ ∅ si en RV

→ ∅ si en RV precedidos por GU

Además, si U en RV, U → ∅

Paso final

Acentos agudos

→ ∅

6.3 Herramientas para *stemming* y lematización

Si bien se pueden usar expresiones regulares para lematización y *stemming*, lo mejor es utilizar las herramientas que se encuentran disponibles.

El *Natural Language Toolkit* de *Python* ofrece el *stemmer* de Porter y el de Lancaster, aunque sus resultados suelen variar. (Bird, Klein y Loper, 2009)

```
>>> porter = nltk.PorterStemmer()
>>> lancaster = nltk.LancasterStemmer()
>>> [porter.stem(t) for t in tokens]
['DENNI', ':', 'Listen', ',', 'strang', 'women', 'lie', 'in', 'pond',
'distribut', 'sword', 'is', 'no', 'basi', 'for', 'a', 'system', 'of', 'govern',
'.', 'Suprem', 'execut', 'power', 'deriv', 'from', 'a', 'mandat', 'from',
'the', 'mass', ',', 'not', 'from', 'some', 'farcic', 'aquat', 'ceremoni', '.']
>>> [lancaster.stem(t) for t in tokens]
['den', ':', 'list', ',', 'strange', 'wom', 'lying', 'in', 'pond', 'distribut',
'sword', 'is', 'no', 'bas', 'for', 'a', 'system', 'of', 'govern', '.', 'suprem',
'execut', 'pow', 'der', 'from', 'a', 'mand', 'from', 'the', 'mass', ',', 'not',
'from', 'som', 'farc', 'aqu', 'ceremony', '.']
```

Métodos de lematización

- Una herramienta disponible en Internet es el flexionador y lematizador de palabras del español, elaborado por el Grupo de Estructuras de Datos y Lingüística Computacional de Las Palmas de Gran Canaria, en

<http://www.gedlc.ulpgc.es/investigacion/scogeme02/lematiza.htm>

6.3 Métodos de lematización

Existen diferentes métodos para obtener los lemas de un corpus.

- Eliminación de afijos (*stemming*) que ya fue abordado previamente y
- Búsqueda en diccionarios

6.3 Búsqueda en diccionarios

Este método consiste en utilizar diccionarios para obtener los lemas de un corpus.

- Ventajas:
 - Método sencillo
 - Puede ser rápido si se implementa la búsqueda con árboles
- Desventajas
 - Usar un diccionario requiere almacenamiento adicional que puede ser excesivo.

6.3 Distancia de Levenshtein

Es una medida de similitud entre dos cadenas de caracteres.

Dadas dos cadenas s y t , la distancia es el número de inserciones, substituciones o eliminaciones requeridas para transformar s en t .

- Ejemplo:

Si X es “gato” y Y es “gato” $LD(X,Y) = 0$

Si X es “gato” y Y es “gatote” $LD(X,Y) = 2$

Por las letras “t” y “e”

Por lo tanto, un valor grande de la distancia Levenshtein indica que las cadenas son muy diferentes.

6.3 Distancia de Levenshtein (Algoritmo)

Dadas s y t como cadenas de texto:

- Paso 1:
 - $n \leftarrow \text{length}(s)$; $m \leftarrow \text{length}(t)$
 - if $n == 0$ then return m ; if $m == 0$ then return n ;
 - Crear matriz de $n \times m$
- Paso 2:
 - Inicializar primer renglón con 0 hasta n
 - Inicializar primer columna con 0 hasta m

6.3 Distancia de Levenshtein (Algoritmo)

- Paso 3:
 - for($i = 1$; $i \leq \text{length}(S)$; $i++$)
 - for($j=1$; $j \leq \text{length}(T)$; $j++$)
 - if($S[i] == T[j]$) then $\text{cost} = 0$
 - else $\text{cost} = 1$
 - $a \leftarrow \text{matrix}[i-1,j] + 1$; $b \leftarrow \text{matrix}[i,j-1] + 1$; $c \leftarrow \text{matrix}[i-1,j-1] + \text{cost}$
 - $\text{matrix}[i,j] = \text{minimum}(a,b,c)$
- Paso 4:
 - Después de las iteraciones la distancia Levenshtein queda en $\text{matrix}[n,m]$

Referencias

- Bauer, L. (2004). *A Glossary of Morphology*. Washington D.C.: Georgetown University Press
- Bird, S., Loper, E. and Klein, E. (2009). *Natural Language Processing with Python*. Estados Unidos de América: O'Reilly Media Inc.
- Bosque, I. (1999) El nombre común. En Ignacio Bosque y Violeta Demonte (dirs.): *Gramática descriptiva de la lengua española*. (pp. 5, 6). Madrid: Espasa Calpe
- Di Tullio, Á. (2005) Manual de gramática del español. Buenos Aires: La Isla de la Luna
- Haspelmath, M., & Sims, A. D. (2010). *Understanding morphology* (2nd ed). London: Hodder Education.
- Leonetti, M (1999) El artículo. En Ignacio Bosque y Violeta Demonte (dirs.): *Gramática descriptiva de la lengua española*. (p. 787). Madrid: Espasa Calpe

- Pena, J.(1999). Partes de la morfología. Las unidades del análisis morfológico, En Ignacio Bosque y Violeta Demonte (dirs.): *Gramática descriptiva de la lengua española*. (pp. 4305-4366). Madrid: Espasa Calpe.
- Real Academia Española. (2010). *Nueva gramática de la lengua española. Manual*. Madrid, Espasa Libros.
- snowball.tartarus.org. (2018a). *The Porter stemming algorithm*. Disponible en línea en: <http://snowball.tartarus.org/algorithms/porter/stemmer.html> [13 de enero 2018].
- snowball.tartarus.org. (2018b). *Spanish stemming algorithm*. Disponible en línea en: <http://snowball.tartarus.org/algorithms/spanish/stemmer.html> [13 de enero de 2018].
- Steele, S. (1999). Auxiliaries. En Keith Brown & Jim Miller (dirs.). *Concise Encyclopedia of Grammatical Categories*. (p. 49). Gran Bretaña: Elsevier
- Van der Auwera, J. (1999) Adverbs and Adverbials. En Keith Brown & Jim Miller (dirs.). *Concise Encyclopedia of Grammatical Categories*. (p. 8). Gran Bretaña: Elsevier