

Introducción al Procesamiento de Lenguaje Natural

Grupo de PLN - InCo
2011



Morfología



Morfología

morfología.(De *morfo-* y *-logía*).

- f. Parte de la biología que trata de la forma de los seres orgánicos y de las modificaciones o transformaciones que experimenta.
- f. *Gram.* Parte de la gramática que se ocupa de la estructura de las palabras.



Morfología

f. *Gram.* Parte de la gramática que se ocupa de la estructura de las palabras.

Mecanismos de formación / análisis de la palabras.

Análisis morfológico: Reconocer una palabra y construir una representación estructurada

Gatitos -> gato+Masc+Pl+Dim



Morfología

- Morfema: mínima unidad con sentido
- Raíz: morfema “principal”
- Afijos: dan significado adicional
 - Prefijos: im+posible
 - Sufijos: Gat+ito+s
 - Circunfijos: En+guan+tar (Parasintéticas)
 - Infijos: No hay (hingi => humingi). Tagaloga ...
 - Un-f**ing-believ-able
- Lema: palabra representativa (IR)



Morfología

Dos problemas a resolver:

Formación de las palabras (Morfoláctica)

Las palabras están compuestas por unidades menores
(morfemas)

Los morfemas pueden combinarse de acuerdo a ciertas reglas
Imprescindible , *inelefante, inelefantemente?



Morfología

Dos problemas a resolver:

Alternaciones Ortográficas

Los morfemas pueden cambiar según el contexto

Pez => Pezs => Peces

Maní => Manís => Maníes



Morfología

Formas de combinar morfemas

Flexión

Derivación

Composición

Clitización



Morfología

Morfología Flexional

Mecanismo de producción de palabras dentro de una misma clase (y con igual semántica)

Las flexiones representan conceptos generales

Género / Número

Persona / Tiempo / Modo

Etiquetas Eagle, ejemplo...



Morfología

Morfología Derivativa

Combinar una raíz con un afixo, para generar una palabra de *otra* clase, o con *otro* significado.

estable (adj) => estabilizar (verbo) => estabilización (sustantivo)
=> desestabilización (sustantivo)



Morfología

Morfología Derivativa

La derivación es un mecanismo *productivo*

tokenización

inelefantezcamente

sindudamente (!?)

quizásmente



Morfología

Apenas él le amalaba el noema, a ella se le agolpaba el clémiso y caían en hidromurias, en salvajes ambonios, en sustalos exasperantes. Cada vez que él procuraba **relamar** las incopelusas, se enredaba en un grima-do quejumbroso y tenía que **envulsionarse** de cara al nóvalo, sintiendo cómo poco a poco las arnillas se espejunaban, se iban **apeltronando**, **reduplimiendo**, hasta quedar tendido como el trimalciato de ergomanina al que se le han dejado caer unas fílulas de cariaconcia. Y sin embargo era apenas el principio, porque en un momento dado ella se **tordulaba** los hurgalios, consintiendo en que él aproximara suavemente sus orfelu-nios. Apenas se **entreplumaban**, algo como un ulucordio les encresto-riaba, los extrayuxtaba y **paramovía**, de pronto era el clinón, la esterfu-rosa convulcante de las mátricas, la jadehollante embocapluvia del orgu-mio, los esproemios del merpasmo en una **sobrehumítica** agopausa. ¡Evohé! ¡Evohé! Volposados en la cresta del murelio, se sentían balpa-ramar, perlinos y márulos. Temblaba el troc, se vencían las marioplumas, y todo se **resolviraba** en un profundo pínice, en niolamas de argutendidas gasas, en carinias casi crueles que los ordopenaban hasta el límite de las gunfias. (*Rayuela* - Julio Cortázar)



Morfología computacional

Parsing (análisis): tomar una entrada y devolver algún tipo de estructura lingüística

Parsing Morfológico: tomar una palabra (forma de superficie) y devolver una representación de la raíz y sus morfemas.

gatito <-> *gato*+NCMS00D

Nombre. Común. Masculino. Singular. Diminutivo



Morfología computacional

Métodos:

1. Diccionario de formas flexionadas

amigo -> amigo + N + ms

amiga -> amigo + N + fs

amigos -> amigo +N + mp

...



Morfología computacional

Métodos:

1. Diccionario de formas flexionadas

! Se vuelven enormes

! No permiten palabras nuevas



Morfología computacional

¿Qué pasa con los lenguajes aglutinantes?

uygarlaştıramadıklarımızdanmışsınızcasına

civilizado

+BEC "convertirse en"

+CAUS "causar X"

+NABL "incapaz"

+PART "participio pasivo"...

+PL+P1PL+ABL+PAST+2PL+ASIF ... "etc."

"comportándote que si estuvieras entre aquellos que no pudimos
civilizar"



Morfología computacional

¿Qué pasa con los lenguajes aglutinantes?

El turco tiene 40,000 formas posibles

A eso hay que sumarle los sufijos derivativos

Teóricamente hay infinitas palabras (Causaste que X cause que Y cause que W.... haga Z)

Otros lenguajes aglutinantes: aymarará, euskera, finlandés

En español hay sufijos productivos: -mente, -s



Morfología computacional

Métodos:

2. Parser morfológico dinámico.

pavos -> pavo +N +Masc +Pl

superficie -> lema + rasgos

Necesitamos

1. Un lexicón: lista de raíces y afijos
2. Morfotácticas: modelo de ordenamiento de morfemas; que clases de morfemas pueden seguir a otros.
3. Reglas ortográficas: cambios en una palabra al combinar morfemas



Procesamiento morfológico humano

Hipótesis:

Listado completo.

Sabemos todo.

¿Y los turcos?

Mínima redundancia

Sólo tenemos los morfemas

Procesamos las morfotácticas

"()media loca" , "(*) son bastantes grandes"*



Procesamiento morfológico humano

Errores espontáneos

“Sara me dijo que me iba a llamar cada día y sólo me ha llamado esta vez una tarde”

“Me dijisteis que comprareis el coche”

“Estas arañas no pueden ser de picadura”

“Me labio el muerde”

“Muchas tardes, buenas gracias”

“¿Por qué me fijas tan miramente”

www.artifara.com/Rivista4/testi/errores.asp



Referencias

- J.Martin & D.Jurafsky. Speech and Language Processing - Capítulo 3
- K.Beasley, L.Karttunen. Finite State Morphology

