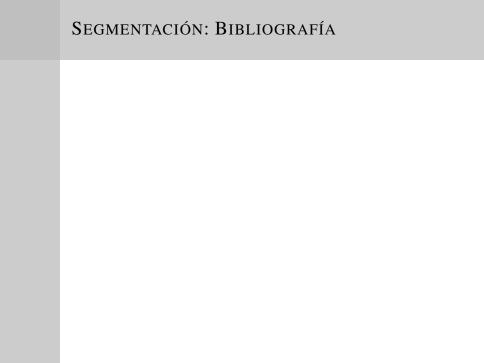
LINGÜÍSTICA COMPUTACIONAL TOKENIZING

Antonio Frías Delgado

Área de Lógica y Filosofía de la Ciencia Departamento de Historia, Geografía y Filosofía Universidad de Cádiz



SEGMENTACIÓN: BIBLIOGRAFÍA

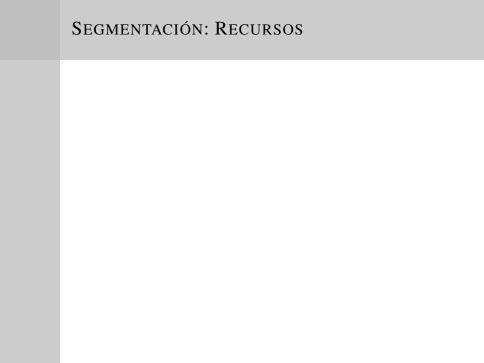
 What is a word, What is a sentence? Problems of Tokenization
Gregory Grefenstette, Pasi Tapanainen

SEGMENTACIÓN: BIBLIOGRAFÍA

- What is a word, What is a sentence? Problems of Tokenization
 Gregory Grefenstette, Pasi Tapanainen
- Text preprocessing David Palmer

SEGMENTACIÓN: BIBLIOGRAFÍA

- What is a word, What is a sentence? Problems of Tokenization
 Gregory Grefenstette, Pasi Tapanainen
- Text preprocessing David Palmer
- Tokenisation and sentence segmentation David Palmer



NLTK

■ import nltk

- import nltk
- nltk.word_tokenize(texto,language='spanish')

- import nltk
- nltk.word_tokenize(texto,language='spanish')
- nltk.wordpunct_tokenize(texto)

- import nltk
- nltk.word_tokenize(texto,language='spanish')
- nltk.wordpunct_tokenize(texto)
- nltk.sent_tokenize(texto,language='spanish')

■ import nltk

- import nltk
- from nltk.corpus import brown

- import nltk
- lacktriangledown from nltk.corpus import brown
- text_brown=brown.raw()

- import nltk
- from nltk.corpus import brown
- text_brown=brown.raw()
- oraciones_brown=brown.sents()

- import nltk
- from nltk.corpus import brown
- text_brown=brown.raw()
- oraciones_brown=brown.sents()
- palabras_brown=brown.words()