

Introducción al Procesamiento del Lenguaje Natural (NLP)

Ricardo Castro
Software Engineer
uBiome



www.linkedin.com/in/ricardo-castro-a386879

Contenidos

1. **Que es NLP ?**
2. **Por que nos interesa NLP ?**
3. **Aplicaciones Básicas**
4. **Aplicaciones Avanzadas**
5. **Desafíos**
6. **Tipos de Análisis**
7. **Preguntas**

Que es NLP ?

- Intersección entre ciencias de la computación, inteligencia artificial y lingüística
- NLP busca “entender” el lenguaje humano con el objetivo de crear aplicaciones interesantes.

Por que nos interesa NLP ?

- característica diferenciadora del reino animal.
- conocimiento humano está representado en texto
- gran número de aplicaciones en la actualidad



Aplicaciones Básicas

1. Construcción de Aplicaciones “user-friendly”
2. Realización de tareas específicas
3. Corrección ortográfica, búsqueda de palabras clave, sinónimos, antónimos.
4. Extraer opiniones de sitios de interés.
5. Análisis de Sentimiento.
6. Identificación de Intenciones

Aplicaciones Avanzadas

1. Traducción automática.
2. Sistemas de conversación (asistentes virtuales, chatbots)
3. Respuestas a preguntas complejas (Watson, Siri, Asana, etc)
4. Búsqueda
5. Reconocimiento de voz
6. Identificación de precios de productos, lugares, fechas, horas, nombres de personas o empresas

Desafíos

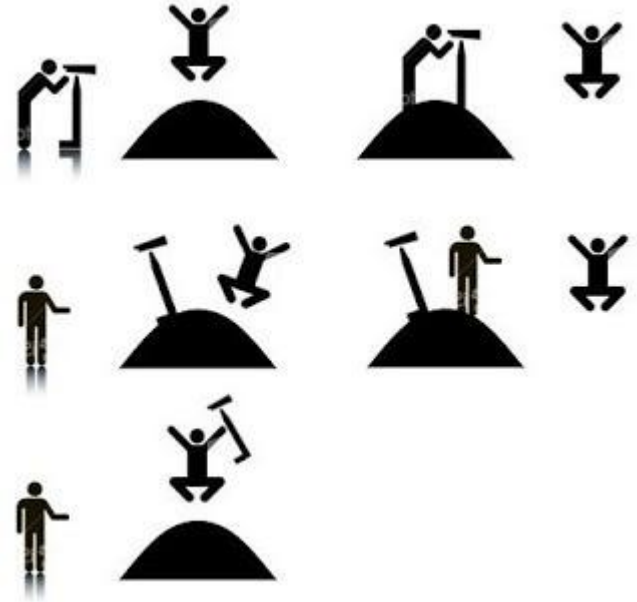
- Problema especial
- Vocabularios enormes (sparsity problem)
- El orden importa
- Complejidad para representar, aprender y razonar

Language ⇅	Dictionary ⇅	Approximate number of words ⇅
Korean	우리말샘 (Woori Mal Saem, 2017)	1,100,373
Turkish	Büyük Türkçe Sözlük	616,767
Swedish	Svenska Akademiens ordbok, Swedish Academy	~600,000
Icelandic	Orðabók Háskólans	~560,000
Japanese	Nihon Kokugo Daijiten	500,000 and more
Lithuanian	Lietuvių kalbos žodynas (Academic Dictionary of Lithuanian)	500,000 and more
Norwegian	Norsk Ordbok	ca. 500,000
Dutch	Woordenboek der Nederlandsche Taal	ca. 400,000
German	Deutsches Wörterbuch	330,000

English	Oxford English Dictionary, Second Edition	171,476
English	Merriam-Webster, Third Edition and Addenda Section	470,000
Spanish	Diccionario de la lengua española de la Real Academia Española, 23th edition, 2014	93,000

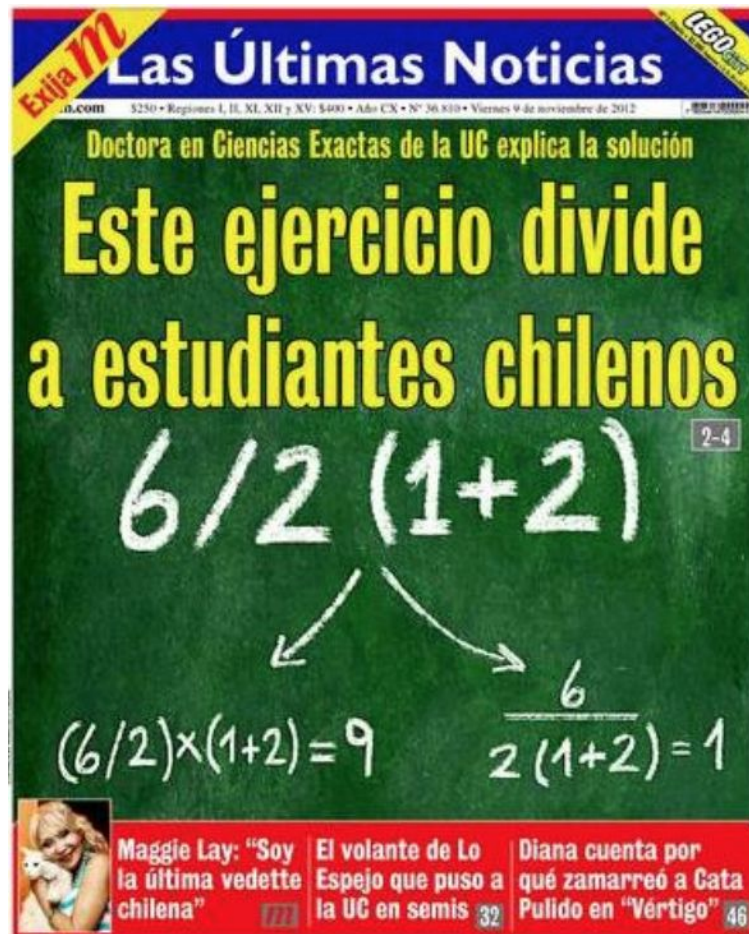
Desafíos

“Observé a un hombre en una colina con un telescopio”

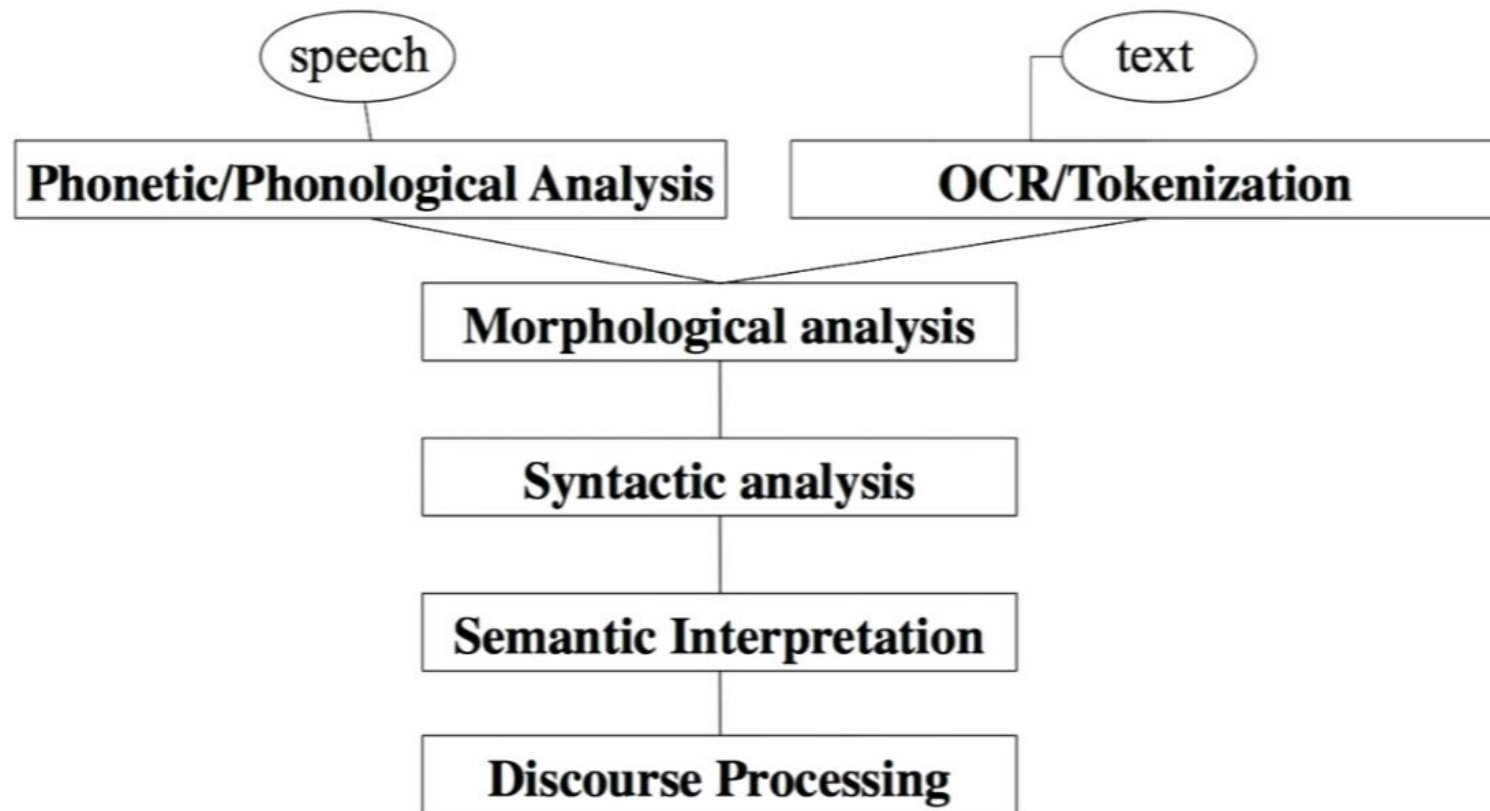


Desafíos

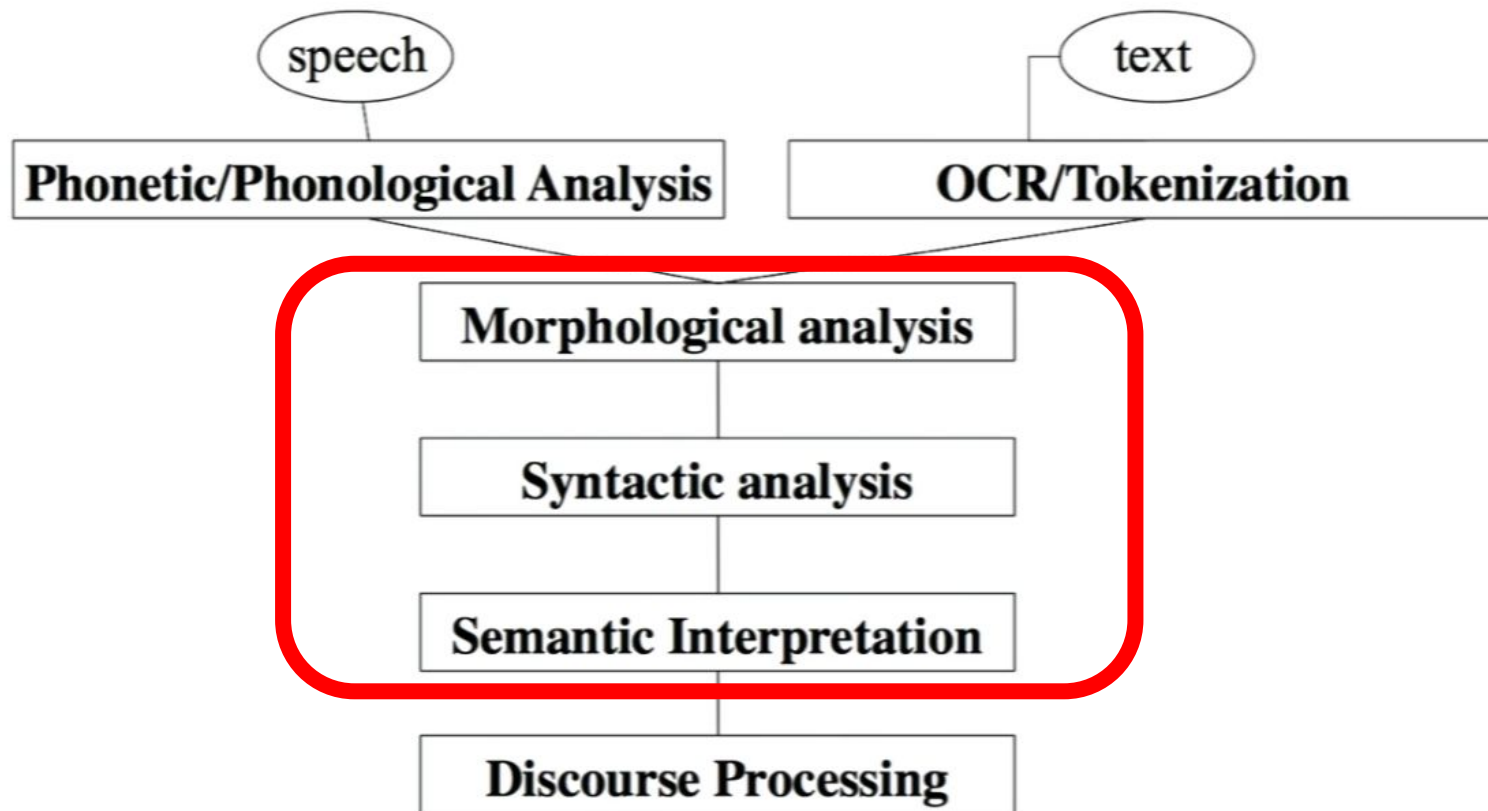
- Ambigüedad, a diferencia de lenguajes formales, lenguajes de programación.
- Sentido común, contextual (Watson example), basado en conocimiento previo.
- Se omite información con tal de que sea un mecanismo eficiente de comunicación



Tipos de Análisis



Tipos de Análisis



Preguntas