

Presentación del curso “Lingüística Computacional”

Igor Boguslavsky
iboguslavsky@fi.upm.es
Bloque 2 D2201

¿Qué es la lingüística?

- Una ciencia que estudia distintas lenguas y el lenguaje natural en general
- La importancia de la lingüística para la ingeniería
 - los sistemas de procesamiento de lenguaje natural (PLN) desempeñan un papel cada vez más importante
 - estos sistemas están basados en los métodos específicos de Lingüística Computacional
 - para desarrollar un sistema de PLN moderno es indispensable tener los conocimientos básicos sobre como funciona el lenguaje natural
- Ilustración: construcción del robot hablante.
 - tiene que entender lo que le dicen (modelo de comprensión)
 - tiene que expresar correctamente en lenguaje natural la información que quiere transmitir (modelo de habla)

HAL 9000

- Película de Stanley Kubrick “2001: La Odisea en Espacio”
- HAL es un agente artificial capaz de comunicarse con la gente en lenguaje natural (inglés).
 - *Dave: Open the pod bay doors, HAL.*
 - *HAL: I am sorry, Dave. I am afraid I can't do that.*
- HAL tiene que analizar la señal audio entrante y recuperar la secuencia de palabras. Para responder: la operación inversa
=> **FONÉTICA y FONOLOGÍA**

¿Qué más?

- Reconocer distintas formas de palabras y comprender la diferencia entre ellas (*door – doors, am – is – are – be*). Hacer/reconocer contracciones (*I am → I'm*. En español: *del -> de el, al -> a el*)

=> **MORFOLOGÍA**

- Reconocer la estructura de lo que se dice y poder ordenar las palabras de su respuesta correctamente
- *I am afraid I can't do that - am can afraid I not that do* (una cadena correcta, otra no)
- *Planes de estudio – Estudio de planes* (ambas frases son correctas pero los significados son distintos)

=> **SINTAXIS**

¿Qué más?

- Reconocer el significado de lo que se dice y elegir las palabras que expresen el significado que se quiere transmitir
=> ***SEMÁNTICA***

No es todo

Dave: What's the problem?

HAL: I think you know what the problem is just as well as I do.

Dave: I don't know what you're talking about.

HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

- Cortesía
- Conciencia de si mismo: un modelo de si mismo, incluyendo planes y objetivos
- Modelo del interlocutor, razonar sobre sus planes y objetivos

=> **PRAGMÁTICA**

HAL: I can tell from the tone of your voice, Dave, that you're upset. Why don't you take a stress pill and get some rest.

[Dave acaba de hacer un dibujo de Dr. Hunter].

HAL: Can you hold it a bit closer?

[Dave lo hace]

HAL: That's Dr. Hunter, isn't it?

Dave: *Yes.*

- Reconocimiento de las emociones por el habla
- Capacidad visual. Reconocimiento de caras.

Rama de la Lingüística	Qué estudia
fonética	la naturaleza acústica y fisiológica de los sonidos del habla
fonología	el modo en el que los sonidos funcionan en la lengua
morfología	palabras, sus partes y sus rasgos
sintaxis	cómo se combinan las palabras, cómo se ordenan en la oración y qué relaciones existen entre ellas
semántica	el significado de palabras y oraciones enteras
pragmática	cómo el contexto influye en la interpretación del texto

Un robot como HAL es todavía ciencia ficción

- Pero algunos pasos importantes en esta dirección se han hecho y se están haciendo.
- Existen modelos formales de muchos aspectos de LN.
- Estos modelos están incorporados en muchas aplicaciones (recuperación de información, traducción automática, respuestas a las preguntas, etc.)
- En la parte de IA también hay muchos avances.

¿Qué capacidades tiene que tener un agente artificial?

- Comprender, explicar y aprender.
- Comprender
 - no sólo el significado directo del texto, sino el significado que sea apropiado al contexto
 - reconstruir significados implícitos
 - interpretar actos de habla indirectos
- Explicar
 - sus acciones, conclusiones y decisiones a los humanos en términos comprensibles. Tenemos que confiar a sistemas inteligentes.

¿Qué capacidades tiene que tener un agente artificial?

- Aprender
 - aprender constantemente como los humanos
 - aprender nuevas palabras, nuevos elementos del modelo del mundo, nuevos metodos de resolver problemas
- Para alcanzar este nivel, el agente tiene que tener modelos adecuados del lenguaje natural y del mundo.
- Sobre eso vamos a hablar en este curso.

Cosas prácticas

- Varias prácticas y ejercicios en grupos de 2.
- La mayor parte de las prácticas están relacionadas con la construcción del modelo computacional del castellano.
 - Vamos a contribuir al sistema de traducción automática del español al inglés
- La nota final se deriva de las notas puestas por estos trabajos.
- El comportamiento activo en clase puede mejorar la nota (y habitualmente mejora).