

# INGENIERIA Y LINGÜÍSTICA: GESTIÓN DEL CONOCIMIENTO

Ponente: Prof. Jesús Cardeñosa  
Univ. Politécnica de Madrid

XIII Jornada de la Asociación Española de Terminología

Alcalá de Henares  
8 de noviembre del 2013

# LOS TÓPICOS...

## INGENIERIA

- Modelos matemáticos
- Tablas, reglas, procedimientos
- ..

**Pensamiento generativo (resolver problemas)**

## LINGÜÍSTICA

- Teorías
- Modelos
- Gramáticas
- ..

**Pensamiento analítico (¿qué problemas?)**

*“una máquina nunca podrá hacer lo que hace un ser humano”*

# .....¿Seguro?

- Se supone que una máquina no tiene inteligencia, pero a veces....lo parece...

## A ESO SE DEDICA LA INTELIGENCIA ARTIFICIAL

Y dentro de ella, asociados a la lengua hay varios “items”:

- Lingüística computacional
- Procesamiento de lenguaje natural
- Ingeniería lingüística
- Reconocimiento y generación de voz
- Y un sinfín de sub-items...

Entonces ¿qué puede hacer realmente una máquina incluso mejor que un ser humano?

# ¿Qué puede hacer una máquina?

**Cualquier tarea sistemática y repetitiva.**

¿Hace el ser humano tareas repetitivas y sistemáticas?

Muchas.....

¿y en lingüística?

También y hay áreas donde es más evidente y una de ellas es la terminología.

Luego veremos algo...

# ¿Confrontaciones?

Lo cierto es que

esas “confrontaciones” entre  
Ingeniería y Lingüística  
son artificiales,  
porque en un futuro no muy lejano

**la Ingeniería Lingüística**

**será una de las industrias más grandes del planeta.**

Hoy por hoy es una “industria oculta” porque ya existe, pero dispersa...

# LA INDUSTRIA OCULTA

- Tiene pendiente su propia **revolución industrial**.
- Del trabajo “artesano (fichas, subrayar textos, mucha lectura, etc) al **trabajo sistemático**, masivo..
- Las necesidades son inmensas, casi inimaginables.
- No hay profesionales “mixtos” en el mercado porque involucran estudios muy “lejanos” como Ingenieros informáticos o de telecomunicación y filólogos, lingüistas..

# Un poco de historia...

## A principios de los 80.....

- El optimismo sobre la solución de todo lo imaginable en PLN y en traducción automática era enorme.
- 30 años más tarde los avances en reconocimiento y síntesis de voz han sido grandes.
- No así en comprender (por parte de una máquina) lo que se dice, ni generar contenidos no preexistentes (a partir de datos por ejemplo)
- Hay quien dice que nunca se podrá.....
- Pero decir eso en el mundo de la ciencia es trazar un objetivo...*(es decir, se acabará pudiendo...)*

# ¿es posible representar el conocimiento?

- Entender (por una máquina) lo que se dice supone tratar de emular el proceso del ser humano.
- El lenguaje es la expresión del conocimiento
- Los llamados “modelos de representación del conocimiento” ¿pretenden representar conocimiento o lenguaje?
- La mayor parte de estos modelos típicos de la IA (redes semánticas, grafos conceptuales, etc..) vienen de otros campos (psicología, lingüística..)
- Desde los años 80 se ha avanzado poco...
- Donde sí se ha avanzado es la parte tecnológica (capacidades enormes de procesamiento y almacenamiento de datos)



# Hay una distorsión..

- Grandes capacidades de procesamiento parecen haber creado grandes novedades. Un ejemplo, la TA.
- Los modelos de tratamiento computacional de la lengua se basaron en reglas (siempre insuficientes) y léxicos muy ricos (y muy costosos y nunca completos)
- Luego (sobretudo en TA) se pasó a las memorias de traducción, pero había que ir haciéndola, (muy costosas)
- Y ahora la “moda” es la SMT (Statistical Machine Translation) basada en *estadística*..(expresión curiosa usada como sinónimo de potencia y calidad)

# Pero....¿y los resultados?

- ¿Se diferencian drásticamente de los de hace 30 años? No...
- ¿Hay nuevos modelos que supongan un avance revolucionario en el tratamiento de la lengua? No...
- Los buscadores de Internet, ¿buscan mucho mejor que hace 15 años por ejemplo? No....( eso sí, ofrecen en 0,24 segundos 7.000.000 de documentos que por supuesto nunca verá)
- ¿Y las herramientas de trabajo de terminólogos, lingüistas de corpus, traductores, lexicógrafos?
- Déjenme que les cuente una anécdota...

# Una anécdota

- Recientemente mi colega lingüista me comentó que una aplicación informática que hemos desarrollado “solo traga” un corpus de 10 millones de palabras.
- Que su grupo para hacer una cosa muy específica (lista de palabras) tiene una que se “traga” 40 millones, y eso lo asocia a “mejor” y “más representativo”.
- Es como todo “burro grande ande o no ande”. La mejoría de representatividad de 40 millones sobre 10 M para lo que hacía era menor de una millonésima.
- Tenemos un artículo empezado sobre “calidad y lingüística” que algún día terminaremos (espero)

# Un paso más allá: Gestión del conocimiento

- Concepto de principios de los 90 para recopilar y gestionar el conocimiento corporativo de las organizaciones.
- Asociada a Bases de Datos, ontologías, buscadores, procesos, etc...
- Solo en ocasiones asociado a la lengua.
- Poco a poco fueron surgiendo conceptos como “descubrimiento del conocimiento” “construcción automática de ontologías”, “minería de textos”,..
- Todos ellos son imposibles sin “tratamiento de la lengua”.

# ¿Cómo se prevé que sea la G.C. futura?

- Lo anterior no ha conseguido una implantación sistemática...falta mucho por hacer y todo pasa por modelos de tratamiento de la lengua, representación de contenidos, multilingüidad...
- Imaginemos.....
- Imaginen un “lugar” donde está el conocimiento de la Humanidad, que crece continuamente con cada cosa nueva, dato nuevo, al que se le pueda preguntar cualquier cosa, desde donde nació Picasso a en qué valor cerró la bolsa de N.York en 1977. Y que le contesta en su idioma venga de donde venga dicha información.
- Ese “lugar” ¿cómo se alimenta de información? ¿Quién y de qué manera le pregunta cosas?

# “Ese lugar...”

- La información está en formatos (texto, imágenes, audio..)
- Pero ¿cómo se sabe donde hay información?
  - *Juan le dijo a María que quería ir al cine*
  - *Juan María que le a dijo cine al ir quería*
  - *Juan Juan Juan Juan*
- En la 1ª hay información, en la 2ªpodría haberla pero “le falta algo” y en la 3ª no hay información apenas.
- Podemos saber que:
  - La 1ª cumple las reglas de una lengua
  - La 2ª no las cumple
  - La 3ª ni cumple ni incumple

# ¿Y las lenguas?

- Cambian las reglas, pero lo que se dice es lo mismo..
- ¿Hay alguna manera de representar el concepto de lo que Juan le dijo a María sin depender de las reglas de una u otra lengua?
- A esa “manera” se la llama “interlingua”
- Se considera hoy por hoy la única forma de atacar el problema del multilingüismo. (contenidos accesibles suponen contenidos que hay que representar)

# Interlinguas

- A “esa manera” de representar lo que se dice de manera independiente de la lengua se le llama “interlingua”
- Se considera la única manera de solventar el problema del multilingüismo masivo
- Es una aproximación de origen japonés que en su día se presentó como una aproximación para la TA “contra” los sistemas “transfer” de pares de lenguas (Europa)
- La zona del mundo desarrollado con más problemas de ese tipo es Europa, que se negó durante décadas a adoptar la “solución asiática”, ni siquiera como alternativa.



# Universal Networking Language (UNL)

- La Universidad de las Naciones Unidas (con sede en Tokio) puso en marcha en 1996 una iniciativa para definir las especificaciones de una interlingua que fuera capaz de traspasar las barreras del multilingüismo en Internet. (causa de la brecha digital)
- Juntó a más de 100 investigadores de 14 países y cobertura idiomática de mas del 90% de la población mundial.
- El UNL es público y para asegurarlo, la Asamblea General de las Naciones Unidas patentaron el UNL (2001) y lo hicieron público. (para evitar su apropiación por nadie)
- Es la única interlingua pública de carácter general que existe.

## ¿y el soporte a la Gestión del Conocimiento?

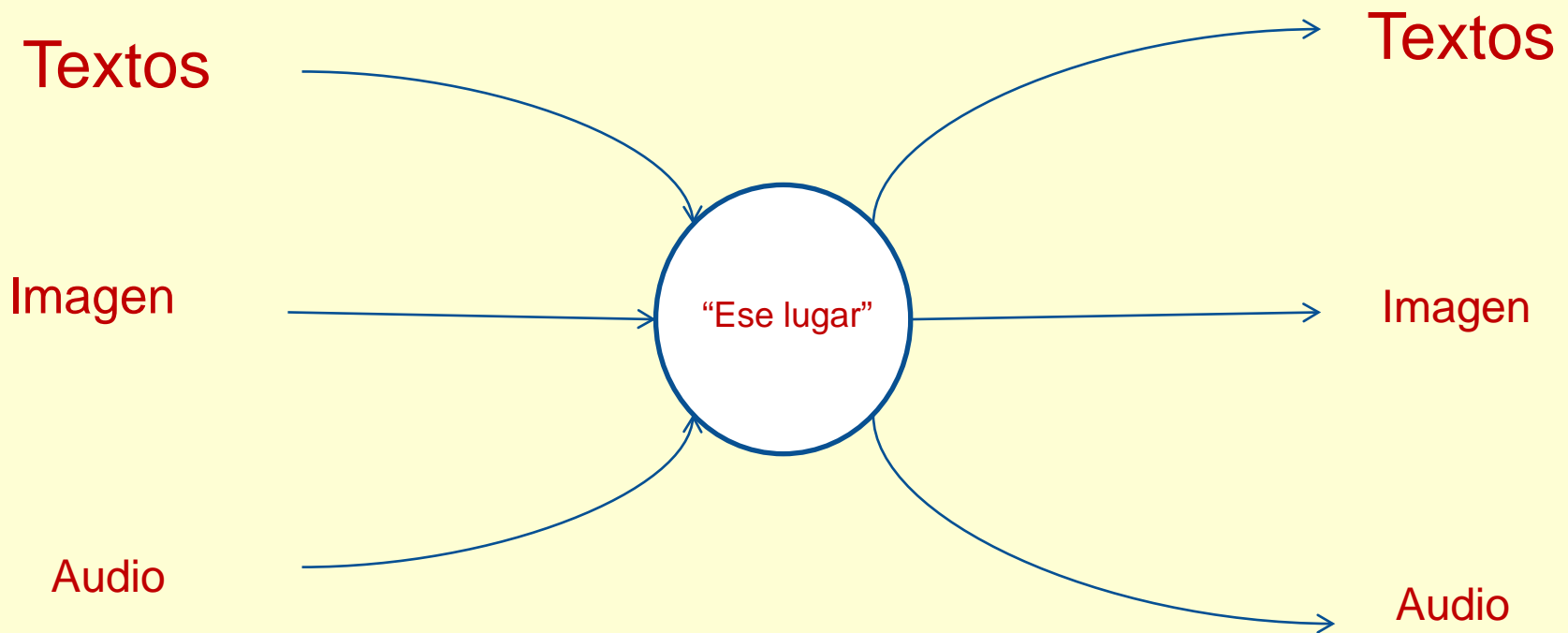
- Se asocia a ontologías, pero para ser capaz de generar contenidos le hace falta el componente lingüístico.
- Apenas hay esbozos hoy en día de sistemas que integren el conocimiento ontológico con el lingüístico.
- Una línea es “ontology learning” es decir, “capturar conocimiento ontológico a partir de textos”.
- Pero la línea opuesta:

Ontología → Lingüística → Texto

en cualquier lengua, no está apenas trabajada.

# Es decir...

En cualquier lengua



Esto...no existe hoy...pero ya hay gente trabajando

## ...Y en todo esto....

- Hay grupos que dicen que para procesar la lengua no hace falta la intervención de lingüistas.
- En nuestra experiencia como ingenieros se comenzó sin lingüistas, pero a medida que íbamos avanzando tuvimos que necesitar la progresiva incorporación de personas de perfil lingüístico
- Para nuestro equipo ([www.vai.dia.fi.upm.es](http://www.vai.dia.fi.upm.es)) fue muy enriquecedor y sinérgico, sobretodo en la terminología de la interlingua UNL. (nos hicimos falta)

# Para construir “ese lugar” .....

- Necesitamos ingenieros y lingüistas.
- Promovemos convenios con grupos de filología.
- Estamos definiendo los contenidos de un profesional de perfil mixto mediante el que será un Máster Internacional de Ingeniería lingüística en el contexto de una cátedra UNESCO.
- Hay demanda en el mercado de ese profesional ...y salvo excepciones individuales no lo hay.
- Termino diciendo:

*“Nos necesitamos. Trabajemos juntos.*

*La tarea es inmensa...”*

Y ahora sí termino...

Déjenme que les enseñe algunos  
juguetes de “ingenieros”.....😊

¡Ah! ¡¡¡Gracias!!!