

# Modelos Computacionales

## Actividad 1

**José Alberto Benítez Andrades**

**71454586A**

**Modelos Computacionales**

**Máster en Lenguajes y Sistemas Informáticos - Tecnologías del Lenguaje en la Web**

**UNED**

**8 de enero de 2011**

---

8 de enero de 2011

## 0. Enunciado

Contenidos:

o El objetivo global del Tema 1 es presentar al alumno aquellos conceptos y conocimientos preliminares sin los que no podría ubicar los contenidos de la asignatura. Se pretende, además, justificar el interés de la asignatura, motivar al alumno en su estudio y presentar las posibles aplicaciones de los contenidos.

o Bibliografía básica: (Jurafski&Martin 2009, cap 1)

Visitar los dos enlaces siguientes, realizar la actividad y responder a la pregunta:

[http://es.wikipedia.org/wiki/Ling%C3%BC%C3%ADstica\\_computacional](http://es.wikipedia.org/wiki/Ling%C3%BC%C3%ADstica_computacional)

[http://es.wikipedia.org/wiki/Procesamiento\\_de\\_lenguaje\\_natural](http://es.wikipedia.org/wiki/Procesamiento_de_lenguaje_natural)

Actividad 1: Escribir un resumen (con vuestras palabras) de ambos conceptos, Ingeniería Lingüística y Procesamiento Automático del Lenguaje Natural.

Pregunta 1: ¿Qué diferencias encuentra entre el concepto de Lingüística computacional y procesamiento de lenguaje natural?

## 1. Resumen de Ingeniería Lingüística y Procesamiento Automático del Lenguaje Natural.

### Ingeniería Lingüística

En primer lugar, antes de hablar sobre la **Ingeniería Lingüística**, voy a definir brevemente lo que significa **lingüística computacional**. La lingüística computacional es un campo interdisciplinario que se ubica entre la lingüística y la informática y que tiene como fin la elaboración de modelos computacionales que reproduzcan distintos aspectos del lenguaje humano. Tiene como particularidad la no centralización en un área en particular, necesitando así la participación de personas que dedican sus estudios a distintas áreas bastante diversas, como por ejemplo: lingüistas, informáticos especializados en la IA (Inteligencia Artificial), psicólogos cognoscitivos y expertos en lógica.

La LC se originó en los 1950 en Estados Unidos, y en la década de los sesenta comenzó a tratar con cierto nivel de comprensión humana y **producción de los lenguajes naturales**.

La **Ingeniería Lingüística**, o también llamada **Lingüística Computacional Aplicada**, es la rama de la Lingüística Computacional que posee una orientación tecnológica, y está hecha para poder realizar simulaciones de la conducta lingüística con medios informáticos. El fin de este tipo de ingeniería, es crear productos informáticos en los que el lenguaje oral o escrito jueguen un papel fundamental. Algunos ejemplos de componentes de este tipo son: las búsquedas automáticas en textos, tanto de forma como de significado (recuperación de información y motores de búsqueda), los traductores los programas de reconocimiento y

---

8 de enero de 2011

síntesis de voz, todo lo relacionado con la minería de datos y extracción de información, como por ejemplo la búsqueda de información en grande de bibliografías o la contestación directa de preguntas sobre la base de grandes bancos de datos, la indexación de literatura, la creación de índices y listas de materias, la producción de resúmenes, también son ayudas relacionadas con la ingeniería lingüística, etcétera.

En general la lingüística computacional tiene 3 problemas bastante claros: la determinación de la semántica correcta de las palabras dependiendo del contexto en el que se encuentren, la existencia de ambigüedad sintáctica y el reconocimiento de la pragmática de una frase.

Los ordenadores son máquinas, y para entender el lenguaje se necesita intelecto, se pretende buscar mediante la IA, que los ordenadores en un futuro puedan llegar a pensar, y la **Ingeniería Lingüística** se encuentra en el camino hacia esa meta.

## Procesamiento Automático del Lenguaje Natural

El **Procesamiento de Lenguaje Natural (PLN)** es una rama de la IA y de la LC (de la cual hemos hablado en los párrafos anteriores). El PLN se encarga de encontrar los caminos posibles que existen en la comunicación entre personas, o entre personas y ordenadores mediante lenguajes naturales. El PLN busca realizar simulaciones o ejecuciones de comunicación mediante programas.

En relación a su origen, todo comenzó en los años 1940 y 1950 con los primeros intentos de traducción fallidos, y más adelante en la década de los sesenta y setenta comenzaron a funcionar los avances que se hicieron produciendo interfaces en lenguaje natural para bases de datos y algunas aplicaciones.

Los problemas que surgen para el procesamiento de lenguaje natural, son los mismos que para la **Ingeniería Lingüística**: la semántica, la ambigüedad sintáctica y la pragmática. Además de estos 3 problemas, se suman la detección de separación entre palabras en el lenguaje hablado y la recepción imperfecta de datos (relacionado con los acentos de cada región, etcétera).

El **PLN** tiene 6 componentes:

- Análisis morfológico, sintáctico, semántico y pragmático.
- Planificación de la frase y generación de la frase.

Además, el **PLN** se aplica a: tareas de síntesis del discurso, análisis del lenguaje, comprensión del lenguaje, reconocimiento del habla, síntesis de voz, generación de lenguajes naturales, traducción automática, respuestas a preguntas, recuperación de la información y extracción de la información.

## 2. ¿Qué diferencias encuentra entre el concepto de Lingüística computacional y procesamiento de lenguaje natural?

Después de leer las 2 páginas de la wikipedia sugeridas en el enunciado, y un número bastante amplio de webs en español y en inglés, mis conclusiones con respecto a las diferencias entre el concepto de lingüística computacional y procesamiento de lenguaje natural son las siguientes:

En la actualidad LC y PLN se tienden a identificar, por lo que ambos términos se pueden considerar sinónimos:

“Computational linguistics: [...] 2. (more usually today) A synonym for natural-language processing” (Trask 1993:53).

Igual que la LC, el PLN considera los ordenadores como un instrumento adecuado para la descripción y explicación de las diferentes facetas o niveles del lenguaje: fonética y fonología, morfología, sintaxis, semántica, pragmática, análisis del discurso, etc. El objetivo general, también común con la LC, es diseñar programas o sistemas informáticos que simulen la conducta lingüística humana en todas o en alguna de sus facetas, programas que sean capaces de utilizar lenguajes naturales:

“[...] Lingüística Computacional y Procesamiento del Lenguaje Natural tratan de lo mismo: del desarrollo de programas de ordenador que simulan la capacidad lingüística humana” (Moreno Sandoval 1998:14).

O en palabras de J. Allen (1995:1): *“The goal of this research is to create computational models of language in enough detail that you could write computer programs to perform various tasks involving natural language. The ultimate goal is to be able to specify models that approach human performance in the linguistic tasks of reading, writing, hearing, and speaking.”*

Así pues, igual que se desprendía de las definiciones de LC, la creación de modelos del lenguaje y el hincapié en la comunicación hombre-máquina también parecen ser los pilares centrales del PLN, orientación que aparece estrechamente vinculada con otra rama de la Informática, la inteligencia Artificial (IA), que se encarga de “codificar en un programa informático facultades cognitivas” (Moreno Sandoval 1998:14), entre ellas el lenguaje.

La doble faceta teórica y aplicada del PLN también se puede rastrear en otro clásico, J. Allen (1995:1-3), quien distingue en el PLN motivaciones científicas y motivaciones aplicadas, de modo paralelo a como R. Grishman (1991 [1986]) había hecho a propósito de la LC. Para J. Allen (1995:1), el objetivo del PLN es “crear modelos computacionales del lenguaje lo suficientemente detallados que permitan escribir programas informáticos que realicen las diferentes tareas donde interviene el lenguaje natural”. Dichos modelos los considera útiles “both for scientific purposes –for exploring the nature of linguistic communication– and for practical purposes –for enabling effective human-machine communication” (ibid.).

8 de enero de 2011

No obstante, hay que destacar que para algunos autores el PLN se centra específicamente en los objetivos aplicados, en los que confluye con la vertiente aplicada de la LC o LC Aplicada

Las diferencias entre LC y PLN se establecen considerando argumentos como los siguientes

a) **Ámbito de procedencia**

- La LC se sitúa en la esfera de la Lingüística y de la Ciencia Cognitiva.
- El PLN está más vinculado a la Informática y a la Inteligencia Artificial,
- El Procesamiento del Lenguaje Natural (PLN) es una parte esencial de la Inteligencia Artificial que investiga y formula mecanismos computacionalmente efectivos que faciliten la interrelación hombre/máquina y permitan una comunicación mucho más fluida y menos rígida que los lenguajes formales.

b) **Motivación**

- En la LC predominan los aspectos teóricos o científicos.
- En el PLN, los aplicados o tecnológicos.

A continuación muestro una tabla relacional resumida de la LC y el PLN:

<b>Lingüística Computacional</b>		<b>Procesamiento de Lenguaje Natural</b>	
Término más usual desde la perspectiva de la Lingüística		Término más común desde la perspectiva de la Informática	
LC Teórica	LC Aplicada	PLN Teórico	PLN Aplicado
Predominio de aspectos teórico		Predominio de aspectos aplicado	