

# Introducción al PLN

## Procesamiento del Lenguaje Natural

José María Gómez Hidalgo

<http://www.esp.uem.es/~jmgomez/>

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Índice

- Definición y aplicaciones
- Historia e hitos
- Ingeniería y evaluación
- Análisis superficial y comprensión
  - Comprensión
  - Análisis estadístico
- El problema de la ambigüedad

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Introducción al PLN

## Definición y aplicaciones

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Definición y aplicaciones

- El lenguaje es un aspecto humano fundamental
- Relacionado con la inteligencia
- Medio de
  - Comunicación
  - Almacenamiento del conocimiento humano

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Definición y aplicaciones

- PLN = tratamiento computacional de los lenguajes naturales (vs. formales) para
  - Desarrollar **programas** que realicen tareas relacionadas con el lenguaje
  - Desarrollar **modelos** que ayuden a comprender los mecanismos humanos relacionados con el lenguaje

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Definición y aplicaciones

- Lingüística computacional
  - Especificar modelos que aproximen la capacidad humana en tareas lingüísticas como *leer, escribir, escuchar y hablar*
- Ingeniería del lenguaje (Unión Europea)
  - Desarrollar sistemas plenamente funcionales y utilizables que realicen tareas relacionadas con el lenguaje

**i + D**

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Definición y aplicaciones

- **Carácter multidisciplinar**
  - Lingüística y psicolingüística
    - Lexicografía y terminología
  - Lenguajes formales y compiladores
  - **Inteligencia artificial**
    - Representación del conocimiento
  - Lógica
  - Ingeniería del software

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Definición y aplicaciones

- Traducción automática
- Recuperación de información
- Interfaces a bases de datos, y sistemas de respuesta a preguntas
- Extracción de información
- Extracción de resúmenes
- Resolución cooperativa de problemas
- Tutores inteligentes

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Introducción al PLN

## Historia e hitos

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Historia e hitos

- **Primeros tiempos (1950-65)**
  - Traducción automática => ALPAC
  - Gramática transformativa (Chomsky)
  - Recuperación de información
- **Ampliación de perspectivas (1960-70)**
  - Sistemas de respuesta a preguntas (BASEBALL)
  - Resolución de problemas (STUDENT)
  - Consulta (ELIZA)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Historia e hitos

- **Desarrollo de formalismos (1965-70)**
  - Augmented Transition Network, ATN (1960)
  - Gramática de casos (1975)
  - Dependencia conceptual (1972)
  - Semántica procedimental (1968)
  - Red semántica (1968)
- **Hitos de los 70**
  - SHRDLU (1972)
  - LUNAR (1972)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Historia e hitos

- **Expansión de dominios de aplicación (1970-85)**
  - Interfaces a BDs (LADDER/LIFER, 1977)
  - Aprendizaje asistido por computadora (SCHOLAR, 1970)
  - Automatización de oficinas (SCHED, 1978)
  - Programación automática (NLPQ, 1976)
  - Procesamiento de textos científicos

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Historia e hitos

- **Tendencias actuales (1985-)**
  - Implementaciones independientes del dominio
  - Resurgimiento de la traducción automática
    - METEO (1977)
  - Comercialización del PLN
    - ROBOT/INTELLECT (1977)
    - INQUERY (1990), Oracle ConText Cartridge (1997)
    - BABYLON (1997)
    - VIAVOICE (1998)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Historia e hitos

- **Tendencias actuales (1985-)**
  - Proliferación de formalismos (1970-)
    - Sintaxis (gramáticas basadas en restricciones)
    - Semántica (gramática de Montague)
  - Intensificación del empiricismo
    - Aplicación del aprendizaje automático
  - Nuevos dominios de aplicación
    - Internet/Web => traducción automática, recuperación de información
  - Bibliotecas digitales

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Introducción al PLN

## Ingeniería y evaluación

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Ingeniería y evaluación

- Técnicas genéricas de Ingeniería del Software
- Reutilización de herramientas y módulos
  - Lenguajes declarativos (Prolog, DCG)
  - Recursos léxicos - WordNet, EuroWordNet
  - Oracle ConText Cartridge
  - Smart
- Nosotros usaremos
  - Java, Prolog Cafe, Weka

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Ingeniería y evaluación

- Evaluación
  - Objetivo => determinar la calidad de un sistema de PLN
  - Mala fama desde ALPAC (1966)
  - Evaluación desde la Ingeniería del Software
    - International Standard ISO/IEC 9126
    - Caja negra vs. caja de cristal

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Ingeniería y evaluación

- EAGLES (1992)
  - Expert Advisory Groups for Language Engineering Standards - Evaluation Working Group
  - Evaluación de progreso, **adecuación** y diagnóstico
  - Enfoque “informe del consumidor”
  - Definición de una metodología de evaluación

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Ingeniería y evaluación

- **EAGLES (1992)**
  - Características de calidad (ISO 9126)

Características	Subcaracterísticas
Funcionalidad	Adecuación, precisión, interoperabilidad, conformidad, seguridad
Fiabilidad	Madurez, tolerancia a fallos, recuperabilidad
Facilidad de uso	Comprensibilidad, facilidad de aprendizaje, facilidad de operación
Eficiencia	Comportamiento respecto al tiempo y memoria
Facilidad de mant.	Analizabilidad, capacidad de modificación y prueba, estabilidad
Transportabilidad	Adaptabilidad, facilidad de instalación y mantenimiento, capacidad de ajuste

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Ingeniería y evaluación

- **ARPA/DARPA (1985-)**
  - (Defense) Advanced Research Projects Agency
  - Aportan recursos
  - Dependientes de la tarea
    - Traducción automática
    - Acceso BDs con habla (ATIS)
    - Recuperación de textos (TREC)
    - Extracción de información (MUC)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Ingeniería y evaluación

- **Recursos para la evaluación**
  - **Córpora (Brown Corpus)**
    - Colecciones de textos de origen natural en formato electrónico
  - **Conjuntos de casos de prueba**
    - Conjunto de entradas construidas artificialmente para probar un sistema respecto a un fenómeno
  - **Colecciones de evaluación (Ohsumed)**
    - Conjunto de entradas en lenguaje natural y las correspondientes respuestas deseables

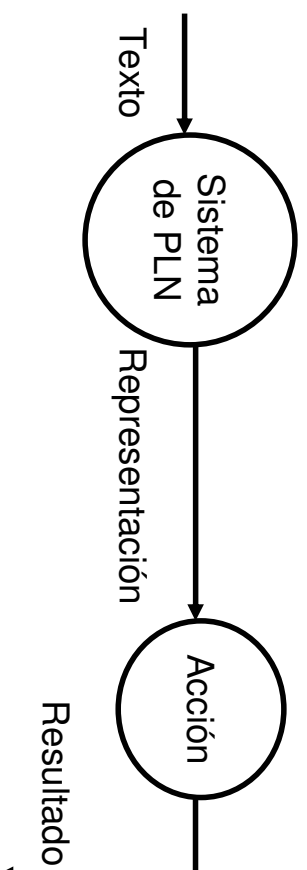
Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Introducción al PLN

### Análisis superficial y comprensión

## Análisis superficial y comprensión

- PLN = traducción a lenguaje interno de representación



Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis superficial y comprensión

- Complejidad de la representación
  - Alta (ej. lógica de primer orden) => comprensión
  - Baja (ej. Listas de términos) => análisis superficial
- Corresponde a dos corrientes históricas
  - Corriente formalista = comprensión = PLN basado en conocimiento
  - Corriente empiricista = análisis superficial = PLN estadístico
- Actualidad => tendencia a la integración

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis superficial y comprensión

- Características de los sistemas

	Comprensión	Análisis superficial
Dominio	<i>Dependientes</i>	<i>Independientes</i>
Coste de desarrollo	<i>Alto</i>	<i>Bajo</i>
Efectividad	<i>Alta</i>	<i>Media</i>

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Introducción al PLN

### Análisis superficial y comprensión

#### Comprensión

# Comprensión

## Niveles del lenguaje

- **Nivel fonológico**
  - Trata de cómo las palabras se relacionan con los sonidos que las representan
- **Nivel morfológico**
  - Trata de cómo las palabras se construyen a partir de unas unidades de significado más pequeñas llamadas morfemas
  - Por ejemplo: rápida + mente => de manera rápida

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Comprensión

## Niveles del lenguaje

- **Nivel sintáctico**
  - Trata de cómo las palabras pueden unirse para formar oraciones, fijando el papel estructural que cada palabra juega en la oración y que sintagmas son parte de otros sintagmas
- **Nivel semántico**
  - Trata del significado de las palabras y de cómo los significados se unen para dar significado a una oración
  - Se refiere al significado independiente del contexto, es decir, de la oración aislada

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Comprensión

## Niveles del lenguaje

- Nivel pragmático
  - Trata de cómo las oraciones se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones
  - Se suele reconocer un subnivel: discursivo
    - Trata de cómo el significado de una oración se ve afectado por las oraciones inmediatamente anteriores

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Comprensión

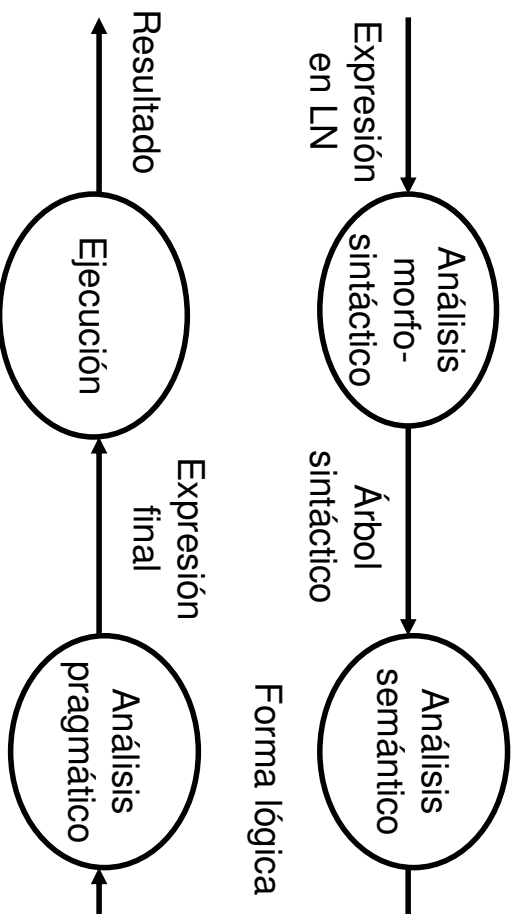
## Arquitectura de un sistema de PLN

- Análisis = traducción de un lenguaje a otro
- Cada nivel se corresponde con un tipo de conocimiento y por tanto un lenguaje
- Separación en fases

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Comprensión

## Arquitectura de un sistema de PLN



Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Comprensión

## Arquitectura de un sistema de PLN

- Separación en fases favorece la **modularidad**
- Integración de fases favorece la **eficiencia**

## Comprensión

### Representación del conocimiento

- Se han propuesto numerosos formalismos de **representación (lenguajes formales)**
  - Gramáticas de estructura de frase (syntaxis)
  - Gramáticas basadas en restricciones (syntaxis, semántica)
  - Lógica de primer orden y otras (semántica)
  - Sistemas de marcos - frames (semántica, pragmática)
  - Dependencias conceptuales (semántica)
  - Scripts, planes y objetivos (pragmática)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Comprensión

### Representación del conocimiento

- Tipos de conocimiento
  - Conocimiento lingüístico
  - Conocimiento del mundo
    - Situación del discurso o de la historia
    - Conocimiento de la aplicación

# Comprensión

## Representación del conocimiento

- El papel del Prolog como **marco unificador**
  - Capacidad para representar conocimiento morfo-sintáctico (Gramáticas de Cláusulas Definidas)
  - Capacidad para representar conocimiento semántico y del mundo (Lógica de primer orden)
  - Operatividad (como lenguaje de programación)
    - Simultáneamente formalismo y analizador
    - Ejecución

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Introducción al PLN

## Análisis superficial y comprensión

### Análisis estadístico

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Motivación

- Era de la información (electrónica)
  - Crecientes cantidades de información y otros recursos en formato electrónico
    - Empresas, agencias gubernamentales, individuos enfrentados a un problema de *sobrecarga de información*
    - WWW - 270 millones de páginas (mediados 1998)
    - De 1970 a 1980 se pusieron en circulación 2000000 de informes anuales (7000 diarios)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Motivación

- Carencia de métodos efectivos de procesamiento de información para sistemas prácticos
  - PLN basado en conocimiento efectivo sólo en dominios limitados
- Disponibilidad de recursos
  - Grandes cantidades de texto en formato electrónico => posibilidad de realizar estudios estadísticos efectivos del uso del lenguaje
  - Diccionarios y datos en formato electrónico
- Por todo ello, *creciente interés en técnicas estadísticas*

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## ¿Qué es?

- Enfoque simplista
  - Técnicas basadas fundamentalmente en contar palabras
- Enfoque general
  - Sistemas prácticos de efectividad razonable basados en técnicas simples (tareas de clasificación)
  - Descubrimiento de generalizaciones (teorías) sobre el uso del lenguaje a partir de datos contrastables empíricamente

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## ¿Qué es?

- Intersección de varios campos
  - Recuperación de información
  - Aprendizaje automático (*machine learning*)
  - Teoría de la probabilidad
  - Estadística
  - Teoría de códigos
  - Teoría de la información
  - etc.

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Tareas de clasificación

- Casi todas las tareas resueltas por técnicas estadísticas admiten una formulación como “clasificación de texto”
- Clasificación = agrupamiento de entidades
- Clasificación de texto = agrupamiento de expresiones en lenguaje natural

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Tareas de clasificación

- Básicamente dos tipos de tareas
  - Clasificación de documentos
    - Expresiones en LN = documentos / casi documentos
    - Grano grueso
  - Clasificación de términos
    - Expresiones en LN = palabras / expresiones muy cortas
    - Grano fino

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Tareas de clasificación

- Tareas de clasificación de documentos
  - Recuperación de documentos (IR - *information retrieval*)
  - Categorización de documentos (*text categorization*)
  - Agrupamiento de documentos (*clustering*)
  - Filtrado/enrutamiento de documentos (SDI - *selective dissemination of information*)
  - Alineamiento de texto
  - Segmentación de texto

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Tareas de clasificación

- Tareas de clasificación de términos
  - Categorización de términos
    - Etiquetado sintáctico (POS *tagging* - *part-of-speech tagging*)
    - Desambiguación (WSD - *word sense disambiguation*)
  - Agrupamiento de términos
    - Creación de *thesauri* - diccionarios de sinónimos/campos semánticos

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Recursos disponibles

- Tres tipos de recursos
  - Colecciones de texto
    - *Corpora*
    - Colecciones de evaluación
  - Recursos léxicos
    - Diccionarios electrónicos y *thesauri*
    - Bases de datos léxicas
  - Programas
    - Recuperación de información
    - Aprendizaje automático

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Recursos disponibles

- Corpora (sing. corpus)
  - Colecciones de texto, no necesariamente documentos
  - Puede tener las siguientes características
    - Representatividad (orientado al estudio de la lengua)
      - El corpus constituye una muestra representativa del (fragmento del) lenguaje que se trata de representar
      - Brown Corpus - muestra representativa del inglés americano moderno - 1M pal, elaborado 1960-70 a partir de reportajes periodísticos, ficción, texto científico y legal, etc.

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Recursos disponibles

- Anotación
  - El corpus poseen anotaciones (etiquetas) que proporcionan algún tipo de información
  - Penn Treebank - etiquetado con marcas sintácticas que indican el árbol de análisis de cada oración - extraído del Wall Street Journal
- Multilingüidad
  - El corpus puede tener textos en varios idiomas
  - Canadian Hansards - bilingüe inglés francés alineado (mismos textos en dos idiomas) - actas del parlamento canadiense

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Recursos disponibles

- Colecciones de evaluación
  - Conjunto de textos en lenguaje natural y las correspondientes respuestas deseables para la tarea elegida
  - *Recuperación de documentos*
    - Un conjunto de documentos + un conjunto de solicitudes de información (consultas) + un conjunto de juicios de relevancia (qué documentos deberían recuperarse para cada consulta)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Recursos disponibles

- Algunos ejemplos (colecciones clásicas)
  - CACM - resúmenes de artículos de Communications of the ACM - 3204 docs, 52 consultas
  - CRAN - aerodinámica e Ingeniería Aeronáutica - 1398 docs, 225 consultas
  - MEDLINE - Medicina - 1033 docs, 30 consultas
  - LISA, NPL, CRAN, Time, etc.
- Más ejemplos (recientes)
  - OHSUMED - Extraídas del canal MEDLINE - 348566 docs, 106 consultas

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Análisis estadístico

## Recursos disponibles

- Categorización de documentos
  - Un conjunto de documentos + un conjunto de categorías + un conjunto de juicios de relevancia (qué documentos están en cada categoría)
  - Algunos ejemplos
    - Reuters-21578 - noticias de la agencia Reuters, categorías basadas en el contenido - 21578 docs, 135 categorías
    - OHSUMED - categorías Medical Subject Headings (MeSH) - cerca de 3000 categorías
    - Yahoo! Science, Industry Sector, Newsgroups, WebKB, Spambase, etc.

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico Recursos disponibles

- Desambiguación (WSD)
  - Un conjunto de documentos + un diccionario + un conjunto de juicios de relevancia (cuál es el significado de cada aparición de una palabra respecto al diccionario)
  - Algunos ejemplos
    - Semcor - subconjunto del Brown Corpus y “The Red Badge of Courage” - 250000 palabras, etiquetadas con significados de WordNet
    - Interest - 2369 oraciones con la palabra “interest”, etiquetadas con significados de WordNet

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico Recursos disponibles

- Diccionarios electrónicos
  - Diccionarios tradicionales en formato electrónico
  - Algunos ejemplos
    - LDOCE - Longman Dictionary of Contemporary English
    - Internet Dictionary Project

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Recursos disponibles

- Thesauri (sing. thesaurus)
  - Tradicionalmente, diccionarios de sinónimos
  - En la actualidad, colecciones de palabras organizadas en grupos relacionados semánticamente
- Algunos ejemplos
  - Roget's Thesaurus

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Recursos disponibles

- Bases de datos léxicas
  - Sistemas con información relativa al léxico de uno o varios idiomas
  - Diccionarios estructurados conceptualmente
- Algunos ejemplos
  - WordNet - inglés
  - EuroWordNet - inglés, español, italiano, holandés, etc.
  - EDR - inglés y japonés

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Recursos disponibles

- Programas de recuperación de información
  - Diseñados para la tarea de recuperación de documentos
  - Frecuentemente, útiles para otras tareas como categorización, filtrado, etc.
  - Algunos ejemplos
    - Smart - modelo del espacio vectorial
    - Inquiry - redes bayesianas de inferencia
    - Oracle ConText - tecnología propietaria

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico

### Recursos disponibles

- Programas de aprendizaje máquina
  - Construyen clasificadores *automáticamente*
  - Aplicables a múltiples tareas (categorización, desambiguación, etiquetado sintáctico, filtrado...)
  - Algunos ejemplos
    - Rainbow - aprendizaje bayesiano
    - SVM<sup>light</sup> (Support Vector Machines) - aprendizaje de funciones lineales con umbral
    - Ripper - aprendizaje inductivo de reglas
    - C4.5 - aprendizaje de árboles de decisión
    - WEKA - múltiples métodos en Java

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico Evaluación

- Dos criterios básicos
  - *Eficiencia*
    - Cómo es de rápido el sistema (pruebas empíricas, complejidad teórica)
  - *Eficacia o efectividad*
    - Cuál es la calidad de la clasificación (métricas de efectividad)
    - Suele ser el único aspecto considerado
  - Se suelen ignorar otros aspectos (p. ej. facilidad de uso) aunque en general, las técnicas son *portables y escalables*

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## Análisis estadístico Evaluación

- Métricas de efectividad
  - Típicamente provenientes del campo de la recuperación de información y del aprendizaje automático
  - Basadas en contar el número de aciertos o fallos al asignar categorías a expresiones

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# Introducción al PLN

## El problema de la ambigüedad

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## El problema de la ambigüedad

- Uno de los grandes problemas en el PLN
- Se produce cuando una expresión en LN posee más de una interpretación (es decir, cuando en el lenguaje de destino se le pueden asignar dos o más expresiones distintas)
- Se presenta en todos los niveles del lenguaje

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# El problema de la ambigüedad

- Ambigüedad sintáctica
  - Categoría sintáctica
    - "Loro" = N,V
  - Estructural
    - "Juan vio (a María (con el telescopio))"
    - "Juan vio (a María) (con el telescopio)"

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# El problema de la ambigüedad

- Ambigüedad semántica
  - Significado de las palabras (polisemia)
    - "Banco"
      - Entidad financiera.
      - Asiento en un parque.
      - Acumulación de arena en un río.
  - **Ámbito de los cuantificadores**
    - "Todo chico ama a un perro"
      - $\forall X$  (perro(X) &  $\exists Y$  (chico(Y)  $\rightarrow$  ama(X,Y)))
      - $\exists Y$  (chico(Y)  $\rightarrow \forall X$  (perro(X)  $\rightarrow$  ama(X,Y)))

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## El problema de la ambigüedad

- Ambigüedad semántica
  - Roles o casos
    - "Pedro corrió los mil quinientos en (los Juegos Olímpicos/tres minutos y medio)"
    - "En" expresa tiempo y lugar
- Ambigüedad pragmática
  - Ambigüedad referencial
    - "Cogí la tarta de la mesa y me la comí"
    - ¿"la" se refiere a la tarta o a la mesa?

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## El problema de la ambigüedad

- Ambigüedad pragmática
  - Ambigüedad en el discurso (Literalidad)
    - "(1) ¿Puedes abrir la puerta? (2) Tengo frío"
    - ¿Interpretamos (1) literalmente o como acto expresivo indirecto?
    - Literalmente: Al volver de un paseo por la calle y llegar a casa, (1) es una pregunta "sí/no" y la respuesta (2) no es adecuada
    - Indirectamente: En una habitación con calefacción, (1) es una solicitud, y (2) es una respuesta adecuada

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## El problema de la ambigüedad

- La ambigüedad en un nivel del lenguaje se puede resolver usando conocimiento de otro nivel
  - Ambigüedad estructural
    - "Pedro mató (al ciervo (con el rifle))"
    - "Pedro mató (al ciervo) (con el rifle)"
    - Se puede resolver usando conocimiento semántico (un rifle sólo puede ser un instrumento) o conocimiento del mundo (los ciervos no usan rifles)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

## El problema de la ambigüedad

- Ambigüedad referencial
  - "Meí la camiseta en la lavadora y la lavé"
  - ¿"la" se refiere a la camiseta o a la lavadora?
  - Se puede resolver usando conocimiento semántico (la lavadora es un instrumento y la camiseta el objeto) o conocimiento del mundo (las lavadoras no se lavan, se limpian)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea Madrid

# El problema de la ambigüedad

- Disminución de la efectividad en todas las aplicaciones
  - Traducción automática
    - “car” => automóvil/vagón
  - Recuperación de información
    - Banco => documentos financieros o sobre geología o etc.