

Métodos estadísticos en PLN

ITALICA

Universidad de Sevilla

Víctor J. Díaz Madrigal

Fundamentos de Lingüística

- Categorías sintácticas
- Morfología
 - Nombres y pronombres
 - Determinantes y adjetivos
 - Verbos
 - Adverbios, Preposiciones y conjunciones
- Sintaxis
 - Frases
 - Gramáticas sintagmáticas
 - Dependencias
 - Ambigüedad sintáctica
- Semántica y Pragmática

Categorías sintácticas (gramaticales)

- Categorías (POS, Part of Speech): Conjunto de palabras con comportamiento sintáctico-semántico similar.
- Ejemplos: nombres, adjetivos, verbos, preposiciones
- Test de sustitución: El hombre {canta, come, vive, ...}
- Problemas: palabras con múltiples categorías: bajo
- Las categorías se dividen en dos grupos:
 - abiertas (léxicas): nombres, verbos, adjetivos, adverbios
 - cerradas (funcionales): preposiciones, determinantes
- Diccionario (Lexicon): Conjunto de pares (Palabras, POS)
- POS tags: Nombres de categorías (Estándar: Brown)

Morfología

- Morfología: Estudio de la formación de palabras
- Importancia: Inferir la categoría de una palabra desconocida
- Procesos morfológicos: Inflexión, Derivación y Composición
- Inflexión: No suelen alterar la categoría pero si los rasgos
 - Patrón: Prefijo-Raíz-Sufijo
 - Formación de plurales -s
- Derivación: Alteran la categoría y el significado
 - Pasar de adjetivo a adverbio mediante sufijo mente
- Composición: Agregación de palabras
 - Separadas: New York
 - Unidas: carricoche

Nombres NN y Pronombres NNP

- Representan entidades (personas, animales y cosas)
- Inflexiones para nombres:
 - Género {masculino,femenino,neutro}
 - Número {singular,plural}
- Subclases:
 - Clíticos (Phrasal affix)
 - Pronombres: Es una subcategoría de nombre y actúan como variables que refieren nombres.
 - Inflexiones para pronombres:
 - Caso {nominativo/sujeto, acusativo/objeto}
 - Persona {primera, segunda,tercera}
 - Pronombres reflexivos y posesivos/genitivo
- Anáforas

Determinates D y Adjetivos JJ

- Suelen acompañar y modificar a los nombres
- Determinates: Son una descripción particular de un nombre
 - Artículos AT: (determinado e indeterminado)
 - Demostrativos
 - Cuantificadores
 - Pronombres interrogativos
 - Partículas interrogativas o de relativo
- Adjetivos: Califican a un nombre:
 - Atributivos: Calificación directa
 - Predicativos: Calificación mediante el verbo ser
 - Comparativos y superlativos
- Inflexiones para D y JJ: (Concordancia-Agreement)
 - Número, género y caso

Verbos VB

- Describen acciones
- Inflexiones:
 - persona
 - número
 - tiempo {presente, pasado, futuro}
 - aspecto (perfecto, imperfecto}
 - modo {indicativo, subjuntivo}
 - voz {activa, pasiva}
- Según la regularidad en las inflexiones se dividen en regulares e irregulares
- Verbos modales y auxiliares

Adverbios RB, Preposiciones IN y Conjunciones

- Los adverbios suelen acompañar y modificar a los verbos respecto al tiempo, espacio, manera y grado. Pueden modificar también a adjetivos y otros adverbios
- Las preposiciones y conjunciones acarrean poca semántica y juegan el papel de nexo
- Preposiciones: suelen expresar relaciones espaciales
 - Partículas RP son preposiciones que también juegan el papel acompañan a los verbos (Pharsal verbs)
- Conjunciones: Unen palabras con igual categoría
 - Coordinantes CC: (No hay frase principal ni secundaria)
 - Subordinantes CS: (Existe una frase principal y secundaria)
 - Complementadores

Frases (Phrase Structure)

- Los lenguajes imponen restricciones en el orden de las palabras (word order) que originan las frases (sintagmas) .
- La sintaxis estudia la formación de frases de una lengua.
- Constituyentes: Agrupaciones que se comportan de forma uniforme y que tienen capacidad de cambiar de posición. Suelen girar en torno a una palabra núcleo (head)
- Clases de frases:
 - Nominal NP: D+JJ+NP (Cláusulas de relativo)
 - Preposicional PP: IN+NP
 - Verbal VP: VB+ NP
 - Adjetival AP:

Gramáticas sintagmáticas

- Estudian la formación y composición de sintagamas:
- Restricciones de construcción:
 - Free word order
 - SVO (Subject-Verb-Object)
- Configuraciones:
 - Declarativa
 - Inversión interrogativa (Wh-movement)
 - Imperativas
- Estrategia basada en reglas de reescritura (CFG)

Dependencias

- Clasificación de dependencias
 - sintácticas: (Sujeto, Objeto Directo, Objeto Indirecto, ...)
 - semánticas (roles) (Agente, Paciente, Instrumento, ...)
- Dependencias más frecuentes
 - Locales
 - No locales (Recurción, Movimiento)
 - Argumentos (Sujeto, Complementos obligatorios)
 - Adjuntos (Subordinadas, Objetos opcionales)
 - Subcategorización (Transitivos, Intransitivos, Ditransitivos)
- El problema de los nodos nulos
- Las dependencias son capturadas mejor en un enfoque racionalista que empirista.

Ambigüedad sintáctica

- Representación de la estructura sintáctica:
 - Árboles sintácticos
 - Corchetes (Bracketing)
- Ambigüedad: Varias representaciones
 - Attachment ambiguity
- Garden paths (Retrotazado por análisis incorrectos)
 - The horse raced past the barn fall
 - El vino de Jerez ayer
- Distinción entre incorrección y anomalía semántica

Semántica

- Estudio del significado de las palabras (semántica léxica) y de las construcciones donde intervienen.
- Relaciones semánticas:
 - Hiperónimo/ Hipónimo: pez-animal animal-peiz
 - Antónimo: frío-calor
 - Merónimo/ Holónimo: hoja-árbol árbol-foja
 - Sinónimos: coche-automóvil
 - Homónimo/ Polisemia/Homófonos (Ambigüedad léxica) bajo/rama/?
- Problemas:
 - Composicionalidad (Estricta combinación)
 - Colocaciones (Combinación+Uso)
 - Expresiones idiomáticas (No hay composición)
 - Ámbito (Cuantificación)

Pragmática

- Estudio de la interacción del lenguaje y el mundo.
- Problemas:
 - Representación del conocimiento
 - Relaciones anafóricas
- Aplicaciones:
 - Análisis del discurso (Sistemas conversacionales)
 - Extracción de información