

Especificación del espacio de accesibilidad anafórico de las descripciones definidas en español para sistemas de resolución computacional de la anáfora: propuesta estructural y tipología

SEL 2001

Borja Navarro Colorado

Grupo de investigación en
Procesamiento del Lenguaje y Sistemas de Información

PLSI

Grupo de Investigación en Procesamiento del
Lenguaje y Sistemas de Información

LS#

Departamento de Lenguajes y Sistemas
Informáticos

 **Universitat d'Alacant**
Universidad de Alicante

Universidad de Alicante



Índice

- Introducción.
- Fundamentos teóricos:
 - Las descripciones definidas: definición y tipología.
 - La estructura del discurso. Problemas para el procesamiento automático del texto.
 - Las unidades textuales y los espacios de accesibilidad estructural.
- Experimentación y resultados.
- Conclusiones y trabajos futuros.



Resolución computacional de la anáfora

- Localización automática del antecedente de una expresión anafórica.
- Estrategias:
 - Fuentes de conocimiento:
 - morfológico, sintáctico, semántico (Ferrández 1998).
 - Espacio de accesibilidad (Martínez-Barco 2001):
 - Completo.
 - Ventanas deslizantes.
 - Estructural.



Espacio de accesibilidad

- Espacio de accesibilidad completo:
 - Se procesan gran cantidad de candidatos inútiles.
 - Se genera gran coste computacional.
 - La posibilidad de producir error en el sistema aumenta exponencialmente.
- Ventan deslizante de oraciones:
 - No hay criterio estable para definir el tamaño de las ventanas: depende del dominio y del corpus de estudio utilizado.



Espacio de accesibilidad

- Espacio de accesibilidad estructural (I):
 - Textos tipo diálogo (Palomar y Martínez-Barco 2001 y Martínez-Barco 2001):
 - Relación anáfora – estructura del discurso (Fox 87).
 - Mayor consistencia: basado en criterios estructurales y textuales.
 - Mayor exactitud: no establece un número fijo de oraciones.
 - Reduce el tiempo de respuesta del sistema y la posibilidad de generar errores.



Espacio de accesibilidad

- Espacio de accesibilidad estructural (II):
 - Texto tipo monólogo (*Navarro et. al. 2001*):
 - Adaptación a una estructura del discurso diferente.
 - Datos concluyentes para los diferentes tipos de expresiones anafóricas planteados: pronombres, adjetivos, alias y expresiones temporales
 - Excepto para las descripciones definidas.



Objetivo del trabajo

- Especificar un espacio de accesibilidad anafórica estructural de las descripciones definidas en textos monologales escritos.
- Tareas:
 - Proponer una clasificación de las DD en español.
 - Determinar unidades textuales formales (independientes del significado):
 - información aportada por las etiquetas HTML.



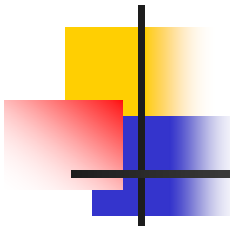
Índice

- Introducción.
- Fundamentos teóricos:
 - Las descripciones definidas: definición y tipología.
 - La estructura del discurso. Problemas para el procesamiento automático del texto.
 - Las unidades textuales y los espacios de accesibilidad estructural.
- Experimentación y resultados.
- Conclusiones y trabajos futuros.



Fundamentos teóricos

- **Descripciones definidas (DD):** sintagmas nominales definidos.
 - *Un sintagma nominal es definido cuando al usarlo el hablante da a entender que aquella entidad a la que se está refiriendo es perfectamente identificable en el contexto de uso (Moreno 1994).*
- **Descripciones definidas anafóricas:** sintagmas nominales definidos que correfieren con un elemento aparecido anteriormente en el discurso (antecedente).



Las descripciones definidas: problemas en LC

- **Problemas** en el tratamiento computacional de las descripciones definidas:
 - A diferencia de otras expresiones anafóricas (pronombre), las descripciones definidas no están vacías de significado.
 - Posibilidad de tener un espacio de accesibilidad muy amplio.
 - Grupo heterogéneo de elementos.
 - Necesidad de una **tipología**.
 - Vieira 1998, Vieira y Poesio 2000, Muñoz 2001.



Las descripciones definidas: clasificación

1. Por su **estructura sintagmática**:
 - 1.1. Según el determinante.
 - 1.2. Según el núcleo.
 - 1.3. Según los complementos.
2. Por su **relación con el antecedente**:
 - 2.1. Relación sintáctico-semántica.
 - 2.2. Relación pragmática.



1. Por su estructura sintagmática

1.1. Según el **determinante**:

1.1.1. DD con determinante artículo (A).

1.1.2. DD con determinante demostrativo (D).

1.1.3. DD con determinante posesivo (POS).

1.2. Según el **núcleo**:

1.2.1. DD con núcleo nombre propio (NPR).

1.2.1. DD con núcleo nombre común.

1.3. Según los **complementos**:

1.3.1. DD con complementos.

1.3.1. DD sin complementos.



2. Por su relación con el antecedente

2.1. Relación **sintáctico-semántica**:

2.1.1. Relación de repetición (A1MN).

2.1.2. Relación léxica: sinonimia, hipo/hiperonimia, proforma léxica (A1RLEX).

2.1.3. Relación con el verbo antecedente (A1VERN).

2.2. Relación **pragmática**:

2.2.1. Explicativas de nombre propios (A2NP).

2.2.2. Relacionadas con el tópico del discurso (A2MT).

2.2.3. Marcos, guiones mentales y conocimiento enciclopédico general –inferencia- (A2INF).



Problema del procesamiento automático del texto

- Texto: unidad comunicativo-funcional (de “uso”).
 - No es una unidad del sistema lingüístico (abstracta), sino del “uso” del sistema.
 - Las características de las unidades textuales y sus relaciones jerárquicas dependen de:
 - cada texto concreto y
 - la situación comunicativa.
 - No se pueden establecer unidades textuales abstractas (Bernárdez 95).
- Problema en Lingüística computacional:
 - cómo especificar las unidades jerárquico-textuales desde un punto de vista general, de tal manera que sea posible aplicarlas a cualquier tipo de texto.



Unidades textuales

- **Párrafo:** unidad textual formal fundamental.
 - Conjunto de oraciones dadas por el productor como unidad independiente (Petöfi 90; Navarro 01).
- **Pareja de párrafo:**
 - Relación núcleo-satélite (Mann y Thompson 76; Pérez 98).



Las unidades textuales.

- **Epígrafe:** determinado por un **subtítulo**.
 - Indica un subtópico o una entidad subtemática de la unidad textual “epígrafe” (Brown y Yule 93).
- **Documento:** determinado por un **título**.
 - Indica el tópico general global o la entidad temática de la unidad “texto/documento” (Brown y Yule 93).



Espacios de accesibilidad

- **Mismo párrafo** donde aparece la descripción definida (MP).
- **Párrafo anterior** (PA).
- **Mismo epígrafe** donde aparece la descripción definida (MEPI).
- **Epígrafe anterior** (EPIANT).
- **Epígrafe anterior al anterior** (EPANT2).
- **Subtítulo** del epígrafe en el que aparece la descripción definida (ST).
- **Título** del texto (T).



Índice

- Introducción.
- Fundamentos teóricos:
 - Las descripciones definidas: definición y tipología.
 - La estructura del discurso. Problemas para el procesamiento automático del texto.
 - Las unidades textuales y los espacios de accesibilidad estructural.
- Experimentación y resultados.
- Conclusiones y trabajos futuros.

Experimentación (I): método



- Análisis de un corpus compuesto por artículos periodísticos.
 - Extraídos de variados periódicos digitales.
 - Ámbitos temáticos diversos: deportes, sociedad, internacional, economía, sucesos, etc.
 - Diferentes formatos de maquetación (títulos y subtítulos).
- Determinación manual del espacio de accesibilidad anafórico estructural (unidad textual) en el que se localiza el antecedente de cada descripción definida.

Experimentación (II): resultados

	TOTAL	MP	PA	MEPI	EPIANT	EPANT2	ST	T
A1MN	290	47	77	60	58	2	19	27
A1RLEX	95	40	26	11	8	0	7	3
A1VERN	3	2	1	0	0	0	0	0
A2NP	25	15	4	1	1	0	2	2
A2MT	13	6	3	0	0	0	1	3
A2INF	20	9	5	1	0	0	2	3
TOTAL	446	119	116	73	67	2	31	38

A	403	101	109	66	61	1	30	35
D	19	12	3	1	1	1	0	1
POS	6	3	1	0	1	0	1	0

NPR	34	9	5	8	5	0	2	5
------------	-----------	----------	----------	----------	----------	----------	----------	----------



Especificación del espacio de accesibilidad estructural

- DD con **relación de repetición** con el antecedente:
 - No tienen un límite de espacio de accesibilidad predefinido.
 - Pueden localizar su antecedente muy alejado.
 - Fácil recuperación.
 - En la mayoría de los casos sigue la regla general de localizar su antecedente en el párrafo anterior o en el mismo párrafo donde aparece la descripción definida anafórica.



Especificación del espacio de accesibilidad estructural

- DD con **relación léxica** (sinonimia, hiperonimia, hiponimia, etc.):
 - Su espacio de accesibilidad queda limitado al epígrafe donde aparece la expresión anafórica (en algunos casos en el epígrafe anterior).
 - Su relación es más compleja de procesar.
 - Necesitan mismo conocimiento activado para poder interpretar la relación léxica.



Especificación del espacio de accesibilidad estructural

- DD con el **verbo** antecedente:
 - Muy pocos casos.
 - Su espacio de accesibilidad queda limitado en el párrafo anterior.
 - La interpretación de la expresión anafórica depende del verbo de la oración antecedente. No puede situarse muy alejado.



Especificación del espacio de accesibilidad estructural

- DD con **relación pragmática con el antecedente nombre propio:**
 - Suelen localizar su antecedente en el mismo párrafo donde aparece la expresión anafórica.
 - Sin embargo, lingüísticamente no tienen límite dado que dependen de un conocimiento previo.



Especificación del espacio de accesibilidad estructural

- DD con **relación pragmática a partir del tópico del discurso:**
 - Su espacio de accesibilidad queda limitado en el párrafo anterior, en el subtítulo o en el título.
 - En el título y/o sub-título se suele enunciar el tópico o subtópico del discurso.



Especificación del espacio de accesibilidad estructural

- DD con **relación pragmática a partir del conocimiento del mundo:**
 - Su espacio de accesibilidad queda limitado al mismo epígrafe, junto a el subtítulo y el título.
 - Es necesario mantener activado el mismo conocimiento.



Especificación del espacio de accesibilidad estructural

- DD por determinante **artículo**:
 - No limita un espacio de accesibilidad
 - Sigue la regla general de preferir el mismo párrafo, seguido del párrafo anterior.
- DD por determinante **demostrativo**:
 - No limita un espacio de accesibilidad
 - La mayoría de las ocasiones su antecedente aparece en el mismo párrafo de la expresión anafórica o el párrafo anterior.



Especificación del espacio de accesibilidad estructural

- DD por determinante **posesivo**:
 - Poco comunes.
 - Su espacio de accesibilidad queda limitado al párrafo anterior.
 - Necesita conocimiento compartido para ser interpretada.
- DD **nombre propio**:
 - No presenta límites en su espacio de accesibilidad.



Índice

- Introducción.
- Fundamentos teóricos:
 - Las descripciones definidas: definición y tipología.
 - La estructura del discurso. Problemas para el procesamiento automático del texto.
 - Las unidades textuales y los espacios de accesibilidad estructural.
- Experimentación y resultados.
- Conclusiones y trabajos futuros.



Conclusiones

- Especificación del espacio de accesibilidad anafórica estructural de las descripciones definidas en español.
 - Implantable en un sistema computacional de resolución de la anáfora.
- Mejoras:
 - Estable: está basado en criterios lingüístico-textuales.
 - Más exacto en la delimitación del EA de cada expresión anafórica (tipología).



Trabajos futuros

- Implantación en un sistema completo de resolución de la anáfora:
 - Módulo previo que reduce los candidatos a procesar.
 - Especifica, antes de la actuación de cualquier otro método de localización del antecedente, qué tipo de antecedente buscar y dónde.
 - Esta información puede ser también incorporada en el módulo de restricciones.