

# La Teoría Sentido Texto

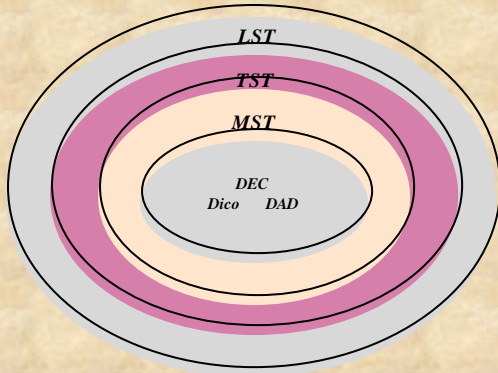
## Introducción

John Jairo Giraldo  
Seminario “Léxico, Terminología y Discurso Especializado”  
Institut Universitari de Lingüística Aplicada  
Marzo 13 de 2003

### Teoría Sentido Texto

#### Contenido

1. Esquema LST-DEC
2. Antecedentes
  - Observatorio de Lingüística Sentido-Texto (OLST)
3. Teoría Sentido-Texto (TST)
  - Ideabística
  - Características
  - Postulados
4. Modelo Sentido-Texto (MST)
  - Componentes/Niveles de representación
5. Lexicología explicativa y combinatoria (LEC)
6. El Diccionario explicativo combinatorio del francés contemporáneo o (DECFC)
  - Generalidades
  - Esquema de elaboración
  - Rasgo particular
  - Características
  - Componentes de la entrada
  - Secciones principales
  - Tipos de Funciones léxicas
  - Diccionario de coocurrencias (DiCo) y Léxico activo del francés (LAF)
  - Usos prácticos de LDEC
7. Dictionnaire analytique de la distribution
  - Generalidades
  - Pros y contras de las FLs



### Antecedentes



#### Observatorio de Lingüística Sentido-Texto (OLST)

- Grupo investigación
- Igor Mel'cuk (Dir.)
- Depto. Ling. & Trad. (U. Montréal)
- Promueve la investigación descriptiva hecha en el marco de la Lingüística Sentido-Texto
- Miembros: Jean Dancette, Marie Claude L'homme, Alain Polguère, entre otros

Antecedentes (2)

OLST

### Objetivos del OLST

- Desarrollo de gramáticas formales basadas en: la semántica, la sintaxis de dependencia y la morfología, postuladas por la TST
- Desarrollo de léxicos formales codificados según los principios de la LEC postulada por la TST
- Implantación informática de los MST

Antecedentes (3)

OLST

### OLST

#### Líneas de investigación

- en lexicología, lexicografía, sintaxis, etc.
  - ⇒ Diccionario explicativo combinatorio del francés contemporáneo (DECFC) o (DEC)
  - ⇒ Base de datos léxica (DiCo) y Léxico activo del francés (LAF)
- en terminología y traducción (Grupo “Éclectik”)

## La Teoría Sentido-Texto (TST)

### Idea básica

“Cualquier acto de comunicación lingüística presupone tres componentes principales:

- Un **contenido**, comunicado por señales lingüísticas, llamado **sentido**.
- Ciertas **formas** o fenómenos físicos (acústicos o gráficos), utilizados para comunicar el contenido, llamado **texto**.
- Un conjunto de **correspondencias** entre sentidos y textos, lo que constituye la **lengua**”.

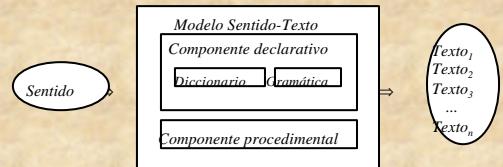
La correspondencia entre estos componentes es multívoca:

- un sentido puede ser expresado por varios textos
- un texto puede expresar varios sentidos.

### Idea básica (2)

(TST)

- La TST es una teoría lingüística que apunta a la descripción de la correspondencia Sentido  $\leftrightarrow$  Texto por medio de la construcción de modelos formales
- Tales modelos se consideran mecanismos lógicos virtuales del tipo:



### Idea básica (3)

(TST)

#### Sentido

*Norm aime sa femme Marie de façon très intense* ∪

#### Textos

- *Norm aime follement sa femme Marie.*
  - *Norm aime sa femme Marie à la folie.*
  - *Norm aime sa femme Marie comme un fou.*
  - *Norm éprouve un amour fou pour sa femme Marie.*
  - *Norm ressent un amour immense pour sa femme Marie.*
- etc.

### Características (1)

(TST)

- Una característica importante es su orientación hacia la síntesis (producción de habla, de sentidos a textos), orientación no corriente en otras teorías lingüísticas.
- Se da un carácter privilegiado al hablante frente al destinatario
- Esta también es la orientación del DEC, lo que constituye una innovación en el campo de los diccionarios monolingües.

### Características (2)

(TST)

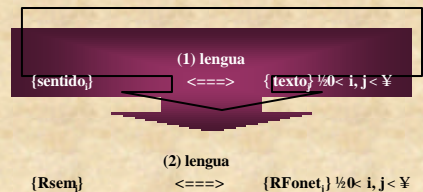
- Da cuenta de la asociación que todo hablante de la lengua L es capaz de hacer entre un sentido dado y el conjunto de enunciados parafrásticos de L que expresan ese sentido → *Norm aime sa femme Marie de façon très intense.*
- Es universal, es decir, que se fundamenta en principios generales que se aplican a todas las lenguas
- Es lingüística en el sentido en que permite construir modelos lingüísticos específicos para cada lengua
- Es formal. Permite construir modelos “calculables” (activables por sistemas lógicos como los programas informáticos)

### Postulados (1)

(TST)

#### Postulado 1

Una lengua natural L es un conjunto finito de CORRESPONDENCIAS multívocas entre un conjunto enumerable de sentidos y un conjunto enumerable de textos



**Postulado 2**

Una lengua natural L, considerada como (2) puede ser descrita sólo por medio de un modelo funcional.

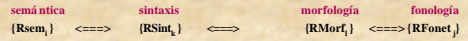
- Un modelo funcional de una L es un sistema de reglas formales que simula el comportamiento lingüístico de los hablantes de L
- Este modelo asocia con una RSem dada de L todas las RFonets que expresan el sentido correspondiente; con una RFonet dada de L todas las RSEms que el texto correspondiente pueda expresar
- A este modelo se le llama **Modelo Sentido-Texto (MST)**
- La TST es una teoría que construye MST de las lenguas actuales, que especifica las correspondencias del tipo (2) para una lengua dada.

**Postulado 3**

Para describir la correspondencia {Rsem<sub>i</sub>} <====> {RFonet<sub>j</sub>} en L, se introducen dos niveles intermedios de representación:

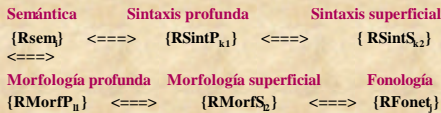
- Representación sintáctica y Representación morfológica

⇒ Se deduce de los postulados 1-3 que un MST es un dispositivo lógico o conjunto de reglas que tiene como estructura general:



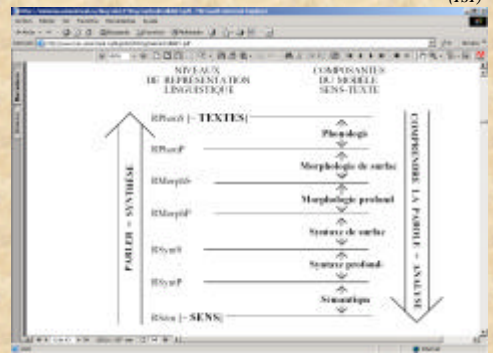
y....

... como estructura detallada:



Todas las representaciones son descritas por diferentes lenguaje s formales:

- Rsem → grafo o red
- RSint → árbol de dependencias
- RMorf y RFonet → cadena de símbolos



La LEC es un estudio de los lexemas de las lenguas naturales que:

- es multilateral (cada lexema está asociado a una información semántica, sintáctica y léxico-combinatoria)
- apunta a la elaboración de diccionarios

- El objetivo de la lexicografía es modelizar el léxico mediante la descripción de los vocablos de la lengua
- La lexicología ofrece una base teórica a la lexicografía



**Generalidades**

- Elaborado a partir de los principios de la LEC, postulada por la TST
- Presenta decenas de vocablos franceses en todas sus facetas (semántica, sintáctica, léxico-combinatoria, morfológica, etc.)
- Es un diccionario de la lengua actual
- En proceso de elaboración (4 vols., 508 vocablos)
- Dicc. idealizado (no tiene en cuenta las consideraciones de tiempo ni de costo)

### Esquema de elaboración

DEC

- **Redacción** { formalidad  
coherencia  
uniformidad  
exhaustividad } → = *Rigurosidad*
- **Nomenclatura:** { disponibilidad  
grado de lexicalización }
- **Datos lingüísticos :** { principio de la primacía del hablante (*la intuición del hablante prima sobre los datos textuales*)  
principio de la constitución obligatoria de BD textuales informatizadas }
- **Estructura de la entrada:** { microestructura  
macroestructura  
lexía (lexema o frasema)  
entrada }

### Rasgo particular

DEC

- No es un diccionario común, está ideado para proporcionar toda la información necesaria para expresar un sentido dado en un contexto dado.
- Se trata de un diccionario orientado hacia “cómo decir X” y no a “qué quiere decir X”.

### Características (1)

DEC

1. Está concebido como **parte integrante de una descripción teórica** de una lengua natural dada
  - La base de esta **descripción** es el **MST**
  - La hipótesis fundamental del MST es que “la lengua natural es una especie de **transformador** que asegura la correspondencia entre un conjunto enumerable de **sentidos** y un conjunto enumerable de **textos** en una lengua dada”.

### Características (2)

DEC

2. El DEC está concebido como una **fuerza completa de información** acerca de las palabras de una lengua dada
  - Es **explicativo** pues hace énfasis en una explicación semántica, exhaustiva y rigurosa de cada lexema inventariado
  - Es **combinatorio** porque describe todos los casos de coocurrencia léxica restringida<sup>1</sup>
  - <sup>1</sup>La coocurrencia léxica restringida se da cuando un lexema 'L' (*ira*, 'emoción negativa...') y un lexema 'L<sub>1</sub>' (*mortal*, que significa 'intenso', como en *odio mortal*) no pueden combinarse para expresar el sentido compuesto 'L+L<sub>1</sub>' (*\*ira mortal* por 'ira intensa'), sin que esto esté prohibido por la sintaxis.

### Características (3) 1/2

DEC

3. Un artículo del DEC está asociado a **un solo lexema**

“La unidad de descripción lexicográfica es un lexema, provisto de un artículo completo y no de una palabra polisémica (=vocablo), como en los demás diccionarios”.

### Características (3) 2/2

DEC

- Cada lexema debe formar al menos con otro lexema un puente semántico
- Dos lexemas se encuentran en relación de polisemia cuando pertenecen a un mismo vocablo
- Dos lexemas se encuentran en relación de homonimia cuando tienen el mismo significante (pronunciación y grafía igual), pero pertenecen a dos vocablos distintos
- Al interior de un mismo vocablo, los números de los lexemas intentan reflejar las distancias semánticas entre los lexemas
  - Números romanos (distancia máxima)
  - Números arábigos (distancia media)
  - Letras minúsculas (distancia mínima)



4. Dado su carácter teórico, el DEC prevé una **estructura rígida** para cada artículo del diccionario

- Un artículo del DEC comprende tres secciones principales:

- Sección semántica o **definición**
- Sección sintáctico-combinatoria o **esquema de régimen**
- Sección léxico-combinatoria o **funciones léxicas**

La descripción de las lexías se organiza en 10 zonas:

1. Zona de la entrada (lexía de entrada, variante ortográf.)
2. Zona fonológica (pronunciación y/o su prosodia particular)
3. Zona morfológica (cat. gram., tipo de declinación/conjugación, formas irregulares, etc.)
4. Zona estilística (marcas de uso)
5. **Zona semántica (definición y connotaciones)**
6. **Zona sintáctico-combinatoria (régimen, restricción sobre la coocurrencia de los diferentes actantes, ejemplos)**
7. **Zona léxico-combinatoria restringida (FLs)**
8. Zona de ejemplos
9. Zona fraseológica
10. Zona de *Nota Bene* (información no formalizada: especificación pragmático-cultural)

1. Sección semántica o Definición

\* Debe ser **adecuada**, i.e., necesaria y suficiente, que identifique el lexema de forma única en todos los empleos posibles.

\* Los **círculos viciosos** son excluidos sistemáticamente de las definiciones, cosa que no ocurre en otros diccionarios como el Robert, ej. *Plaisir*

PLAISIR: "Sensation ou émotion AGRÉABLE, liée à la SATISFACTION d'une tendance, d'un besoin..."

AGRÉABLE: "Qui fait PLAISIR"

SATISFACTION: "PLAISIR qui résulte de l'accomplissement de ce qu'on attend..."

\* Las definiciones deben usar lexemas "desambiguados" por medio de números distintivos entre sí. Otros diccionarios usan palabras polisémicas. Ej: definición de 'cabeza' como parte del cuerpo...

Sin precisar de cuál acepción de las palabras polisémicas *parte* y *cuerpo* se trata.

\* Las definiciones de los lexemas de una misma familia semántica se hacen siempre según un mismo esquema para presentar sistemáticamente sus semejanzas y diferencias.

\* Se procura presentar dos vocablos de un mismo campo semántico (ej. *cabeza* y *corazón*) bajo un mismo esquema.

2. Sección sintáctico-combinatoria o Esquema de régimen

Se presenta en forma de tabla para:

- a. explicitar todos los actantes sintácticos del lexema
- b. especificar para c/u su forma superficial (infinitivo, sintagma preposicional con x preposición, etc)
- c. indicar el actante semántico que le corresponde
- d. mencionar todas las restricciones coocurrenciales de los diferentes actantes sintácticos de un mismo lexema

3. Sección léxico-combinatoria o Funciones léxicas

\* Las FL presentan el conjunto de coocurrencias léxicas restringidas pertinentes al lexema en cuestión.

\* Las FL son una innovación lexicográfica que permite describir sistemáticamente un amplio grupo de locuciones ± fijas que no son expresiones idiomáticas *stricto sensu*. Ej.:

- una FIRME intención
- un argumento DE PESO
- un ruido INFERNAL
- un deseo ARDIENTE
- DAR una lección
- DAR un paso
- COMETER un crimen, etc.

\* El DEC incluye aprox. 60 FL

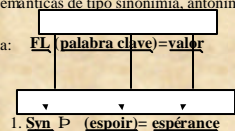
1. FLs paradigmáticas (derivación semántica)

⇒ Son funciones que conectan semánticamente una lexía con otra lexía de la lengua.

⇒ Agrupan todas las relaciones semánticas de tipo **sinonimia**, **antonimia**, **hiperonimia**, etc.

⇒ Se representan bajo el esquema: **FL (palabra clave)=valor**

Ejemplos: Sinonimia (4 tipos)



2. Syn ⊂ (+ estrecho) ⇒ (respeto)= veneración

3. Syn ⊃ (+ amplio) ⇒ (veneración)=respeto

4. Syn ∩ (intersección de sentidos) ⇒ (secundar)= ayudar



## Otros ejemplos:

- Antonimia (Anti)  $\Rightarrow$  (respeto)=irrespeto
- Contrastiva (Contr)  $\Rightarrow$  (mar)=tierra
- Genérica (Gener)  $\Rightarrow$  (armario)=mueble
- Derivada sintáctica
  - Nominalización ( $S_0$ )  $\Rightarrow$   $S_0$  (comprar)= compra
  - Verbalización ( $V_0$ )  $\Rightarrow$  ( $V_0$ ) (promesa)= prometer
  - Adjektivización ( $A_0$ )  $\Rightarrow$  ( $A_0$ ) (escuela)= escolar
  - Adverbialización ( $Adv_0$ )  $\Rightarrow$  ( $Adv_0$ ) (honestidad)= honestamente

## 2. FLs sintagmáticas (colocaciones)

- Conexión en un mismo plano de las diferentes lexías. No se refiere a la frecuencia de uso de una expresión sino a la forma en que el hablante las produce.

## • Tipos:

FLs adjetivales: intensificador (Magn)  
FLs verbales: Verbos soporte (Oper<sub>r</sub>), etc.



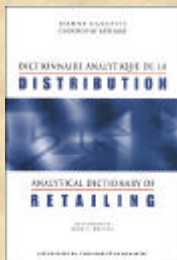
Magn (ruido)= infernal  
Oper<sub>r</sub> (crimen)=cometer



- Se ha pretendido que el DEC llegue a ser un diccionario totalmente informatizado
- Por el momento los autores han desarrollado una versión simplificada del DEC que incluye principalmente las descripciones de los fenómenos de coocurrencia: el DiCo
  - El DiCo es una BD léxica formal que hace énfasis en la descripción de las propiedades combinatorias de las unidades léxicas
- El LAF da cuenta del mismo fenómeno pero adaptado para el uso del público general.
- Se genera totalmente a partir del contenido del DiCo

## Según Mel'cuk:

- Enseñanza de LE, especialmente en niveles avanzados (necesidad de presentación sistemática de locuciones fijas y semifijas según el sentido)
- Tratamiento automático del lenguaje: TA, generación de textos, consulta en BDs
- Empleo como obra de referencia: es un diccionario de sinónimos y antónimos, es fraseológico, analógico y de régimen
- Útil para fines pedagógicos, traductológicos y de edición



## Generalidades

- Dirigido por Jean Dancette (OLST) & Christophe Réthoré
- Orientado a traductores, profesores, estudiantes y profesionales de las áreas comercio y ventas al detal
- Es un diccionario especializado
- Consta de 350 conceptos y 3.500 términos relacionados (FR-EN)
- Está en vías de sistematización

## El Proyecto de sistematización del diccionario busca...

- Pasar de la entrada 'plana' a un modelo relacional
- Estructuras separadas para:
  - Almacenamiento de inf. relacionada con las UTs (inf. gramatical, uso, contextos)
  - Almacenamiento de inf. conceptual (definiciones, rel. semánticas entre términos, equivalencias entre lenguas)
- Integrar otras lenguas (español, alemán e italiano) en forma de módulos
- Activar las **referencias cruzadas** de manera que el usuario tenga acceso ágil a hiperónimos, antónimos, colocaciones, derivaciones, etc.
- Las referencias cruzadas permiten, además, clasificar los tipos de relaciones mediante las FLs.

