



# Seminario de análisis del discurso

**Dr. César Antonio Aguilar**  
**Facultad de Lenguas y Letras**  
**11/10/2010**

**CAguilar@ingen.unam.mx**

# Resumen: grados de centrado (1)

Retomando lo que vimos en la clase pasada, hagamos un resumen respecto a las relaciones de transición que plantea la teoría del centrado:

	$BW(S_{n+1}) = BW(S_n)$ or $BW(S_n)$ undef.	$BW(S_{n+1}) \neq BW(S_n)$
$BW(S_{n+1}) =$ highest in $FW(S_n)$	Center continuation	Smooth shift
$BW(S_{n+1}) \neq$ highest in $FW(S_n)$	Center retaining	Rough shift

De acuerdo con el esquema, tenemos entonces 2 tipos de transiciones básicas: continuas y retenidas. Si son retenidas, hay dos variantes: suaves y bruscas.

## Resumen: grados de centrado (2)

Retomando lo que vimos en la clase pasada, hagamos un resumen respecto a las relaciones de transición que plantea la teoría del centrado:

### **Preference over sequences of transitions:**

Continuation > Retaining > Smooth Shift > Rough Shift

Como vimos la clase pasada, a partir de estas relaciones podemos establecer una escala de gradación que pondera dentro de un texto las transiciones continuas. En caso de que esto no ocurre, entonces se priorizan las transiciones retenidas. Si éstas ocurren, lo idóneo sería que se dieran transiciones suaves. En caso de que no ocurran estas transiciones, entonces se tendrán transiciones duras.

# Teoría de centrado y anáfora (1)

Como hemos en clases previas, un fenómeno característico en toda estructura textual es la manifestación de relaciones anafóricas. Éstas pueden presentar distintos rasgos:

**1. Uso de pronombres con claros rasgos anafóricos:** es uno de los rasgos canónicos esperados en toda relación anafórica: el que un pronombre mantenga una concordancia sintáctica y semántica con su antecedente:



**Jaime** llegó corriendo a **su** casa tan rápido como **[Ø]** pudo

## Teoría de centrado y anáfora (2)

Sin embargo, no en todos los casos se dan estos rasgos de concordancia entre unidades anafóricas.

**2. Cuantificadores y anáforas:** los cuantificadores son un tipo de unidad pronominal que se caracterizan por circunscribir semánticamente los participantes de un evento. Sin embargo, no siempre son del todo claros:



**Todos** tiene los textos que hay que leer, excepto **ellos** [¿?]

## Teoría de centrado y anáfora (3)

La teoría de centrado, si la comparamos con otros modelos que estudian la ambigüedad en anáforas, permite detectarla y solucionarla de forma simple. Veamos:

- 
1. **Juan** le mostró a **Miguel** **la casa** que [Ø] **puso** en renta
2. **Miguel** **le** dijo no estaba seguro de rentar**la**
3. Finalmente, optó por hablar con **su esposa** para ver**la** mañana

**Pregunta:** ¿Qué ambigüedades presentan estas oraciones?

## Teoría de centrado y anáfora (4)

Una de las ventajas que ofrece la teoría de centrado es que podemos hacer uso de otras propuestas, como la teoría del ligamiento, en aras de buscar una solución a estos problemas de ambigüedad.

En este caso, nos conviene recordar los Principios A, B y C que desarrolla Büring:

**Principio A:** Una anáfora está ligada en su categoría regente.

**Principio B:** Un elemento pronominal está libre en su categoría regente.

**Principio C:** Toda expresión-R es libre

## Teoría de centrado y anáfora (5)

Junto con estos principios, retomemos entonces los rasgos de lo que son unidades pronominales y expresiones referenciales (esto es, expresiones-R):

<b>Tipo de partícula</b>	<b>Rasgo anafórico</b>	<b>Rasgo pronominal</b>
Pronombre	— anafórico	+ pronominal
Recíprocos y reflexivos	+ anafórico	— pronominal
Expresiones referenciales (R-Expressions)	— anafórico	— pronominal



# Teoría de centrado y anáfora (6)

Ahora, tomando en cuenta tanto los principios de la teoría de ligamiento, más los rasgos que tienen nombres, pronombres y marcas anafóricas, sumemos también una serie de criterios propuestos por la teoría del centrado:

## **Recency:**

*The doctor found an old map. Jim found an even older map. It described an island.*

## **Grammatical roles determine salience**

Salience hierarchies: Subjects > direct objects > indirect objects

More salient entities are more likely to be referred to again.

## **Repeated mention:**

If an entity has been mentioned several times, it is more likely to be referred to again

## **Parallelism:**

*John went with Jim to Paris. Bill went with him to Rome.*

## **Verb semantics and plausible rhetorical relations:**

*John called Bill. He lost the laptop.*

*John criticized Bill. He lost the laptop.*

## **Selectional preferences:**

*John parked his car in the garage after driving it around for hours*

**Nota:** estos criterios son planteados por Julia Hockenmaier, de la Universidad de Illinois.

# Teoría de centrado y anáfora (7)

¿Cuál es entonces la explicación que damos respecto a las relaciones anafóricas que tenemos en nuestras oraciones 1, 2 y 3? Veamos por partes

1. **Juan** le mostró a **Miguel** **la casa** que **[Ø]** puso en renta

FW: {**Juan**, **Miguel**, **la casa**,  
**[Ø]** marca de flexión}

BW: {No hay antecedentes  
previos}  
Transición = continua

2. **Miguel** **le** dijo no estaba seguro de rentar**la**

FW: {**Miguel**, **le** (clítico), **la**  
**(clítico)**}

BW: {**Miguel**, **Juan**, **la casa**}  
Transición = retenida suave

# Teoría de centrado y anáfora (8)

1. **Juan** le mostró a **Miguel** **la casa** que [Ø] **puso** en renta

FW: {**Juan**, **Miguel**, **la casa**,  
[Ø] **marca de flexión**}

BW: {No hay antecedentes  
previos}  
Transición = continua

2. **Miguel** **le** dijo no estaba seguro de rentar**la**

FW: {**Miguel**, **le (clítico)**, **la**  
**(clítico)**}

BW: {**Miguel**, **Juan**, **la casa**}  
Transición = retenida suave

3. Finalmente, optó por hablar con **su familia** para ver**la** mañana

FW: {**Su familia**, **la (clítico)**}

BW: {**Miguel**, **la casa**}  
Transición = retenida suave

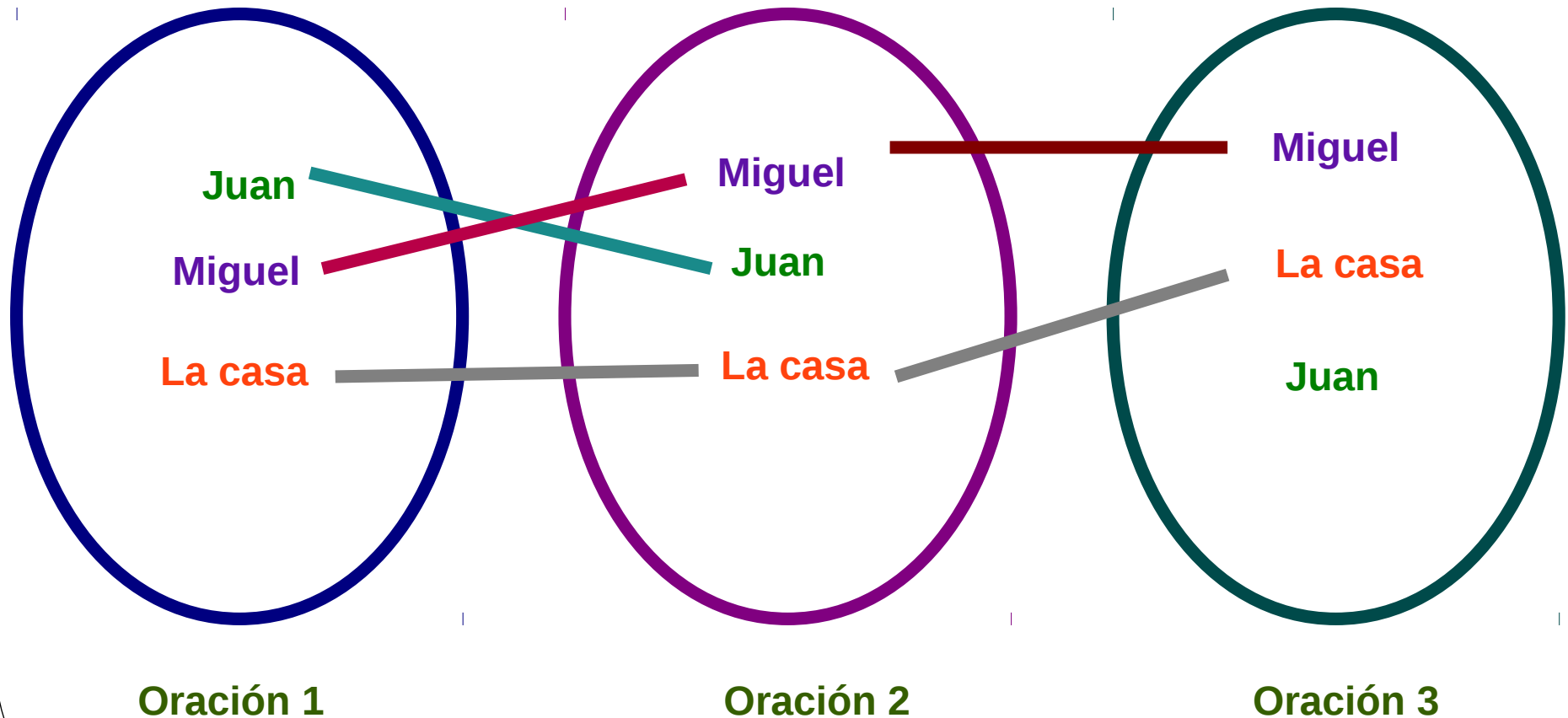
# Teoría de centrado y anáfora (9)

¿Qué otras cosas observamos en estas oraciones?:

1. Entre (1) y (2), los clíticos siguen el principio B: son elementos libres que pueden ser anafóricos con unidades nominales (“Juan”, “la casa”).
2. La unidad referencial relevante es “Miguel”, de modo que en (2) pasa a situarse en posición de Sujeto.
3. En la transición de (1) a (2), “Juan” pasa de ubicarse en posición de Sujeto (+ Topical) a la de Objeto Indirecto (- Topical), dando pie a que “Miguel” asuma esta posición. Por ello, se da una transición retenida suave.
4. Entre (2) y (3) se corrobora esta transición, de modo que “Miguel” es el antecedente referencial de “su familia”, considerando el principio C.
5. Sin embargo, “la casa” sigue una transición continua, si consideramos que se establecer relaciones anafóricas entre dos clíticos y una frase nominal. Con todo, este elemento no tiene un peso topical como el de “Juan” o “Miguel”.

# Teoría de centrado y anáfora (10)

Si vemos así estas relaciones entre transiciones y elementos anafóricos, ¿podemos decir que es equivalente a un mapeo entre conjuntos? Veamos:



# Teoría de centrado y probabilidades (1)

Considerando esta última perspectiva, ver las relaciones de centrado entre elementos nominales y pronominales como un mapeo entre conjuntos, resulta viable aplicar métodos probabilísticos para determinar qué clase de elementos tenemos relacionados anafóricamente entre las oraciones de un texto.

Ahora bien, para hacer esto probabilísticamente, vamos a partir de dos supuestos básicos:

- I. Pensemos que podemos encontrar todas las relaciones correctas de centrado dentro de un texto. Entonces somos muy precisos para localizarlas.
- II. Podemos localizar tales relaciones, pero no son correctas del todo. Quiere decir entonces que, si bien conocemos el fenómeno, nos faltan datos para corroborarlo

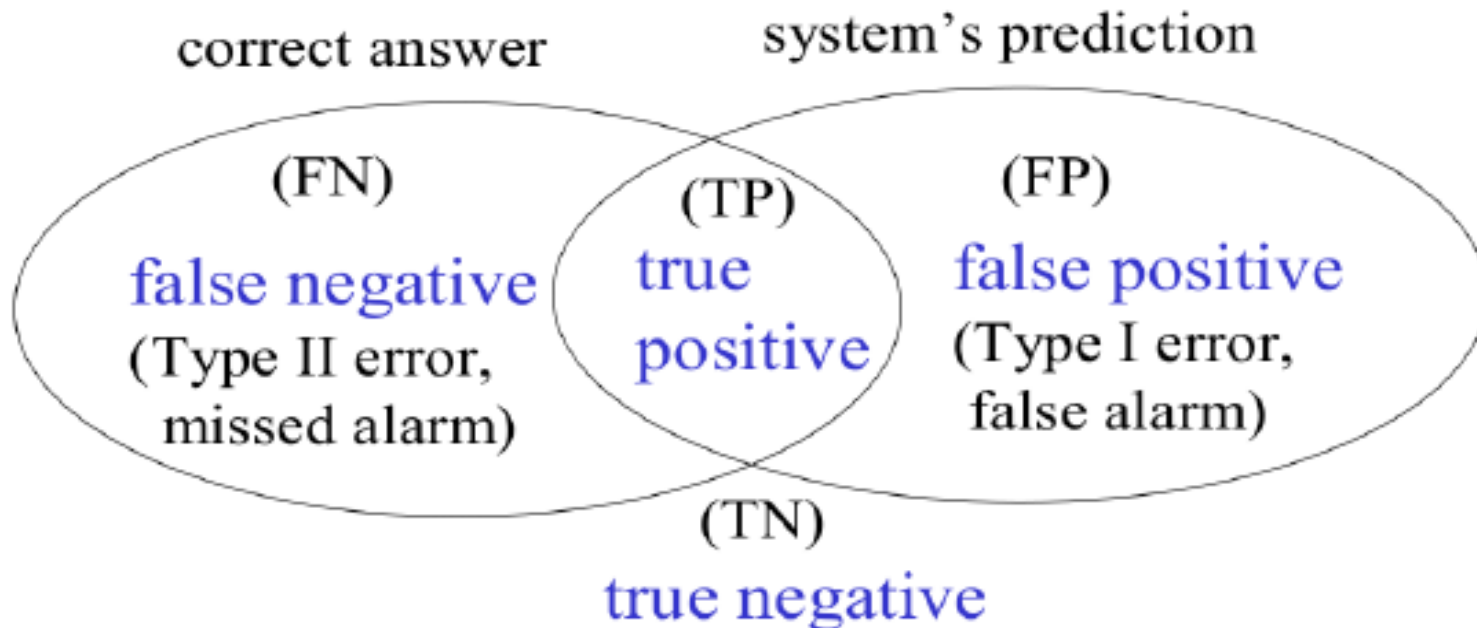
	$H_0$ es cierta	$H_1$ es cierta
Se escogió $H_0$	No hay error	Error de tipo II
Se escogió $H_1$	Error de tipo I	No hay error

$$P(\text{escoger } H_1 | H_0 \text{ es cierta}) = \alpha$$

$$P(\text{escoger } H_0 | H_1 \text{ es cierta}) = \beta$$

## Teoría de centrado y probabilidades (2)

Si partimos de estos dos principios, digamos entonces que lo ideal sería **encontrar todas las relaciones de centrado correctas que hay dentro de un texto**. Ahora, un problema que tenemos es que, si hacemos esto de forma automática, **no podemos determinar todas las relaciones que son correctas o incorrectas**. Vamos a resolver esta duda a partir de aproximaciones, esto es:



# Teoría de centrado y probabilidades (3)

Tenemos entonces 4 opciones:

<b>Tipo de respuesta</b>	<b>Características</b>
Verdaderos positivos	Las relaciones de centrado que localizamos son correctas.
Falsos positivos	Podemos relaciones de centrado, pero éstas no son siempre correctas.
Falsos negativos	Descartamos relaciones de centrado que son correctas.
Verdaderos negativos	Las relaciones de centrado que localizamos son incorrectas.



# Teoría de centrado y probabilidades (4)

Para hacer este análisis, nos basaremos en un modelo probabilístico denominado **Precision & Recall** (Presición y Recuperación):



Keith van Rijsbergen (1943):

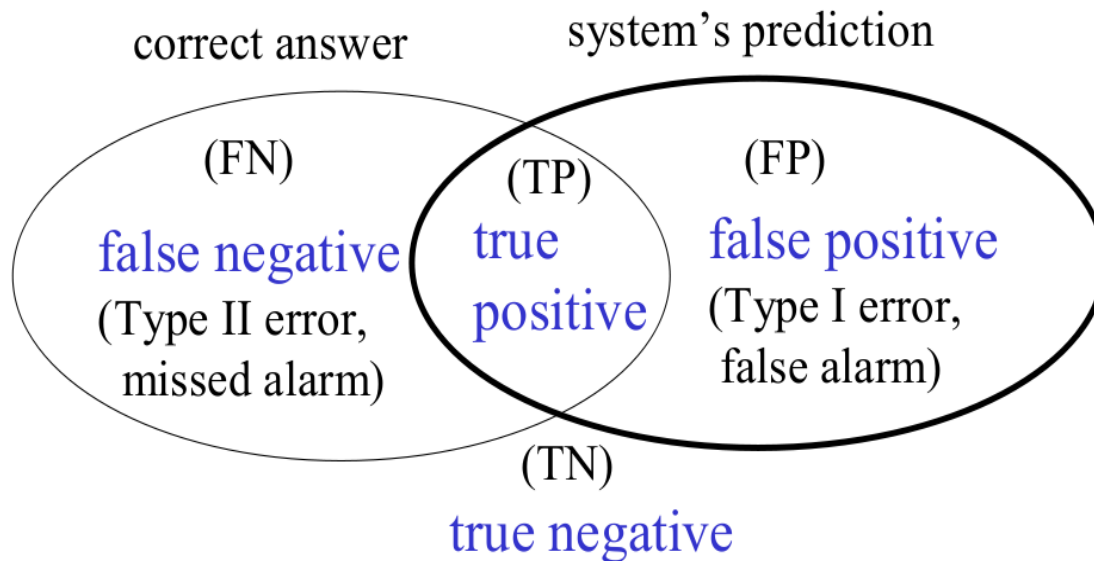
[www.dcs.gla.ac.uk/~keith/](http://www.dcs.gla.ac.uk/~keith/)

Este método fue propuesto para PLN por C. J. “Keith” van Rijsbergen, un lingüista computacional de la Universidad de Glasgow, orientándose sobre todo para resolver tareas de recuperación de documentos. De hecho, si les interesa, tiene un libro electrónico dedicado a este tema, el cual pueden consultar en la siguiente liga:

[www.dcs.gla.ac.uk/Keith/Preface.html](http://www.dcs.gla.ac.uk/Keith/Preface.html)

# Teoría de centrado y probabilidades (5)

De acuerdo con este método, nos movemos entre dos polos: (i) lograr unos resultados positivos ideales de un 100% (lo cual, en la vida real, no sería posible), y unos resultados pésimos de un 0% (lo que sería el peor de todos los mundos posibles). Tomando en cuenta estos polos, cuando decimos que somos “precisos”, queremos dar a entender que:



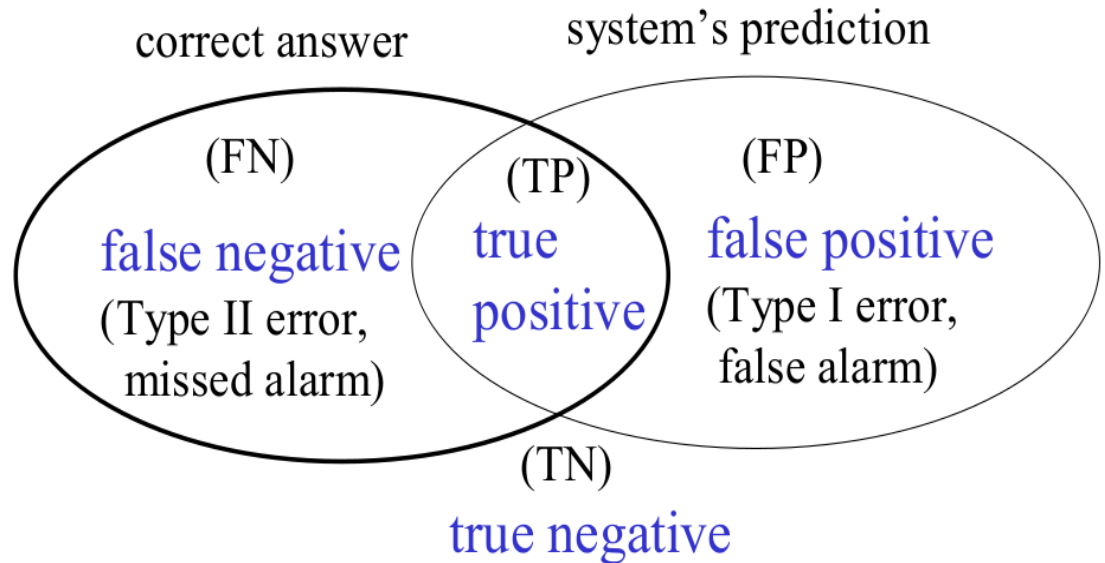
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Nos vamos aproximando a ese resultado ideal, sacando resultados que se aproximan a verdaderos positivos. Si es así, los falsos positivos nos dan pistas para llegar a un buen resultado.

# Teoría de centrado y probabilidades (6)

Y cuando decimos que “recuperamos” una relación de centrado, lo que estamos formulando es que si bien nuestro método de búsqueda logra dar con tales relaciones (imaginen que están usando una herramienta computacional para reconocer estas relaciones de centrado), éstas no necesariamente son correctas.

Lo que requerimos entonces es encontrar un balance entre resultados buenos y malos, para ver qué tanto nos aproximamos a obtener una gran cantidad de los primeros, reduciendo los segundos.



$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

# Teoría de centrado y probabilidades (7)

*Ergo*, podemos elaborar una fórmula como la siguiente:

**Precisión** =  $\frac{|\{\text{RCs relevantes}\} \& \{\text{RCs recolectadas}\}|}{|\{\text{RCs recolectadas}\}|}$

**Recuperación** =  $\frac{|\{\text{RCs recuperadas}\} \& \{\text{RCs relevantes}\}|}{|\{\text{RCs relevantes}\}|}$

## Teoría de centrado y probabilidades (8)

Considerando estas posibilidades de usar métodos probabilísticos para detectar relaciones de centrado, Massimo Poesio ha desarrollado el proyecto **The Gnome Corpus**, el cual consiste en una colección de documentos etiquetados para identificar relaciones de anáfora siguiendo la teoría de centrado.



Massimo Poesio  
University of  
Essex

<http://cswww.essex.ac.uk/poesio/>

Proyectos como el de Poesio permiten explotar la teoría de centrado como una vía para resolver problemas de correferencia y anáfora en textos, a través del uso de herramientas computacionales, así como métodos probabilísticos fiables.

**Gracias por su atención**

**Blog del curso:** <http://discurso-uaq.weebly.com/>