



Seminario de análisis del discurso

Dr. César Antonio Aguilar
Facultad de Lenguas y Letras
04/10/2010

CAguilar@ingen.unam.mx

Lingüística computacional y discurso (1)

En esta nueva unidad, vamos a abordar la revisión de dos modelos teóricos, aplicados a sistemas computacionales, que ayudan a realizar tareas relacionadas con el análisis de textos (o discursos).

1. La teoría del centrado (ing. *Centering Theory*).
2. La teoría de la estructura retórica (ing. *Rhetorical Structure Theory*)

Para ello, vamos a tratar de integrar dos puntos de vista que hemos revisado en las unidades anteriores:

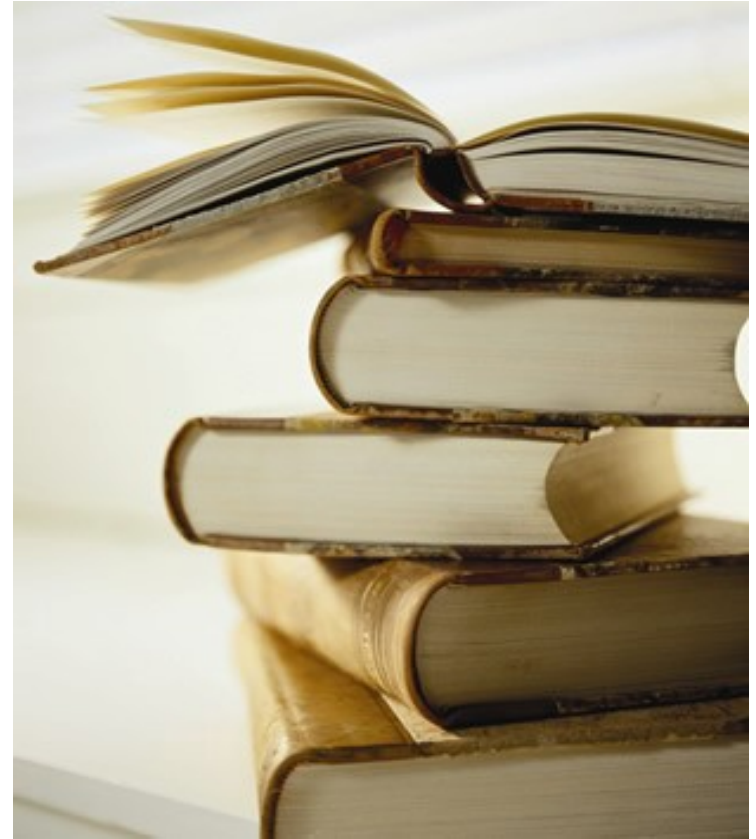
- **Un punto de vista lingüístico:** ¿cuáles son las características principales de un texto como unidad lingüística?
- **Un punto de vista computacional:** ¿cómo podemos reconocer estas características en corpus textuales?

Lingüística computacional y discurso (2)

Retomemos algunos conceptos que hemos visto previamente:

Cohesión: se trata de un fenómeno lingüístico que ocurre en un plano textual, el cual consiste en la conexión de dos o más segmentos (sean palabras, frases y/u oraciones), relacionados temáticamente.

Coherencia: se trata de un proceso mental que nos permite organizar una serie de eventos, situaciones o estados formulados por segmentos textuales, dándoles una estructuración secuencial y lógica.



Lingüística computacional y discurso (3)

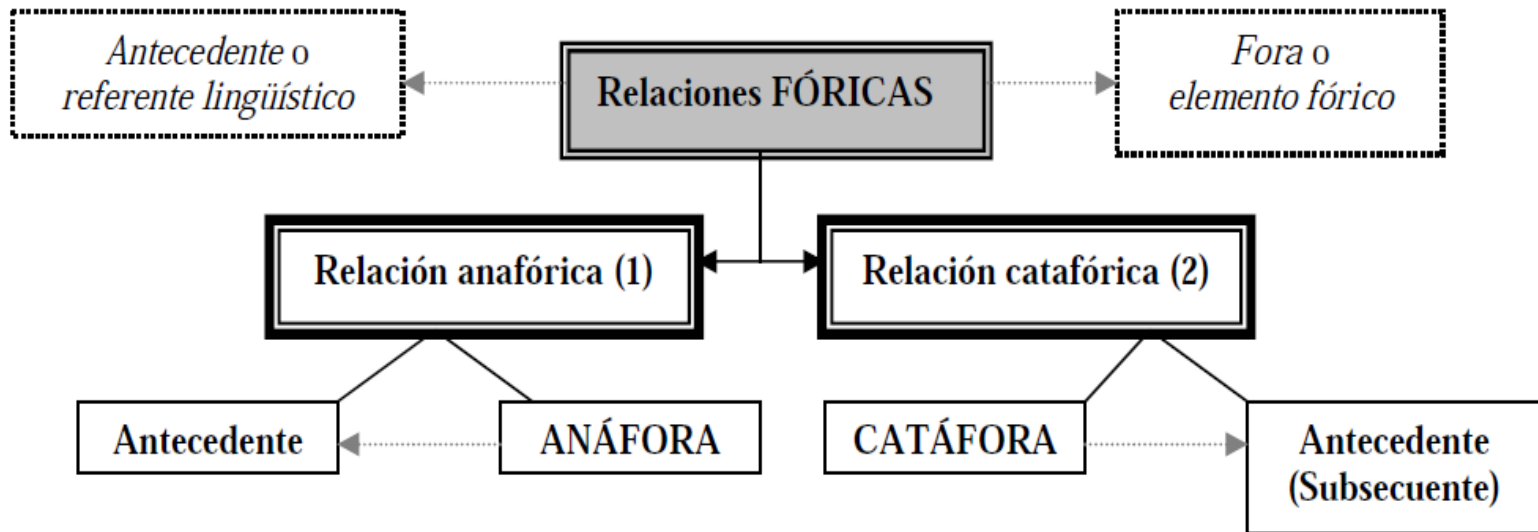
Referencia y correferencia: son dos tipos de relaciones que se dan entre palabras o frases insertas dentro de segmentos textuales, las cuales aluden a una entidad o una situación concreta de la realidad. En el caso de la correferencia, se trata de un vínculo semántico que se establece entre palabras o frases dentro de un texto.

	Tipo de expresión referencial	Ejemplos	Contenido referencial o léxico	Supeditada al discurso
1	Nombres propios	<i>Gómez Orea, M. Godron y G. Merriam, UICN, Colombia.</i>	SI	NO
2	Expresiones definidas (SN con determinante definido)	<i>El apartarrayos, el proceso de cavitación, el micro corrientómetro digital portátil.</i>	SI	NO
3	Expresiones indefinidas (SN con determinante indefinido)	<i>Una bobina, una trampa de línea, un relevador del tipo sencillo.</i>	SI	NO
4	Sintagmas nominales con cuantificador	<i>Algunos resultados, todos los dispositivos de arranque.</i>	SI	NO
5	Expresiones con posesivos	<i>Su actuación, su trayectoria, sus recursos naturales.</i>	SI	SI
6	Expresiones con demostrativo	<i>Este modelo, estos fluidos del ciclo, dicho patrón de búsqueda.</i>	SI	SI
7	Pronombres	<i>Yo, éste, ello, la, mí, conmigo.</i>	NO	SI

Nota: tomado de Benítez (2008: 37)

Lingüística computacional y discurso (4)

Anáfora: es una relación sintáctica y semántica que se establece entre unidades morfológicas, gramaticales y léxicas, en donde se deduce si una entidad o un evento han sido mencionados previamente dentro de un segmento discursivo, o al contrario, si serán referidos de manera explícita en algún segmento próximo.



Nota: tomado de Benítez (2008: 46)

Prominencia (1)

Relacionando los 4 conceptos anteriores, deducimos nuevas relaciones que nos ayuda a ubicar cuáles son los segmentos que contienen la información más relevante de un texto, en contraste con otros segmentos cuya función es simplemente establecer una cohesión textual. A estos segmentos los denominamos **tópicos** y **focos**.

TYPE OF EXPRESSION	CENTRES OF ATTENTION	
	<i>Global</i>	<i>Local</i>
1.1. Simple UTTERANCE	SUBJECT	OBJECT
1.2. Extended UTTERANCE	TOPIC	FOCUS
2. TEXT / DIALOG	GENERAL THEME	PARTICULAR THEME

Como recordarán, tópicos y focos ayudan a darle coherencia un texto, apoyándose en relaciones de correferencia y anáfora establecidas entre segmentos (oraciones, párrafos, escenas, etc.).

Prominencia (2)

¿Cómo lográbamos distinguir un elemento topical de uno focal? Básicamente, podíamos hacer pruebas respondiendo a la pregunta: *¿qué pasa?* P. e.:

¿Qué pasa?

- (1) **Úrsula** se compró un coche hoy
- (2) **El coche** se lo compró Úrsula hoy
- (3) **Hoy mismo** se compró Úrsula el coche
- (4) Se lo compró hoy **con un crédito increíble**
- (5) **Hoy** se lo compró
- (6) Se lo compró hoy

Las partes en negritas indican lo que es tópico o foco dentro de una oración (1-5). En el caso de la última oración, tendríamos una oración netamente declarativa, en donde ningún elemento topical o focal está puesto en perspectiva.

Prominencia (3)

Ahora bien, entre los muchos rasgos que podríamos considerar respecto a esta detección de tópico y foco (cantidad silábica, posición sintáctica dentro de una oración, prosodia, etc.), hay que añadir uno nuevo: el hecho de que una unidad léxica o sintáctica sea **prominente** o no respecto a otras.

[But [[Yemen's]_{med/general} president]_{med/poss}]]*Contrastive* says]_{THEME} [[the FBI] _{old/identity} has told [him] _{old/identity}]_{THEME} [[the explosive material] _{med/set} could only have come from [[the U.S.]_{med/general}, [israel]_{med/general}, or [[two arab countries] _{med/set}]_{med/aggregation}·]_{Adverbial}]_{RHEME} [And to [[a former federal bomb investigator] _{new}·]_{Contrastive}]_{THEME} [[that description] _{old/event} suggests]_{THEME} [[a powerful military-style plastic explosive C-4] _{med/set}]_{Answer} [[that] _{old/relative} can be cut or molded into [different shapes] _{new}·]_{RHEME}

Podemos decir que la prominencia es un rasgo que tienen unidades léxicas y sintácticas de poner en relevancia su contenido de información, lo que las hace situarse precisamente en posiciones topicales y focales.

Prominencia (4)

Veamos un ejemplo. En el siguiente texto, ¿qué unidades léxicas y/o sintácticas considerarían como prominentes?:

“ONE DAY, everything Bill Gates has sold you up to now, whether it's Windows 95 or Windows 97, will become obsolete,” declares Gilbert Amelio, the boss at Apple Computer. “Gates is vulnerable at that point. And we want to make sure we're ready to come forward with a superior answer.”

Bill Gates vulnerable? Apple would swoop in and take Microsoft's customers? Ridiculous! Impossible! In the last fiscal year, Apple lost \$816 million; Microsoft made \$2.2 billion. Microsoft has a market value thirty times that of Apple.

Outlandish and grandiose as Amelio's idea sounds, it makes sense for Apple to think in such big, bold terms. Apple is in a position where standing pat almost certainly means slow death.

It's a bit like a patient with a probably terminal disease deciding to take a chance on an untested but promising new drug. A bold strategy is the least risky strategy. As things stand, customers and outside software developers alike are deserting the company. Apple needs something dramatic to persuade them to stay aboard. A radical redesign of the desktop computer might do the trick. If they think the redesign has merit, they may feel compelled to get on the bandwagon lest it leave them behind.

Prominencia (5)

Veamos algunos candidatos que sobresalen por ser prominentes:

1 APPLE; MICROSOFT

APPLE would swoop in and take MICROSOFT'S customers?

APPLE lost \$816 million;

MICROSOFT made \$2.2 billion.

MICROSOFT has a market value thirty times that of APPLE

it makes sense for APPLE

APPLE is in a position

APPLE needs something dramatic

2 DESKTOP MACHINES; OPERATING SYSTEM

Today's DESKTOP MACHINES, he [Gilbert Amelio] says

Tomorrow's MACHINES must accommodate rivers of data

Time to scrap your OPERATING SYSTEM and start over

The OPERATING SYSTEM is the software that controls

to go with the REENGINEERED OPERATING SYSTEM

3 GILBERT AMELIO; NEW OPERATING SYSTEM

AMELIO, 53, brings a lot of credibility to this task

HIS [Gilbert Amelio] resumé includes

where is AMELIO going to get this NEW OPERATING SYSTEM?

radical redesign in OPERATING SYSTEMS that AMELIO is talking about

Dislocación a la derecha/izquierda (1)

Ahora bien, imaginemos que podemos ligar dos o más párrafos, con miras a construir una larga oración. Hipotéticamente, podríamos decir que el tópico es aquella unidad que se sitúa en la posición inicial de sujeto de esta larga oración. ¿Qué pasa con las unidades que se refieren a dicho tópico, y que aparecen dentro de nuestra oración? Esto es:

El elemento más importante del cromosoma es *la molécula continua de ADN₁. Esta molécula de doble cadena con forma de escalera retorcida₁* está formada por compuestos químicos enlazados.

Antecedente= la molécula continua de ADN

Anáfora de SN directa= Esta molécula de doble cadena con forma de escalera retorcida

Núcleo nominal= molécula

Digamos que son **expansiones** o **dislocaciones** que se sitúan a la derecha o izquierda de nuestro elemento topical (o también focal), los cuales complementan la información referencial que es expresada por dicho tópico.

Dislocación a la derecha/izquierda (2)

Un rasgo interesante de esto es que podemos detectar cuál es la tendencia que sigue una lengua para expandirse dentro de una oración. En el caso del español, pareciera que el patrón normal es extenderse hacia la derecha en esta clase de construcciones:

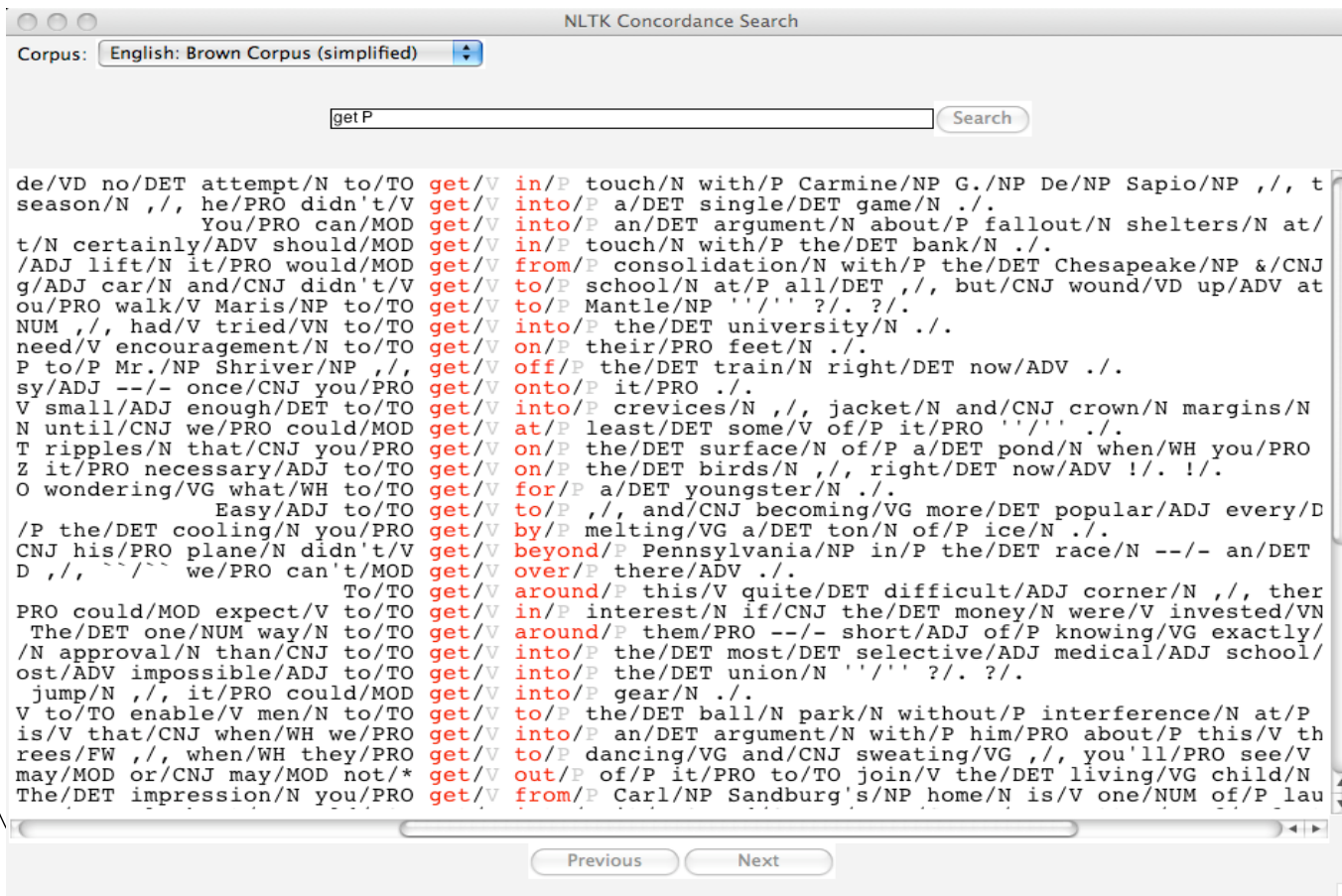
Una computadora ^[Tópico] puede ser definida como **una máquina electrónica** ^[Foco 1] **cuya función** ^[Foco 2] es hacer cálculos lógicos y matemáticos.

Empero, no tenemos problemas para realizar esta clase de expansiones hacia la izquierda:

Si **vienes** ^[Foco 1] de la sala, y pasaste **cerca de la puerta** ^[Foco 2], ¿no viste **las llaves?** ^[Tópico]

La noción de centrado (1)

Dentro de un texto, como hemos visto, se organiza la información de modo que habrá unidades que operarán como puntos nodales, mientras que otras se desempeñarán como dislocaciones a la derecha o a la izquierda de tales nodos.



```
de/VD no/DET attempt/N to/TO get/V in/P touch/N with/P Carmine/NP G./NP De/NP Sapio/NP ,, t
season/N ,, he/PRO didn't/V get/V into/P a/DET single/DET game/N ./
You/PRO can/MOD get/V into/P an/DET argument/N about/P fallout/N shelters/N at/
t/N certainly/ADV should/MOD get/V in/P touch/N with/P the/DET bank/N ./
/ADJ lift/N it/PRO would/MOD get/V from/P consolidation/N with/P the/DET Chesapeake/NP &/CNJ
g/ADJ car/N and/CNJ didn't/V get/V to/P school/N at/P all/DET ,, but/CNJ wound/VD up/ADV at
ou/PRO walk/V Maris/NP to/TO get/V to/P Mantle/NP ''/'' ?/./ ?/./
NUM ,, had/V tried/VN to/TO get/V into/P the/DET university/N ./
need/V encouragement/N to/TO get/V on/P their/PRO feet/N ./
P to/P Mr./NP Shriver/NP ,, get/V off/P the/DET train/N right/DET now/ADV ./
sy/ADJ --/- once/CNJ you/PRO get/V onto/P it/PRO ./
V small/ADJ enough/DET to/TO get/V into/P crevices/N ,, jacket/N and/CNJ crown/N margins/N
N until/CNJ we/PRO could/MOD get/V at/P least/DET some/V of/P it/PRO ''/'' ./
T ripples/N that/CNJ you/PRO get/V on/P the/DET surface/N of/P a/DET pond/N when/WH you/PRO
Z it/PRO necessary/ADJ to/TO get/V on/P the/DET birds/N ,, right/DET now/ADV !/./ !/./
O wondering/VG what/WH to/TO get/V for/P a/DET youngster/N ./
/P the/DET cooling/N you/PRO get/V to/P ,, and/CNJ becoming/VG more/DET popular/ADJ every/D
CNJ his/PRO plane/N didn't/V get/V by/P melting/VG a/DET ton/N of/P ice/N ./
D ,, we/PRO can't/MOD get/V beyond/P Pennsylvania/NP in/P the/DET race/N --/- an/DET
To/TO get/V over/P there/ADV ./
PRO could/MOD expect/V to/TO get/V around/P this/V quite/DET difficult/ADJ corner/N ,, ther
The/DET one/NUM way/N to/TO get/V in/P interest/N if/CNJ the/DET money/N were/V invested/VN
/N approval/N than/CNJ to/TO get/V around/P them/PRO --/- short/ADJ of/P knowing/VG exactly/
ost/ADV impossible/ADJ to/TO get/V into/P the/DET most/DET selective/ADJ medical/ADJ school/
jump/N ,, it/PRO could/MOD get/V into/P gear/N ./
V to/TO enable/V men/N to/TO get/V to/P the/DET ball/N park/N without/P interference/N at/P
is/V that/CNJ when/WH we/PRO get/V into/P an/DET argument/N with/P him/PRO about/P this/V th
rees/FW ,, when/WH they/PRO get/V to/P dancing/VG and/CNJ sweating/VG ,, you'll/PRO see/V
may/MOD or/CNJ may/MOD not/* get/V out/P of/P it/PRO to/TO join/V the/DET living/VG child/N
The/DET impression/N you/PRO get/V from/P Carl/NP Sandburg's/NP home/N is/V one/NUM of/P lau
```

Imaginemos entonces que tenemos una lista de concordancias, y a su derecha e izquierda ubicamos elementos que hacen referencia a tales concordancias.

La noción de centrado (2)

A este fenómeno de establecer puntos nodales dentro de un texto Barbara Grosz (Universidad de Harvard), Aravind K. Joshi y Scott Weinstein (Universidad de Pennsylvania) lo denominaron **centrado** (ingl. *Centering*), y en 1995 escribieron un texto que propone un modelo teórico para su análisis. Para entender mejor, veamos la siguiente secuencia de oraciones:

- 1) a. John went to his favorite music store to buy a piano.
- b. He had frequented the store for many years.
- c. He was excited that he could finally buy a piano.
- d. He arrived just as the store was closing for the day.

La noción de centrado (3)

Siguiendo esta secuencia, no tenemos muchos problemas para ubicar cuáles son los elementos nodales: *John, music store, piano*, más eventos como *to buy* y *to close*.

Ahora, hagamos más formal la relación entre estos nodos: consideremos que cada una de estas oraciones U (ing. *Utterances*) conforma un segmento discursivo DS (ing. *Discourse Segments*), de tal modo que tenemos que cada DS se va a conformar por una secuencia de U :

$$DS = U_1 + U_2 + U_3 + U_n$$

La noción de centrado (4)

Pasemos entonces a analizar lo que ocurre entre las oraciones de nuestro ejemplo. Si consideramos la primera:

John went to **his** favorite music store to buy **a piano**

Consideremos que *John*, *favorite music store* y *a piano* son los tres elementos nodales. ¿Qué relaciones van a establecer con esta oración de base las demás? Pensemos que tienen dos opciones:

a) Aluden a elementos que ya han sido mencionados previamente y establecen vínculos de anáfora (esto es, **backward-looking center**, o C_{bs}).

b) Aluden a elementos que serán mencionados en oraciones subsecuentes, estableciendo relaciones de correferencia o de anáfora, dependiendo de si se tratan de unidades nominales o pronominales (esto es, **forward-looking center**, o C_{fs}).

La noción de centrado (5)

Nuestras relaciones de backward-looking center y forward-looking center quedarían representadas del siguiente modo:

John went to **his favorite music store** to buy **a piano**.

backward-looking center: ? (no previous discourse)

forward-looking center: {**John'**, **store'**, **piano'**}

He had frequented **the store** for many years.

backward-looking center: {**John'**}

forward-looking center: {**John'**, **store'**}

He was excited that **he** could finally buy **a piano**.

backward-looking center: {**John'**}

forward-looking center: {**John'**, **piano'**}

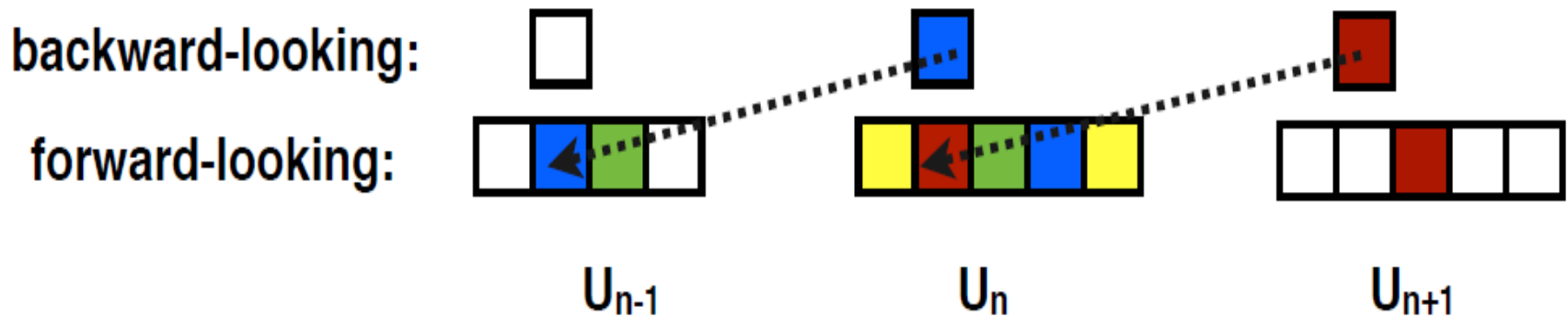
He arrived just as **the store** was closing for the day.

backward-looking center: {**John'**}

forward-looking center: {**John'**, **store'**}

La noción de centrado (6)

En otras palabras, un segmento textual, si es coherente, debe hacer explícito sus segmentos nodales previos y posteriores de una forma clara, de tal modo que podamos rastrearlos y establecer de qué referente se habla previamente, y si este referente tiene una continuación o no en alguna oración subsecuente:



La noción de centrado (7)

Si nuestro segmento textual es coherente, entonces esperamos que tenga el siguiente comportamiento:

$$U_1 = C_b (U_1) = \alpha, C_f + (U_1)$$

$$U_2 = C_b (U_{2+1}) = \alpha, C_f + (U_{2+1})$$

$$U_3 = C_b (U_{3+2+1}) = \alpha, C_f + (U_{3+2+1})$$

Esto quiere decir que nuestras oraciones $U \{1,2,3\}$ comparten un elemento α , el cual mantiene una relación con un antecedente C_b y con un precedente C_f . En la siguiente clase continuaremos viendo a detalle estas relaciones.

Gracias por su atención

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