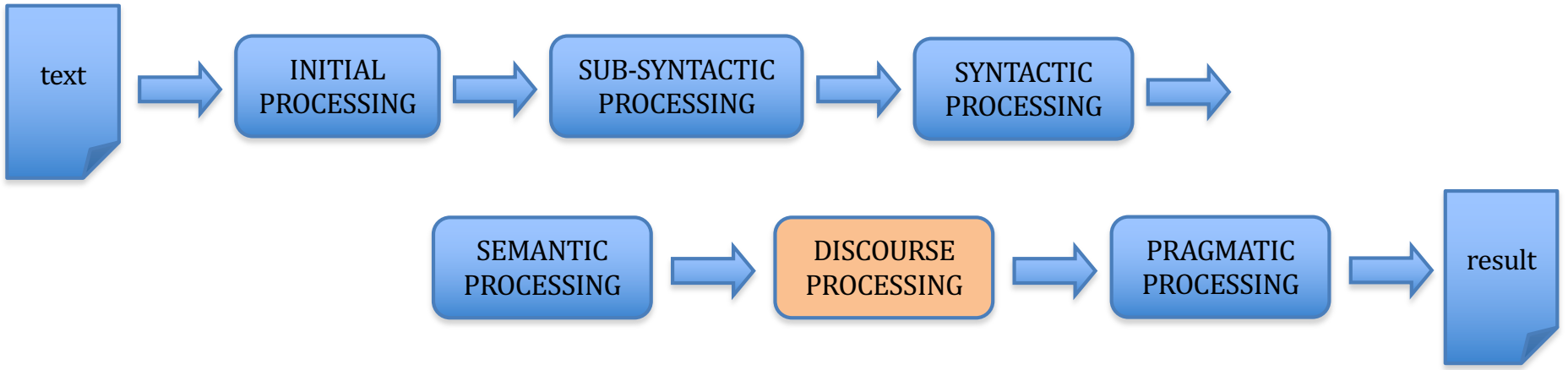
Masteratul de Lingvistică Computațională
Curs: *Introducere in Lingvistica Computațională*

Curs 6
Teorii ale discursului: Centering

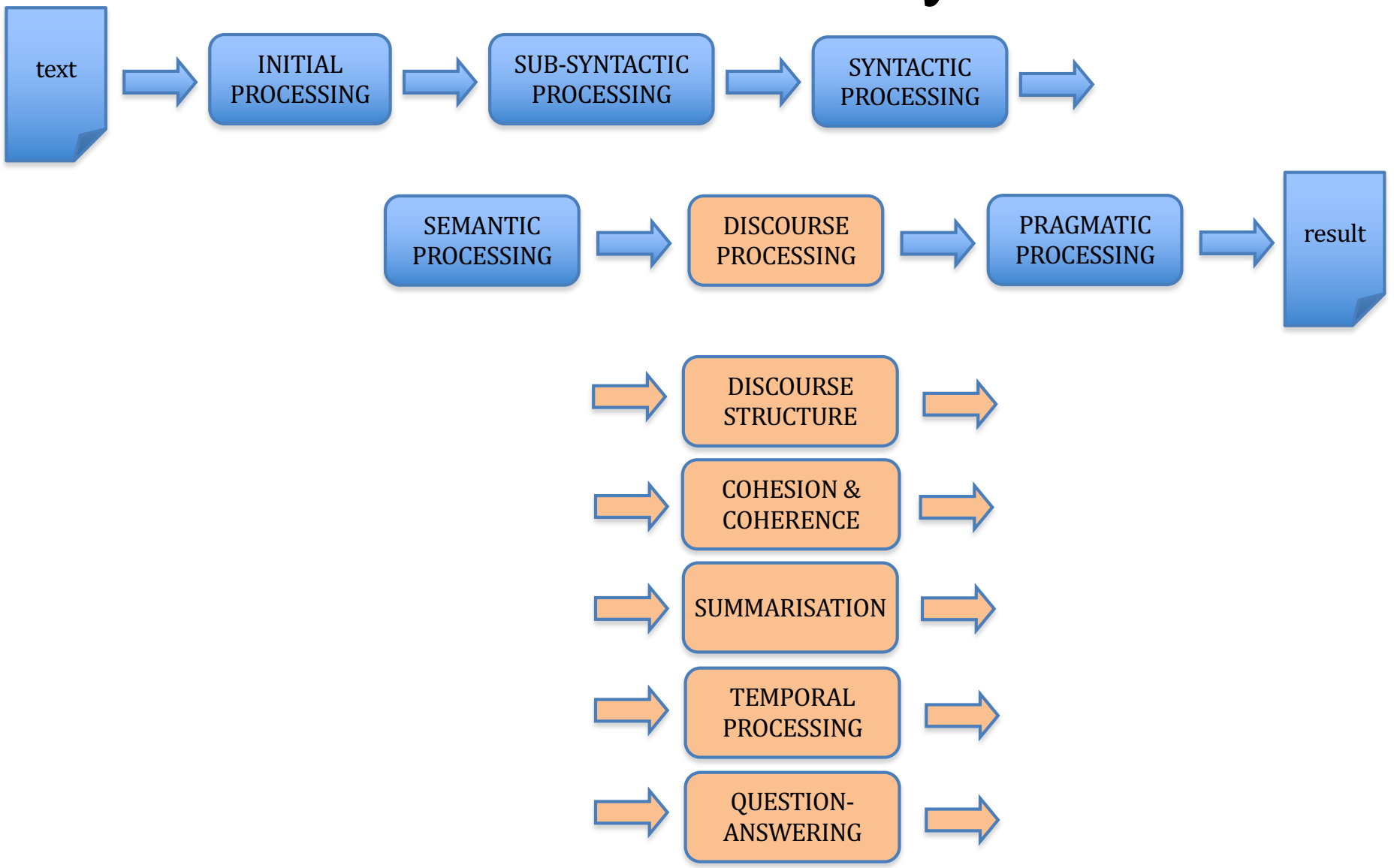
Curs: Dan Cristea

Seminarii & proiect: Mihaela Onofrei, Dan Cristea

The discourse layer



The semantic layer



What is discourse?

Longman: 1. a serious speech or piece or writing on a particular subject: *Professor Grant delivered a long discourse on aspects of moral theology.* 2. serious conversation between people: *You can't expect meaningful discourse when you two disagree so violently.* 3. the language used in particular kinds of speech or writing: *scientific discourse*.

What is discourse?

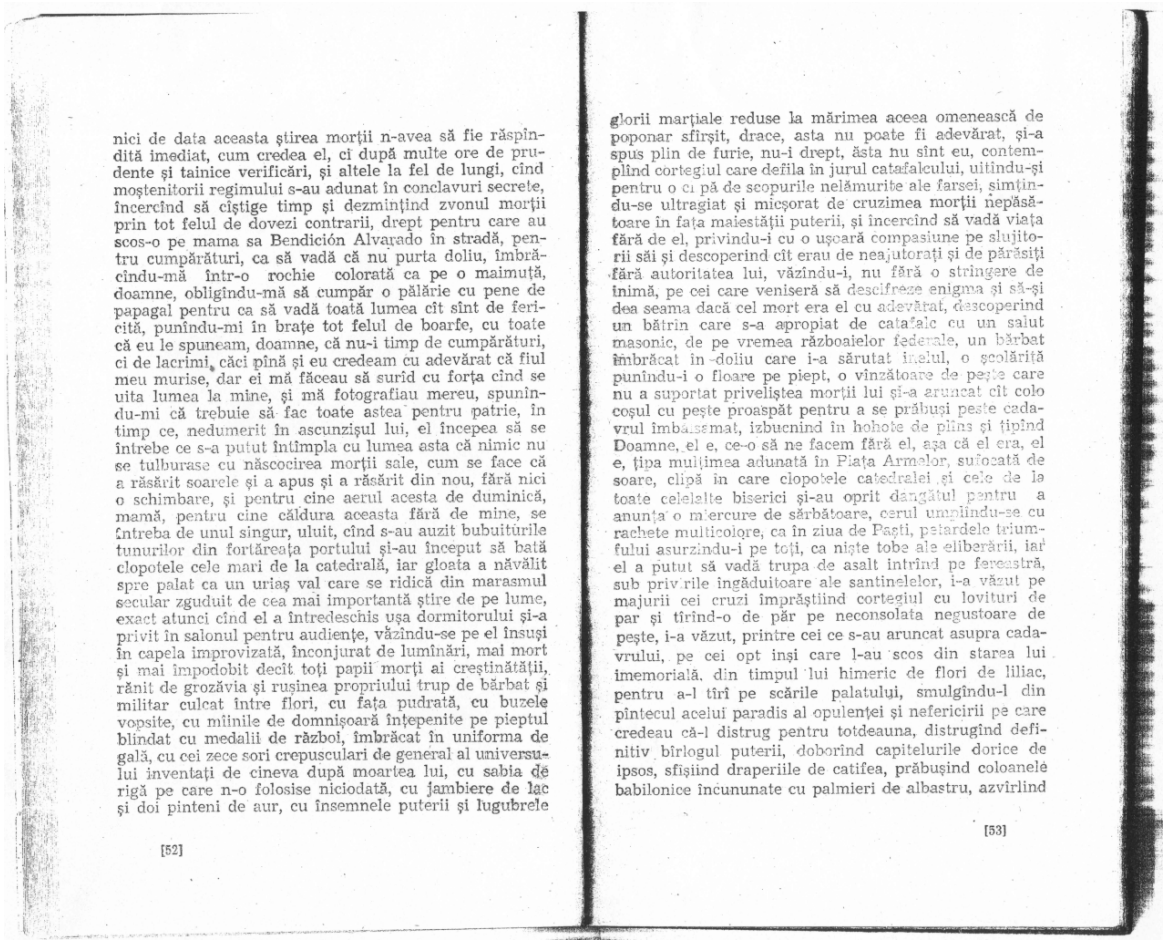
Longman: 1. a serious speech or piece or writing on a particular subject: *Professor Grant delivered a long discourse on aspects of moral theology.* 2. serious conversation between people: *You can't expect meaningful discourse when you two disagree so violently.* 3. the language used in particular kinds of speech or writing: *scientific discourse*.

As you can see, the definition 1 lets aside **the dialogue**. Usually the dialogue is considered closely related to discourse and many techniques devised for discourse can be applied to dialogue as well. In this presentation we will not refer to dialogue.

Text versus discourse

Syntactically – a discourse is more than a single sentence.

From **Garcia Marquez**



nici de data aceasta ştirea morţii n-avea să fie răspîndită imediat, cum credea el, ci după multe ore de prudente şi tainice verificări, şi altele la fel de lungi, cînd moştenitorii regimului s-au adunat în conclavuri secrete, încercînd să cîştige timp şi dezminţind zvonul morţii prin tot felul de dovezi contrarii, drept pentru care au scos-o pe mama sa Bendición Alvarado în stradă, pentru cumpărături, ca să vadă că nu purta doliu, îmbrăcîndu-mă într-o rochie colorată ca pe o maimuţă, doamne, obligîndu-mă să cumpăr o pălărie cu pene de papagal pentru ca să vadă toată lumea cît sînt de fericită, punîndu-mi în braţe tot felul de boarfe, cu toate că eu le spuneam, doamne, că nu-i timp de cumpărături, ci de lacrimi, căci pînă şi eu credeam cu adevărat că fiul meu murise, dar ei mă făceau să surid cu forţa cînd se uita lumea la mine, şi mă fotografiau mereu, spunîndu-mi că trebuie să fac toate astea pentru patrie, în timp ce, nedumerit în ascunzişul lui, el începea să se întrebe ce s-a putut întîmpla cu lumea asta că nimic nu se tulburase cu născocirea morţii sale, cum se face că a răsărit soarele şi a apus şi a răsărit din nou, fără nici o schimbare, şi pentru cine aerul acesta de duminică, mamă, pentru cine căldura aceasta fără de mine, se întreba de unul singur, uluit, cînd s-au auzit bubuiturile tunurilor din fortăreaţa portului şi-au început să bată clopotele cele mari de la catedrală, iar gloata a năvălit spre palat ca un uriaş val care se ridică din marasmul secular zguduit de cea mai importantă ştire de pe lume, exact atunci cînd el a întredeschis uşa dormitorului şi-a privit în salonul pentru audienţe, văzîndu-se pe el însuşi în capela improvizată, înconjurat de luminări, mai mort şi mai împodobit decît toţi papii morţi ai creştinătăţii, rănit de grozăvia şi rusinea propriului trup de bărbat şi militar culcat între flori, cu faţa pudrată, cu buzele vopsite, cu mîinile de domnişoară înţepenite pe pieptul blindat cu medalii de război, îmbrăcat în uniforma de gală, cu cei zece sori crepusculari de general al universului inventaţi de cineva după moartea lui, cu sabia de rigă pe care n-o folosise niciodată, cu jambiere de lac şi doi pinteni de aur, cu însemnele puterii şi lugubrele

[52]

glorii marţiale reduse la mărimea aceea omenească de poponar sfîrşit, drace, asta nu poate fi adevărat, şi-a spus plin de furie, nu-i drept, ăsta nu sînt eu, contemplînd cortegiul care defila în jurul catafalcului, uitîndu-şi pentru o clipă de scopurile nelămurite ale farsei, simţindu-se ultragiat şi micşorat de cruzimea morţii nepăsătoare în faţa maiestăţii puterii, şi încercînd să vadă viaţa fără de el, privindu-i cu o uşoară compasiune pe slujitorii săi şi descoperind cît erau de neajutoraţi şi de părăsiţi fără autoritatea lui, văzîndu-i, nu fără o strîngere de inimă, pe cei care veniseră să descifreze enigma şi să-şi dea seama dacă cel mort era el cu adevărat, descoperind un bătrîn care s-a apropiat de catafalc cu un salut masonic, de pe vremea războaielor federale, un bărbat îmbrăcat în doliu care i-a sărutat inelul, o şcolăriţă punîndu-i o floare pe piept, o vînzătoare de peşte care nu a suportat priveliştea morţii lui şi-a aruncat cît colo coşul cu peşte proaspăt pentru a se prăbuşi peste cadavrul îmbalsamat, izbucnind în hohote de plîns şi ţipînd Doamne, el e, ce-o să ne facem fără el, aşa că el era, el e, ţipa mulţimea adunată în Piaţa Armelor, sufocată de soare, clipă în care clopotele catedralei şi cele de la toate celelalte biserici şi-au oprit dangătul pentru a anunţa o miercure de sărbătoare, cerul umplîndu-se cu rachete multicolore, ca în ziua de Paşti, petardele triumfului asurzindu-i pe toţi, ca nişte tobe ale eliberării, iar el a putut să vadă trupa de asalt intrînd pe fereastră, sub privirile îngăduitoare ale santinelelor, i-a văzut pe majorii cei cruzi împrăştiind cortegiul cu lovituri de par şi tîrînd-o de păr pe neconsolata negustoare de peşte, i-a văzut, printre cei ce s-au aruncat asupra cadavrului, pe cei opt inşi care l-au scos din starea lui imemorială, din timpul lui himeric de flori de liliac, pentru a-l tîrî pe scările palatului, smulgîndu-l din pîntecul acelui paradis al opulenţei şi nefericirii pe care credeau că-l distrug pentru totdeauna, distrugînd definitiv bîrlogul puterii, doborînd capitelurile dorice de ipsos, sfîşiind draperiile de catifea, prăbuşind coloanele babilonice încununate cu palmieri de albastru, azvîrlind

[53]

Text versus discourse

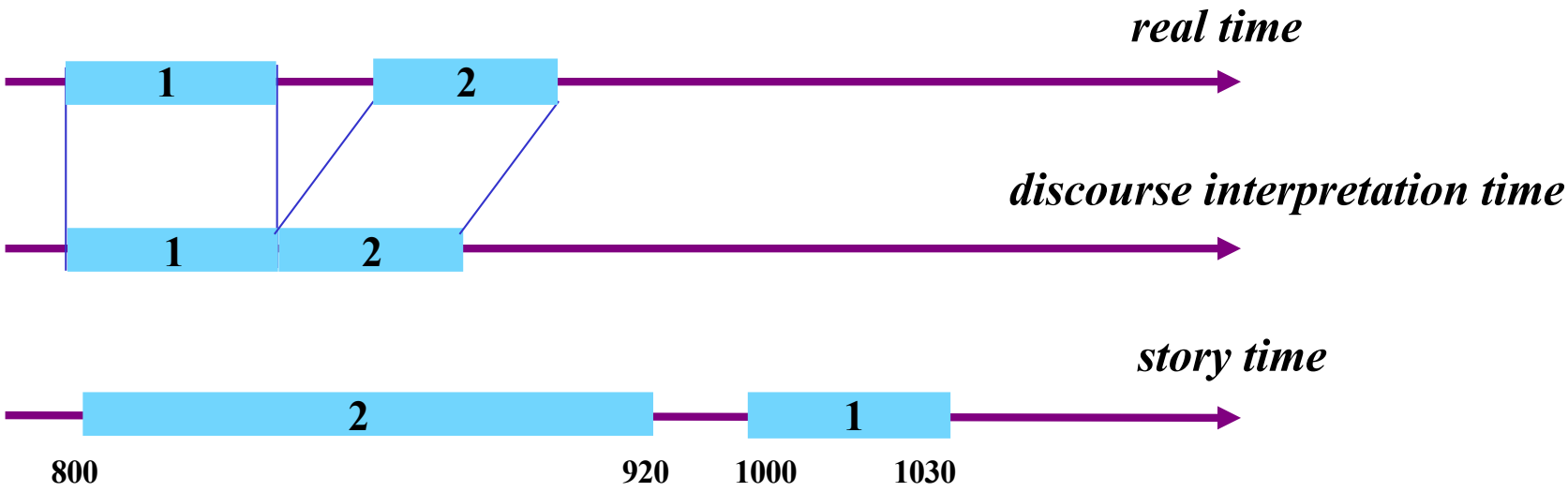
A text is not a discourse!

But it becomes a discourse the very moment
it is read or heard by a human... or a machine.

Time and discourse

Discourse has a dynamic nature

Time axes



Cohesion and coherence

A text manifests **cohesion** when its parts closely correlate.

A text is **coherent** when it makes sense, with respect to an accepted setting, real or virtual.

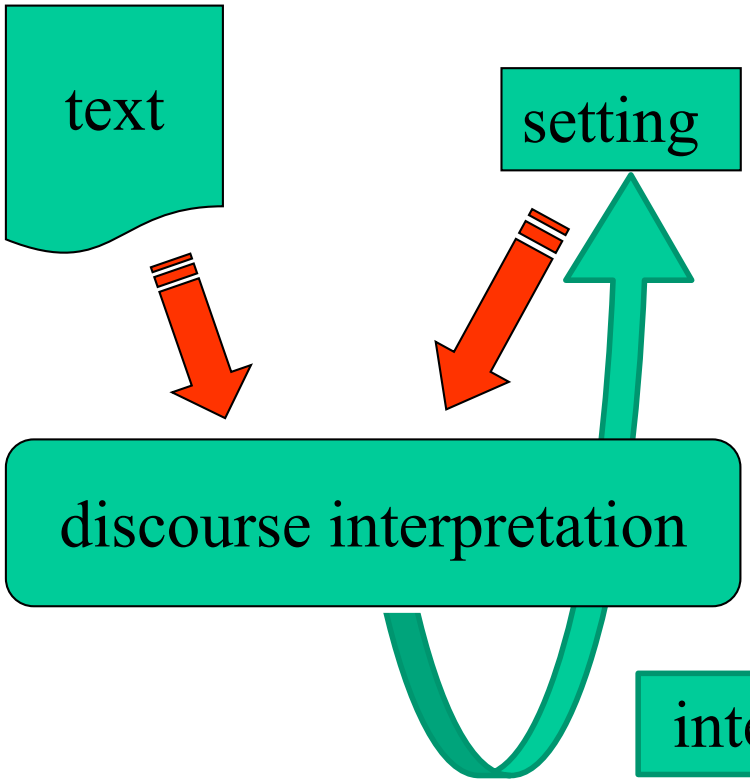
Setting: a dynamic system of conventions.

Cohesion and coherence

- **Cohesion**: usually enforced by anaphoric links, repetitions, etc. (see Haliday and Hassan, 1978)
- **Coherence**: rather easy to decide that a text is coherent, and very difficult to risk a statement of the contrary.

Recently, a friend of mine defied me that I am unable to give him a senseless sentence. So, I uttered the famous Chomskyan sentence "Colorless green ideas sleep furiously." challenging him to find a sense. And he did, because he explained me that this sentence simply says that one night some ideas (colorless, as all ideas) came, during an agitated sleep, to the mind of a politician, a member of the green party… So, the example argues for the necessity of a setting (or a context) according to which to give a meaning to a discourse. Often the key to the interpretation of a discourse comes from finding this setting. This is why to some people a novel like The sound and the fury of William Faulkner is obscure and difficult to read, while for others it is such a relish.

Interpretation of discourse



- knowledge about the language
- knowledge about the situation
- knowledge about the author
- knowledge about the world (real+virtual)

The setting explains why one sentence for different people could have different meanings: each one decodes it in a different setting (context).

What do we expect from a theory of discourse?

- To tell us how is the discourse structured
- To make explicit, using this structure, some discourse phenomena: at least cohesion and coherence
- To explain interruptions, flashbacks
- To explain how the structure can be built from the raw text
- To be easily put at the base of implementations that deal with discourse content

Centering - a theory of local discourse coherence

- Joshi,A.K. and Weinstein,S., 1981: “Control of Inference: Role of Some Aspects of Discourse-Structure Centering“
- Grosz,B.; Joshi,A.K. and Weinstein,S., 1986: “Towards a computational theory of discourse interpretation”
- **Brennan,S.E.; Friedman,M.W. and Pollard,C.J., 1987: “A Centering approach to pronouns“**
- **Grosz,B.; Joshi,A.K. and Weinstein,S, 1995: “Centering: A framework for modeling the local coherence of discourse”**
- Strube,M. and Hahn,U., 1996: “Functional Centering“
- Walker,M.A.; Joshi,A.K. and Prince,E.F. (eds.), 1997: “*Centering in Discourse*“
- Kameyama,M., 1997: “Intrasentential Centering: A Case Study“
- **Poesio, M., Stevenson, R., Di Eugenio, B. and Hitzeman, J. 2004: “Centering: A Parametric Theory and Its Instantiations”**

CT: goals of the theory

- explains why certain texts are more difficult to process than others
- explains why we use the pronouns the way we use them
- anchors a practical approach for anaphora resolution

A smooth discourse

from (Walker, Joshi and Prince, 1997)

a. *Jeff*$_1$ *helped* *Dick*$_2$ *wash the car.*

b. ***He***$_1$ ***washed*** *the windows as* *Dick*$_2$ *waxed the car.*

c. ***He***$_1$ ***soaped*** *a pane.*

He in c. is *Jeff* because **soaping** can only be related to the **washing** event

Se începe prin a se vorbi despre Jeff în (a), în (b) se continuă despre el, iar în (c) constrângerea semantică dictează că e vorba de același Jeff. Procesarea se face lin.

A less smooth discourse

a. *Jeff*$_1$ *helped* *Dick*$_2$ *wash the car.*

b. ***He***$_1$ ***washed*** *the windows as* *Dick*$_2$ *waxed the car.*

c. ***He***$_2$ ***buffed*** *the hood.*

He in c. is *Dick*, because **buffing** can only be related to the **waxing** event.

CT pretinde că acest discurs este mai greu de procesat decât precedentul. O inferență de natură semantică dictează că în (c) *He* se leagă la Dick iar nu la Jeff. Aici apare o comutare a atenției de la Jeff, ce fusese în centrul atenției în primele două fraze, la Dick despre care se vorbește în (c). O comutare a centrului atenției este mai greu de procesat decât o păstrare a lui.

CT: focus helps to disambiguate pronominal anaphora

from (Grosz, Joshi and Weinstein, 1995)

1. *Susan$_1$ is a fine friend.*
2. *She$_1$ gives people the most wonderful presents.*
3. *She$_1$ just gave Betsy$_2$ a wonderful bottle of wine.*
4. *She$_1$ told her$_2$ it was quite rare.*
5. *She$_1$ knows a lot about wine.*

There is no problem with finding who she is in 2 and 3 because no referent other than Susan were introduced by the discourse. 3 introduces Betsy, also a female character. Neither syntactic nor semantic criteria can be applied in 4 to decide who the referent of she and respectively her are. Still we easily link she to Susan and her to Betsy because the focus of the preceding utterance was Susan and it is normal to consider that the focus is preserved. A similar criteria works in 5: more people are inclined to consider she here as being Susan than being Betsy.

CT explains the normality or oddness of some utterances

from (Grosz, Joshi and Weinstein, 1995)

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

"CT investighează interacţiunile ce pot fi stabilite între alegerea expresiilor referențiale, starea atențională, inferențele necesare pentru determinarea interpretărilor unei exprimări într-un segment de discurs și coerență. Pronumele şi descrierile definite nu sunt echivalente relativ la efectul pe care îl au asupra coerenței." (citat din GJW95).

This text has a defect…

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

e. *He was sick and furious at being woken up so early.*

După ce pînă la unitatea (d) focusul a fost menţinut constant – **Terry**, abia în momentul în care *sick* apare cititorul realizează că *he* în (e) nu mai e **Terry** ci **Tony**, pentru că **Terry** n-ar fi putut fi bolnav. Pentru un moment cititorul a fost derutat. El a avut nevoie să facă o inferenţă pentru a repara confuzia.

This text has a defect…

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

e. ***He*** *was sick and furious at being woken up so early.*

This text repairs the proceeding

a. *Terry really goofs sometimes.*
b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*
c. *He wanted Tony to join on a sailing expedition.*
d. *He called him at 6 A.M.*
e. ***Tony*** *was sick and furious at being woken up so early.*

O secvență mai naturală ar fi fost aceea în care în (e) s-ar fi folosit direct *Tony.*

... And we can go on like this:

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

e. *Tony was sick and furious at being woken up so early.*

f. *He told Terry to get lost and hung up.*

But here we have again a problem...

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

e. *Tony was sick and furious at being woken up so early.*

f. *He told Terry to get lost and hung up.*

g. *Of course, he hadn't intended to upset Tony.*

… dar cititorul este din nou derutat în unitatea (*g*). Focusul fusese schimbat din **Terry** în **Tony** în unitățile (*e*) și (*f*) și așteptarea era ca același personaj să rămână în continuare în centrul atenției.

But here we have again a problem...

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

e. *Tony was sick and furious at being woken up so early.*

f. *He told Terry to get lost and hung up.*

g. *Of, course **he** hadn't intended to upset Tony.*

Finally, we are done!

a. *Terry really goofs sometimes.*

b. *Yesterday was a beautiful day and he was excited about trying out his new sailboat.*

c. *He wanted Tony to join on a sailing expedition.*

d. *He called him at 6 A.M.*

e. *Tony was sick and furious at being woken up so early.*

f. *He told Terry to get lost and hung up.*

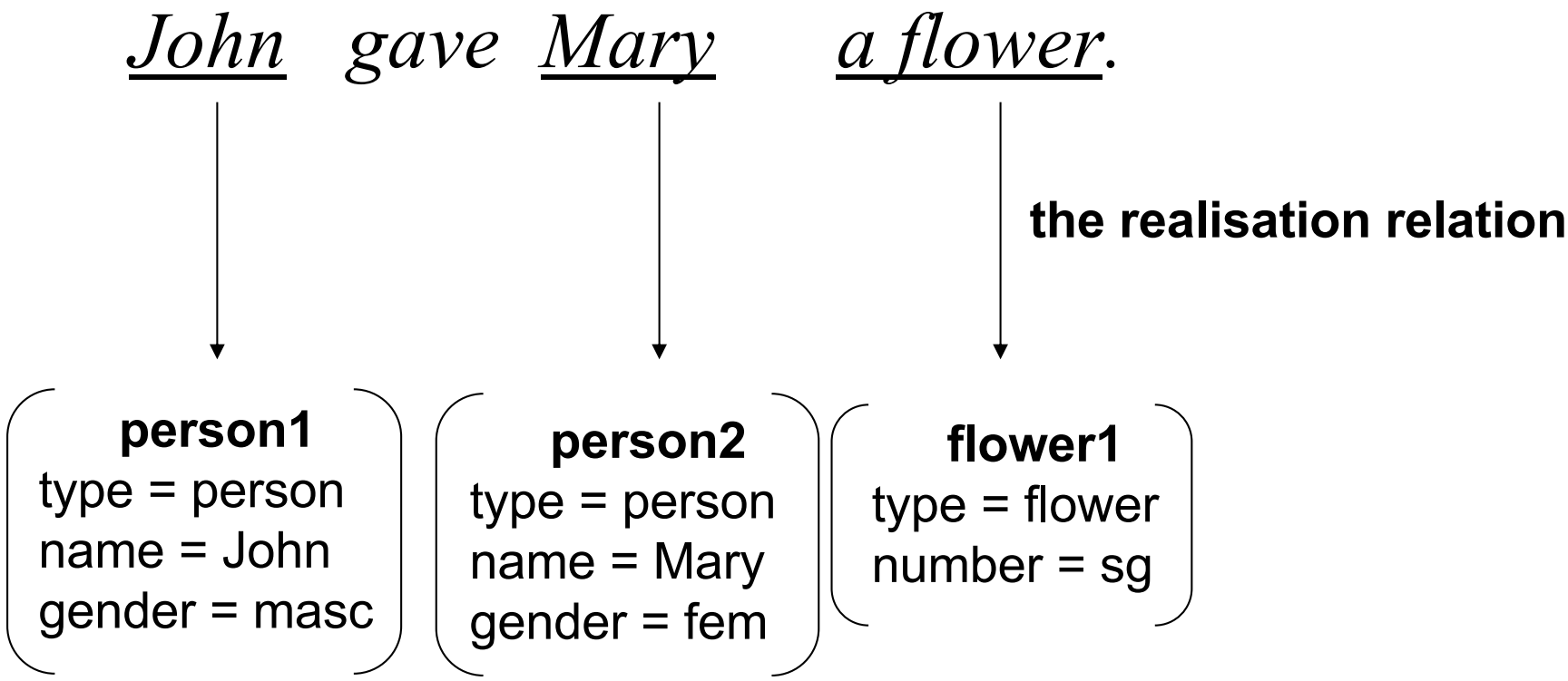
g. *Of, course **Terry** hadn't intended to upset Tony.*

În sfârșit deruta este eliminată în această variantă.
Conjectura CT este că forma de expresie într-un discurs influențează direct cererile de resurse utilizate la descifrarea lui. Este cunoscut că identificarea referenților grupurilor nominale într-un discurs presupune un anumit proces inferențial. CT afirmă că forma în care sunt exprimate aceste grupuri nominale poate introduce o **încărcare inferențială** mai mare ori mai mică în cititor.

CT: the main lines

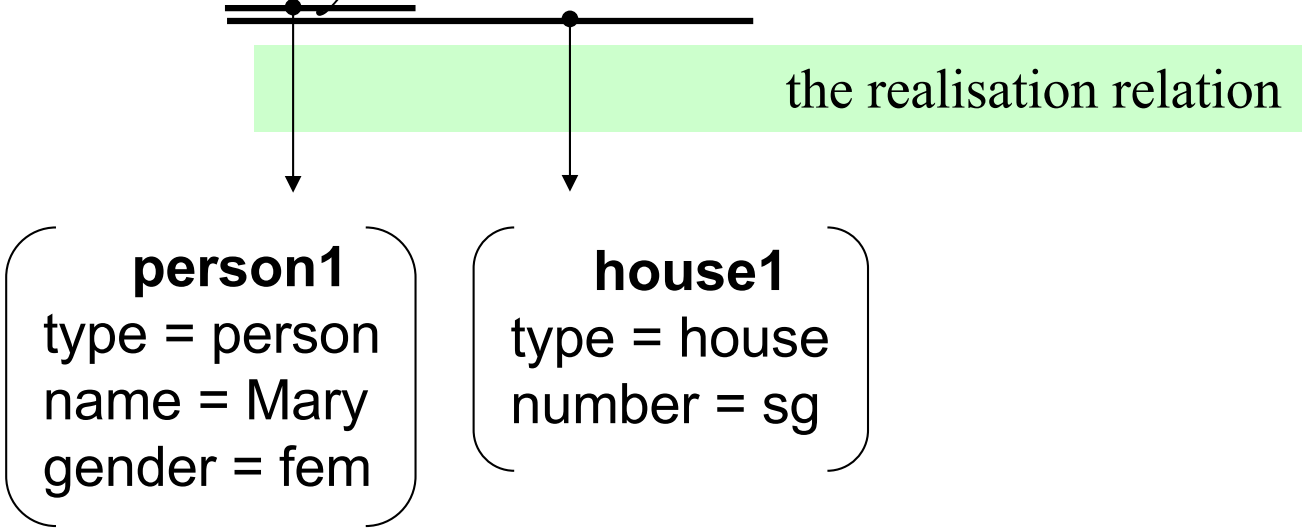
- Applies to just one segment of discourse
  - refers to Grosz&Sidner‘s Attentional State Theory
- Sees the segment drawn up of adjacent utterances (sentences)
- Discourse entities are called **centers**

Centers


John gave Mary a flower.
the realisation relation
person1
type = person
name = John
gender = masc
person2
type = person
name = Mary
gender = fem
flower1
type = flower
number = sg

What is a center?

John went to Mary's house



He met her down the street.

He found it down the street.

For each utterance, compute:

- a list of forward-looking centers:

  $C_f(u_i) = \langle e_1, e_2, \ldots e_k \rangle$

  ranking: subject > direct-object > indirect-object > others

- a backward-looking center:

  $C_b(u_i)$ = highest-ranked element of $C_f(u_{i-1})$ that is realised in u_i

- a prefered center:

  $C_p(u_i) = e_1$

Rule 1: pronoun realisation

- **If some element of $C_f(u_{i-1})$ is realised as a pronoun in u_i, than so is $C_b(u_i)$**
  - it captures the intuition that pronominalisation is one way to indicate discourse salience
  - if there are multiple pronouns in a sentence realising discourse entities from the previous utterance, than C_b must be one of them
  - if there is just one pronoun, then the pronoun must be the C_b

Rule 1 observed

a. *Terry really goofs sometimes.*

C_f = ([**Terry**])

b. *Yesterday was a beatiful day and he was excited about trying out his new sailboat.*

C_f = (*he*=*his*=[**Terry**], [**the sailboat**])

C_b = [**Terry**]

c. *He wanted Tony to join on a sailing expedition.*

C_f = (*he*=[**Terry**], [**Tony**], [**the expedition**])

C_b = [**Terry**]

d. *He called him at 6 A.M.*

C_f = (*he*=[**Terry**], *him*=[**Tony**])

C_b = [**Terry**]

Rule 1 still observed

a. *Terry really goofs sometimes.*

C_f = ([**Terry**])

b. *Yesterday was a beatiful day and he was excited about trying out his new sailboat.*

C_f = (*he*=*his*=[**Terry**], [**the sailboat**])

C_b = [**Terry**]

c. *He wanted Tony to join on a sailing expedition.*

C_f = (*he*=[**Terry**], [**Tony**], [**the expedition**])

C_b = [**Terry**]

d. *He called Tony at 6 A.M.*

C_f = (*he*=[**Terry**], [**Tony**])

C_b = [**Terry**]

Rule 1 disobserved

a. *Terry really goofs sometimes.*

C_f = ([**Terry**])

b. *Yesterday was a beatiful day and he was excited about trying out his new sailboat.*

C_f = (*he*=*his*=[**Terry**], [**the sailboat**])

C_b = [**Terry**]

c. *He wanted Tony to join on a sailing expedition.*

C_f = (*he*=[**Terry**], [**Tony**], [**the expedition**])

C_b = [**Terry**]

d. *Terry called him at 6 A.M.*

C_f = ([**Terry**], *him*=[**Tony**])

C_b = [**Terry**]

Rule 2: transitions

Following (Grosz, Joshi and Weinstein, 1995)
and (Brennan, Friedman and Pollard, 1987)

	$C_b(u) = C_b(u\text{-}1)$	$C_b(u) \neq C_b(u\text{-}1)$
$C_b(u) = C_p(u)$	***CONTINUING***	***SMOOTH SHIFT***
$C_b(u) \neq C_p(u)$	***RETAINING***	***ABRUPT SHIFT***

CON > RET > SSH > ASH

Rule 2: a point of discontinuity

a. *Terry really goofs sometimes.*

C_f = ([**Terry**])

CONTINUING

b. *Yesterday was a beatiful day and he was excited about trying out his new sailboat.*

C_f = (*he*=*his*=[**Terry**], [**the sailboat**])

C_b = [**Terry**]

CONTINUING

c. *He wanted Tony to join on a sailing expedition.*

C_f = (*he*=[**Terry**], [**Tony**], [**the expedition**])

C_b = [**Terry**]

CONTINUING

d. *He called Tony at 6 A.M.*

C_f = (*he*=[**Terry**], [**Tony**])

C_b = [**Terry**]

SMOOTH SHIFT

e. *Tony was sick and furious at being woken up so early.*

C_f = ([**Tony**])

C_b = [**Tony**]

Rule 2: further analysis

...

d. *He called Tony at 6 A.M.*

C_f = (*he*=[**Terry**], [**Tony**])

C_b = [**Terry**]

CONTINUING

e. *Tony was sick and furious at being woken up so early.*

C_f = ([**Tony**])

C_b = [**Tony**]

SMOOTH SHIFT

f. *He told Terry to get lost and hung up.*

C_f = (*he*=[**Tony**], [**Terry**])

C_b = [**Tony**]

CONTINUING

g. *Of, course Terry hadn't intended to upset Tony.*

C_f = ([**Terry**], [**Tony**])

C_b = [**Tony**]

RETAINING

Centering hints on pronominal anaphora

a. *I haven't seen Jeff for several days.*

C_f = (*I*=[**I**], [**Jeff**])
C_b = [**I**]

b. *Carl thinks he's studying for his exams.*

C_f = ([**Carl**], *he*=[**Jeff**], [**Jeff´s exams**])
C_b = [**Jeff**]

c. *I think he? went to the Cape with Linda.*

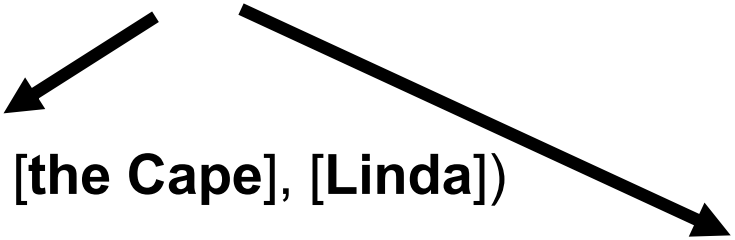
from (Grosz, Joshi & Weinstein, 1983)

Centering explains why we understand *he* in unit c as Jeff

b. *Carl thinks he's studying for his exams.*

C_f = ([**Carl**], *he*=[**Jeff**], [**Jeff´s exams**])
C_b = [**Jeff**]

c. *I think he? went to the Cape with Linda.*



C_f = (*I*=[**I**], *he*=[**Jeff**], [**the Cape**], [**Linda**])
C_b = [**Jeff**]
RETAINING

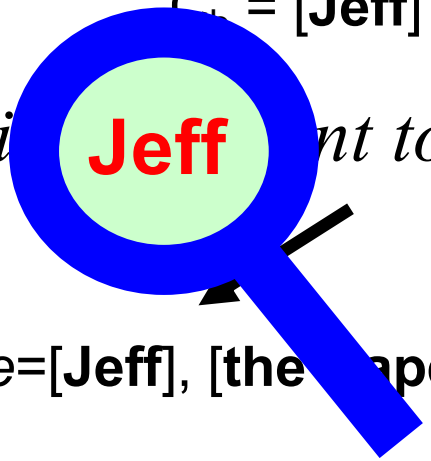
C_f = (*I*=[**I**], *he*=[**Carl**], [**the Cape**], [**Linda**])
C_b = [**Carl**]
ABRUPT SHIFT

Centering explains why we understand *he* in unit c as Jeff

b. *Carl thinks he's studying for his exams.*

C_f = ([**Carl**], *he*=[**Jeff**], [**Jeff´s exams**])
C_b = [**Jeff**]

c. *I thi nt to the Cape with Linda.*



C_f = (*I*=[**I**], *he*=[**Jeff**], [**the pe**], [**Linda**])
C_b = [**Jeff**]
RETAINING

C_f = (*I*=[**I**], *he*=[**Carl**], [**the Cape**], [**Linda**])
C_b = [**Carl**]
ABRUPT SHIFT

Attentional state theory (AST)

(Barbara Grosz & Candence Sidner, 1987)

Models the linguistic structure of the discourse

Gives an account on intentions and how they are combined

Explains the shift of attention during discourse interpretation

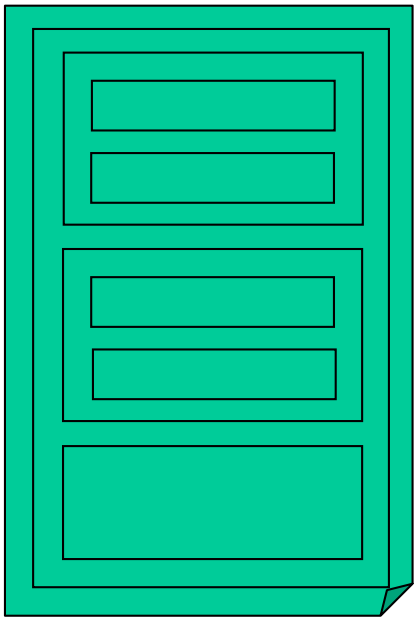
Explains interruptions and flash-backs

Puts in evidence a dynamic domain of referentiality

3 components

AST: 1st component

- a linguistic structure:
  - more sentences are aggregated in the same segment
  - segments display a recursive structure



What locality means in CT view?

- CT acceptă cadrul teoretic al teoriei AST care explică discursul la nivel global. CT a fost elaborată pentru a explica ce se întîmplă în interiorul unui segment. Asta înseamnă că ea va fi în stare să explice de ce anumite tranziţii între unităţile unui segment sînt mai uşor de procesat decît altele. Dar, după cum ştim de la AST, în interiorul unui segment putem regăsi relaţii de dominare, ceea ce înseamnă că el are în componenţă alte subsegmente. Definiţia recursivă a segmentului ridică însă un semn de întrebare relativ la interpretarea restricţiei de localitate în care a fost definită CT: atunci cînd, trecem graniţa dintre ultima unitate a segmentului B şi prima unitate a segmentului C, aflate între ele în relaţie de satisfacere-precedenţă, atunci cînd amîndouă sînt dominate de un segment mai cuprinzător A, pe de o parte trecem graniţa unui segment deci nu putem aplica CT, dar pe altă parte rămînem în interiorul segmentului A, deci ar trebui să putem aplica CT…

Centering: other problems?

- still a local theory (applies inside a segment)
- ranking of C_f ellements
  - on what criteria (surface-order, syntactic role, functional (Strube&Hahn, 1996)) ➔ language dependant
- null pronouns
  - Italian, Japanese, Romanian
- clitics: doubling references
  - Romanian
- intrasentensial centering (Kameyama, 1997)

Centering - a parametric theory (Poesio et al, 2004)