

Interazione Uomo - Macchina II:

Laboratorio di Interfacce Intelligenti

Fiorella de Rosis

Esercitazione 2

Generazione di linguaggio naturale (NLG)

Tutor esercitazione: Irene Mazzotta

Prerequisiti

Concetti di base su **Teorie** e **Metodi** per la Generazione dei messaggi in Linguaggio Naturale

Materiale

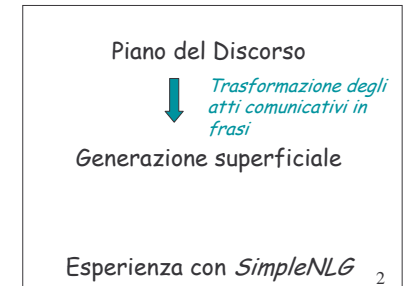
- Dispense del corso, Unità 3 e 4 (reperibili sul sito web)

Obiettivo

Esercitazione 2



Esercitazione 4



Esercitazione 2: Analisi della struttura di un testo e traduzione in piano del discorso

Obiettivi

- Decomposizione del testo in Discourse Segment (DS) e definizione della *Struttura Intenzionale del Discorso*.
- Definizione della *Struttura Retorica del testo* mediante l'individuazione delle relazioni tra i vari DSs.
- Traduzione delle Struttura Retorica del testo in un *Piano del Discorso* (DP).
- Traduzione del DP in una struttura XML rappresentata con il linguaggio di markup *DPML*
- Outline di un algoritmo di esplorazione del DP per la generazione del messaggio

Esempio che svilupperemo: Le previsioni del tempo - da Repubblica.it -

- Previsioni giornaliere
- Con diversi livelli di dettaglio
- Testi con strutture simili

... *tipico caso in cui la
generazione automatica è
giustificata*

N.B.: Il sito web delle previsioni meteo di Repubblica.it ha cambiato look... ma la struttura dei testi è rimasta la stessa!

Le previsioni

lunedì 26 dicembre

Nord: molto nuvoloso o coperto sulla Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e Basso Veneto con precipitazioni sparse, occasionalmente nevose anche in pianura, specie nella seconda parte della giornata. Nuvolosità più attenuata sulle rimanenti zone del Nord con addensamenti più consistenti sulle zone alpine, specie sui versanti settentrionali ove potranno verificarsi delle nevicate.

Centro e Sardegna: molto nuvoloso o coperto, con piogge diffuse. Nevicate sulle zone interne a quote collinari specie su Toscana, Marche ed Umbria.

Nuvolosità e fenomeni più attenuati sulla Sardegna.
Sud e Sicilia: generalmente molto nuvoloso a tratti coperto con precipitazioni diffuse e persistenti su Campania, Calabria Tirrenica e Sicilia settentrionale. Neve sui rilievi interni oltre gli 800-1.000 metri di quota. Molto nuvoloso ma con fenomeni isolati sul resto del Sud, pur con tendenza al peggioramento.
Temperature: in generale diminuzione, più marcata sulle regioni settentrionali.

Venti: deboli o moderati settentrionali al Nord; moderati da Ovest-Sud Ovest al Centro ed al Sud, con rinforzi sulla Sardegna, sulle coste joniche e sul Salento.

Mari: da poco mossi a mossi il Mar Ligure e l'Alto Adriatico; mossi o molto mossi tutti gli altri mari.

Passi nello svolgimento della traccia

1. Definizione della Struttura Intenzionale del testo
 - 1.1 Decomposizione del testo in Segmenti del discorso (DS)
 - 1.2 Analisi della Struttura Intenzionale

Ma un messaggio non è una semplice lista di segmenti.

L'insieme delle relazioni fra i segmenti ne assicura la coerenza.

2. Definizione della Struttura Retorica (RS) del testo
 - 1.1 Analisi delle relazioni retoriche tra i DSs
 - 1.2 Rappresentazione della RS

Abbiamo tutti gli strumenti per generare il Piano del Discorso (DP): il DP, infatti, riflette la struttura intenzionale e retorica del testo

3. Rappresentazione del DP mediante tree structure
 - 3.1 Traduzione dei DSs in goal o sottogoal e (a livello delle foglie) in atti comunicativi
 - 3.2 Rappresentazione mediante tree structure della RS
4. Traduzione della struttura ad albero in DPML

Metodo adottato:

Decomposizione del problema in sottoproblemi più semplici, fino a ridurlo ad una lista di problemi elementari affrontabili con un solo operatore.

5

Struttura del testo:

Data
 Descrizione per aree geografiche
 Nord
 Centro e Sardegna
 Sud e Sicilia
 Temperature
 Venti
 Mari

In ogni area geografica:

Descrizione, 'per zone omogenee':

Nuvolosità
come: 'molto nuvoloso o coperto'
dove: 'sulla...'
 Precipitazioni
 Nevicate

Le previsioni

lunedì 26 dicembre

Nord: molto nuvoloso o coperto sulla Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e Basso Vento con precipitazioni sparse, occasionalmente nevose anche in pianura, specie nella seconda parte della giornata. Nuvolosità più attenuata sulle rimanenti zone del Nord con addensamenti più consistenti sulle zone alpine, specie sui versanti settentrionali ove potranno verificarsi delle neviccate.

Centro e Sardegna: molto nuvoloso o coperto, con piogge diffuse. Nevicate sulle zone interne a quote collinari specie su Toscana, Marche ed Umbria.

Sud e Sicilia: generalmente molto nuvoloso a tratti coperto con precipitazioni diffuse e persistenti su Campania, Calabria Tirrenica e Sicilia settentrionale.

Neve sui rilievi interni oltre gli 800-1.000 metri di quota. Molto nuvoloso ma con fenomeni isolati sul resto del Sud, pur con tendenza al peggioramento.

Temperature: in generale diminuzione, più marcata sulle regioni settentrionali.

Venti: deboli o moderati settentrionali al Nord; moderati da Ovest-Sud Ovest al Centro ed al Sud, con rinforzi sulla Sardegna, sulle coste joniche e sul Salento.

Mari: da poco mossi a mossi il Mar Ligure e l'Alto Adriatico; mossi o molto mossi tutti gli altri mari.

6

Struttura del testo:

DS0 Previsioni Meteo
 DS1 Data
 DS2 Descrizione per aree geogr
 DS2.1 Nord
 DS2.2 Centro e Sardegna
 DS2.3 Sud e Sicilia
 DS3 Temperature
 DS4 Venti
 DS5 Mari

In ogni area geografica:

Descrizione, 'per zone omogenee':

DS2.1 Nord
 DS2.1.1 Zona Omogenea1
 DS2.1.2 Zona Omogenea2
 DS2.2 Centro e Sardegna
 DS2.2.1 Zona Omogenea1
 DS2.2.2 Zona Omogenea2
 DS2.3 Sud e Sicilia
 DS2.3.1 Zona Omogenea1

Struttura del testo:

DS0 Previsioni Meteo
 DS1 Data
 DS2 Descrizione per aree geogr
 DS2.1 Nord
 DS2.2 Centro e Sardegna
 DS2.3 Sud e Sicilia
 DS3 Temperature
 DS4 Venti
 DS5 Mari

In ogni area geografica:

Descrizione, 'per zone omogenee':

DS2.1 Nord
 DS2.1.1 Zona Omogenea1
 DS2.1.2 Zona Omogenea2
 DS2.2 Centro e Sardegna
 DS2.2.1 Zona Omogenea1
 DS2.2.2 Zona Omogenea2
 DS2.3 Sud e Sicilia
 DS2.3.1 Zona Omogenea1

In ogni zona omogenea:

DS2.1.1 Zona Omogenea1
 DS2.1.1.1 Fenomeno1
 DS2.1.1.2 Fenomeno2
 DS2.1.1.3 Fenomeno3

Nel nostro esempio

In ogni zona omogenea:

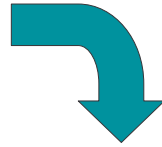
DS2.1.1 Zone Nuvolose
 DS2.1.1.1 nuvolosità
 DS2.1.1.2 precipitazioni
 DS2.1.1.3 neviccate

7

8

In ogni zona omogenea:

- DS2.1.1 Zone Nuvolose
 - DS2.1.1.1 nuvolosità
 - DS2.1.1.2 precipitazioni
 - DS2.1.1.3 neviccate

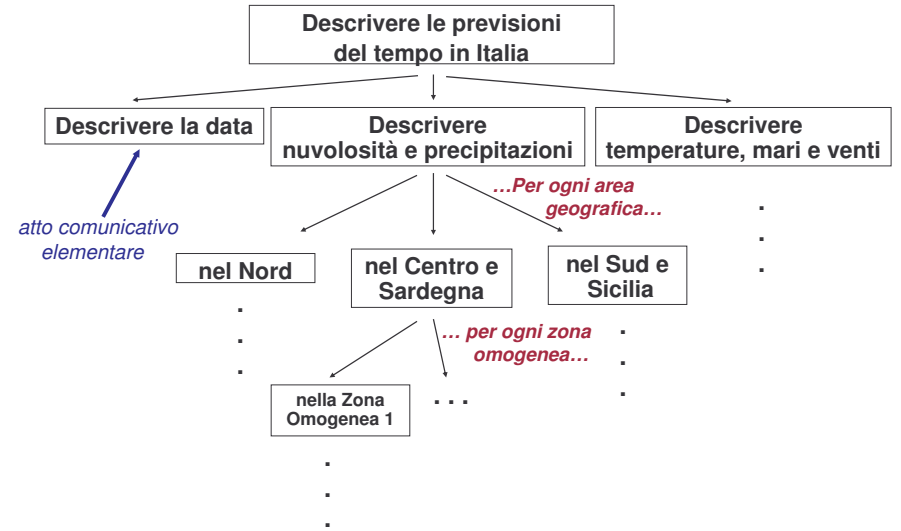


Specifichiamo i DSs

In ogni descrizione del fenomeno:

- DS2.1.1 Zone molto nuvolose
 - DS2.1.1.1 Molto nuvoloso o coperto su Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e basso Veneto
 - DS2.1.1.2 con precipitazioni sparse
 - DS2.1.1.3
 - DS2.1.1.3.1 occasionalmente nevose
 - DS2.1.1.3.2 anche in pianura
 - DS2.1.1.3.3 specie nella seconda parte della giornata

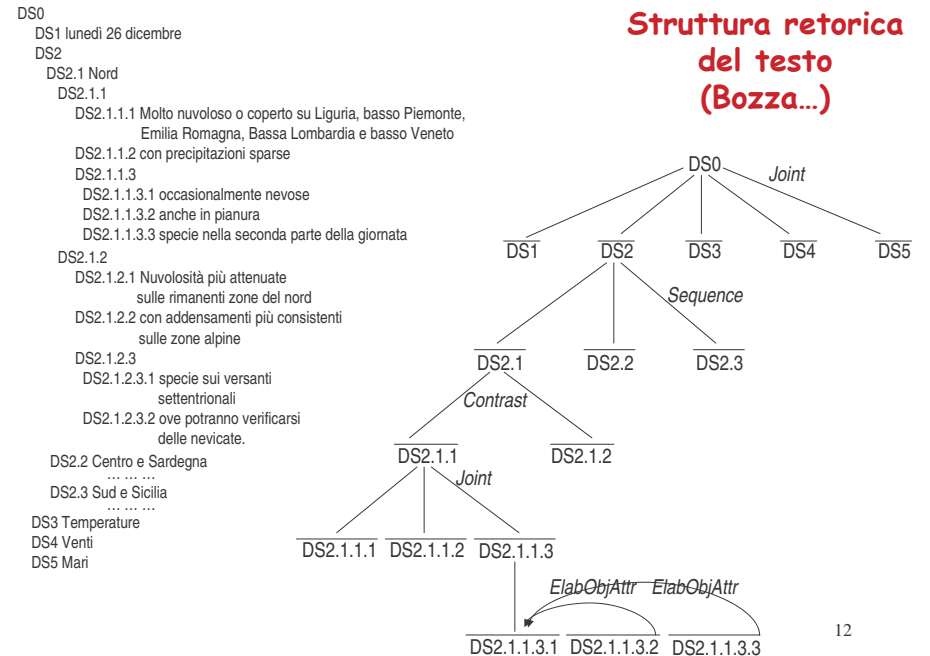
Outline del piano:



Struttura intenzionale del testo

- DS0 <----- Previsioni meteo
- DS1 lunedì 26 dicembre
- DS2 <----- Descrizione per aree geografiche
 - DS2.1 Nord <----- Zone molto nuvolose
 - DS2.1.1.1 Molto nuvoloso o coperto su Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e basso Veneto
 - DS2.1.1.2 con precipitazioni sparse
 - DS2.1.1.3
 - DS2.1.1.3.1 occasionalmente nevose
 - DS2.1.1.3.2 anche in pianura
 - DS2.1.1.3.3 specie nella seconda parte della giornata
 - DS2.1.2 <----- Zone meno nuvolose
 - DS2.1.2.1 Nuvolosità più attenuate sulle rimanenti zone del nord
 - DS2.1.2.2 con addensamenti più consistenti sulle zone alpine
 - DS2.1.2.3
 - DS2.1.2.3.1 specie sui versanti settentrionali
 - DS2.1.2.3.2 ove potranno verificarsi delle neviccate.
- DS2.2 Centro e Sardegna
- DS2.3 Sud e Sicilia
- DS3 Temperature
- DS4 Venti
- DS5 Mari

Struttura retorica del testo (Bozza...)

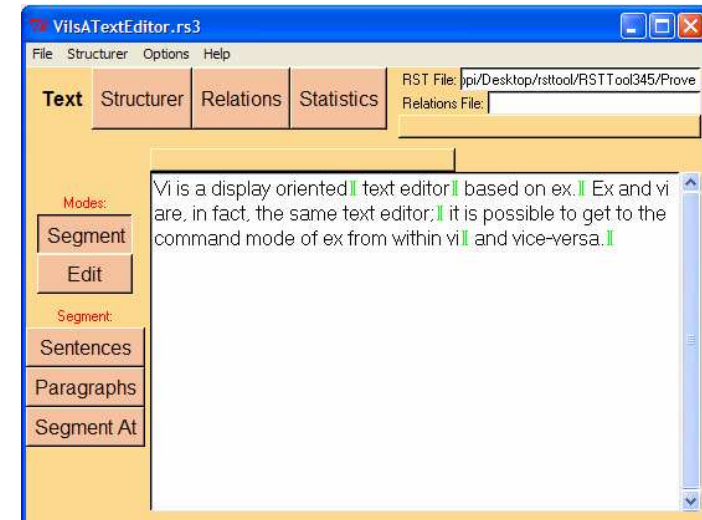


Impariamo ad usare RSTTool <http://www.wagsoft.com/RSTTool/>

1. Creiamo o importiamo il file .txt contenente il testo da analizzare
2. Segmentiamo il testo in frasi
3. Stabiliamo le relazioni fra le parti del discorso e, se necessario, aggiorniamo la lista di RR.
5. Facciamo un quadro riassuntivo delle RR nel nostro testo

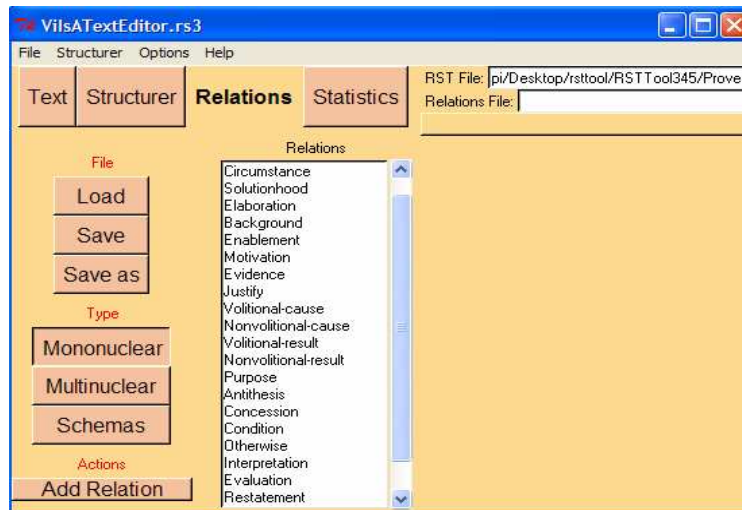
13

Segmentiamo il testo in frasi



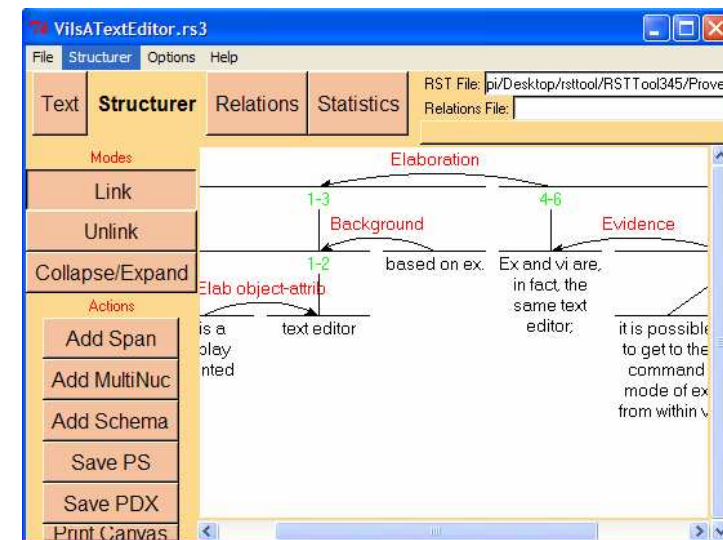
14

Aggiorniamo la lista di RR



15

Stabiliamo le relazioni fra le parti del discorso



16

Facciamo un quadro riassuntivo delle RR nel nostro testo

relation	N	Mean	S^N:1
Total Relations: 6 (RST Only, Counting Multinuclear Nodes)			
Background	1	16.7%	
Elab object-attrib	1	16.7%	
Elaboration	1	16.7%	
Evidence	1	16.7%	
Joint	1	16.7%	
top	1	16.7%	

17

Proviamo ad applicare RSTTool

- L'esempio del messaggio di help
- L'esempio del messaggio persuasivo
- L'esempio del messaggio di help in word
- L'esempio delle previsioni del tempo

18

L'esempio del messaggio di help

Vi is a display oriented text editor based on ex. Ex and vi are, in fact, the same text editor; it is possible to get to the command mode of ex from within vi and vice-versa.

La Struttura Intenzionale

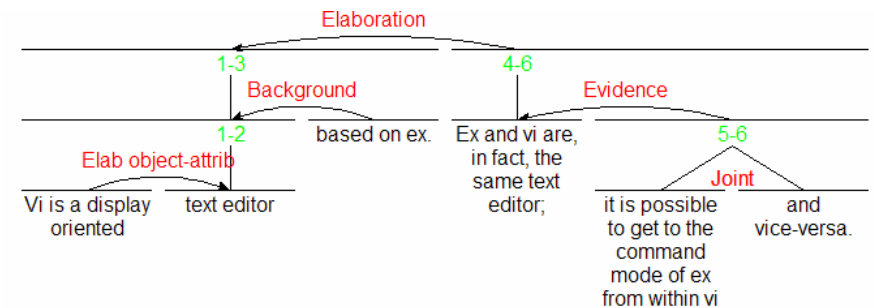
- DS1
- DS1.1 *Vi is a text editor* ← Parte principale del testo
 - DS1.2
 - DS1.2.1 *display oriented* ← Specifica le caratteristiche di text editor
 - DS1.2.2 *based on ex.*
- DS2
- DS2.1 *Ex and vi are, in fact, the same text editor;*
 - DS2.2
 - DS2.2.1 *It is possible to get to the command mode of ex from within vi*
 - DS2.2.2 *and vice-versa*

Descrive la relazione fra Vi ed Ex e la giustificazione

19

L'esempio del messaggio di help

La Struttura Retorica



20

L'esempio del messaggio persuasivo

- Migliore postura

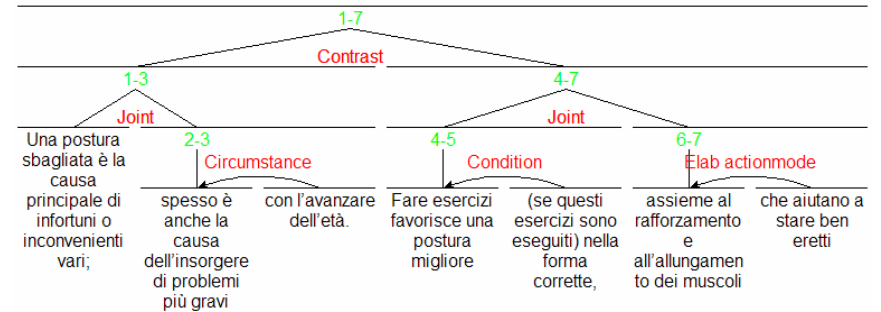
Una postura sbagliata è la causa principale di infortuni o inconvenienti vari; spesso è anche la causa dell'insorgere di problemi più gravi con l'avanzare dell'età. Fare esercizi nella forma corretta favorisce una postura migliore, assieme al rafforzamento e all'allungamento dei muscoli che aiutano a stare ben eretti

----- La Struttura Intenzionale -----

DS0			
DS1	Migliore postura	<-----	Titolo
DS2		<-----	Parte persuasiva
DS3		<-----	Effetti negativi
DS3.1	Una postura sbagliata è la causa principale di infortuni o inconvenienti vari		
DS3.2			
DS3.2.1	Spesso è anche la causa dell'insorgere di problemi più gravi		
DS3.2.2	con l'avanzare dell'età		
DS4			
DS4.1		<-----	Effetti positivi
DS4.1.1	Fare esercizi favorisce una postura migliore		
DS4.1.2	(se questi esercizi sono eseguiti) nella forma corretta		
DS4.2			
DS4.2.1	assieme al rafforzamento e all'allungamento dei muscoli		
DS4.2.2	che aiutano a stare ben eretti		

L'esempio del messaggio persuasivo

----- La Struttura Retorica -----



L'esempio del messaggio di help in word

Scegliere Guida in linea Microsoft Word dal menu ?.

Se è abilitato, verrà visualizzato l'Assistente.

Se l'Assistente è disabilitato, verrà visualizzata la finestra della Guida.

Per digitare una domanda nella finestra della Guida, fare clic sulla scheda Ricerca libera.

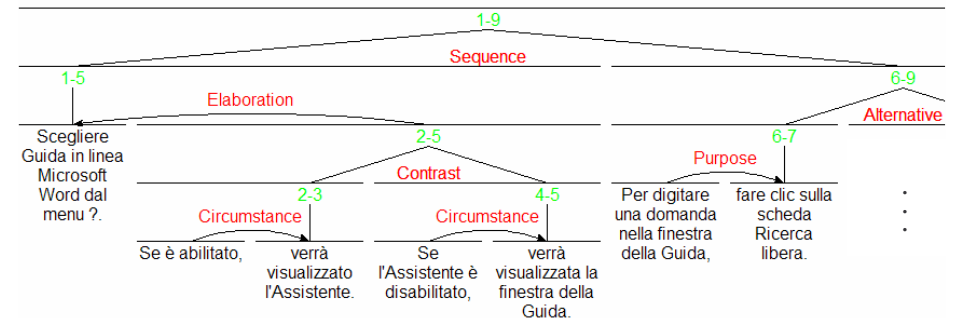
Per sfogliare il sommario della Guida, fare clic sulla scheda Sommario. ...

----- La Struttura Intenzionale -----

DS0		<-----	Come usare l'Assistente
DS1		<-----	Inizializzare l'assistente
DS1.1	Scegliere Guida in linea Microsoft Word dal menu ?.		
DS1.2			
DS1.2.1	Se è abilitato, verrà visualizzato l'Assistente.		
DS1.2.2	Se l'Assistente è disabilitato, verrà visualizzata la finestra della Guida.		
DS2		<-----	Scegliere tra i vari task
DS2.1			
DS2.1.1	Per digitare una domanda nella finestra della Guida,		
DS2.1.2	fare clic sulla scheda Ricerca libera.		
DS2.2			
DS2.2.1	Per sfogliare il sommario della Guida,		
DS2.2.2	fare clic sulla scheda Sommario. ...		

L'esempio del messaggio di help in word

----- La Struttura Retorica -----



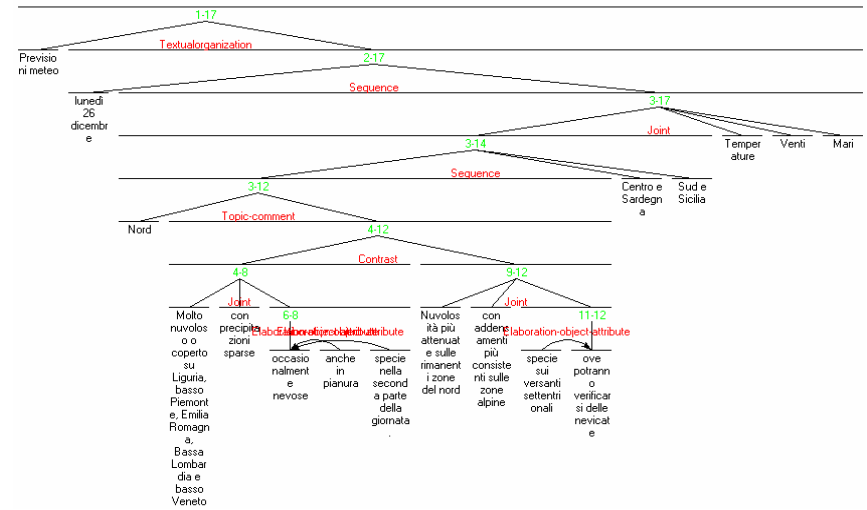
L'esempio delle previsioni meteo

La Struttura Intenzionale

- DS0 Previsioni Meteo
- DS1 lunedì 26 dicembre
- DS2
 - DS2.1 Nord
 - DS2.1.1
 - DS2.1.1.1 Molto nuvoloso o coperto su Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e basso Veneto
 - DS2.1.1.2 con precipitazioni sparse
 - DS2.1.1.3
 - DS2.1.1.3.1 occasionalmente nevose
 - DS2.1.1.3.2 anche in pianura
 - DS2.1.1.3.3 specie nella seconda parte della giornata
 - DS2.1.2
 - DS2.1.2.1 Nuvolosità più attenuate sulle rimanenti zone del nord
 - DS2.1.2.2 con addensamenti più consistenti sulle zone alpine
 - DS2.1.2.3
 - DS2.1.2.3.1 specie sui versanti settentrionali
 - DS2.1.2.3.2 ove potranno verificarsi delle nevicate.
 - DS2.2 Centro e Sardegna
 - DS2.3 Sud e Sicilia
 - DS3 Temperature
 - DS4 Venti
 - DS5 Mari

L'esempio delle previsioni del tempo

La Struttura Retorica



Most Nuclear Part (MNP) - L' "essenza" del discorso -

La MNP è l'insieme dei DSs che si ottiene esplorando l'albero RST del testo (T) a partire dalla radice e 'potando', ad ogni passo, i nodi figlio che corrispondono a 'satelliti' della RR.

MNP(T)

- se la radice di T è una foglia allora leggi il DS associato al nodo altrimenti
- se la radice di T è un Nucleo allora esegui MNP su ogni sottoalbero in ordine da sx a dx.

L'applicazione di un algoritmo di ricerca della most nuclear part genera automaticamente una sintesi del testo.

Uno Speudo-algoritmo ricorsivo per la ricerca della MNP

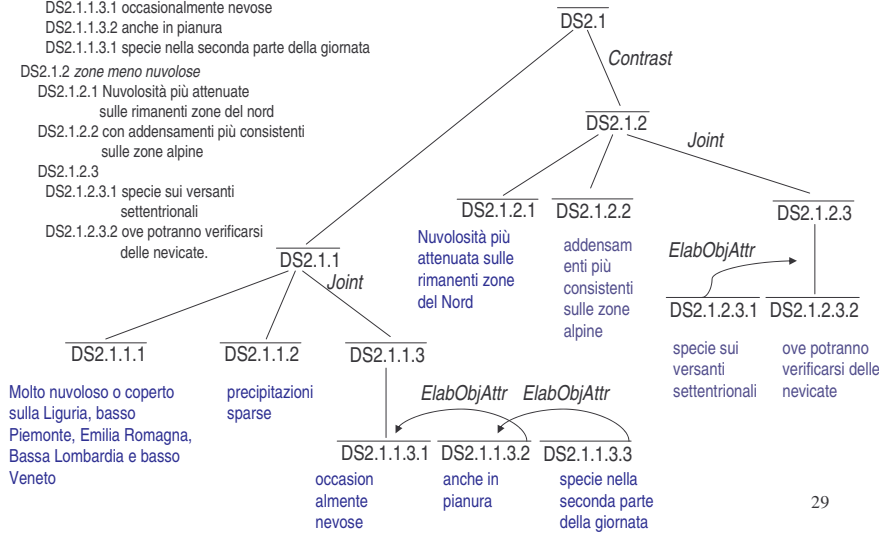
Siano T: RST tree; u, c: nodes;

```

MNP (T, u)
begin
  if Foglia(u, T) then {leggi il DS associato al nodo}
  else begin
    if RR_Nodo(u, T) = nucleo then begin
      c:= Primogenito (u,T);
      MNP(T, c);
      while not (UltimoFratello(c,T) do begin
        c:= SuccessivoFratello(c,T);
        MNP(T,c)
      end
    end
  end
end
end.
    
```

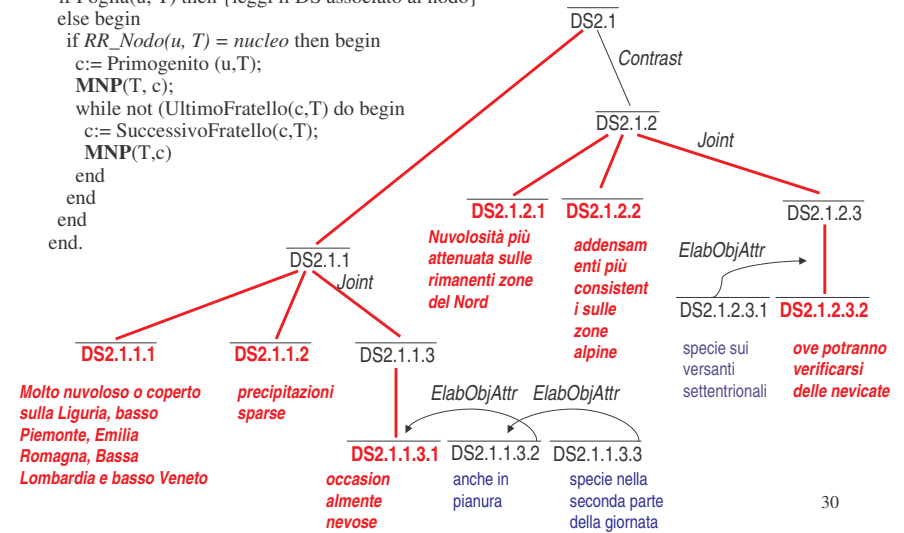
DS2.1 Nord
 DS2.1.1 zone molto nuvolose
 DS2.1.1.1 Molto nuvoloso o coperto su Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e basso Veneto
 DS2.1.1.2 con precipitazioni sparse
 DS2.1.1.3
 DS2.1.1.3.1 occasionalmente nevole
 DS2.1.1.3.2 anche in pianura
 DS2.1.1.3.1 specie nella seconda parte della giornata
 DS2.1.2 zone meno nuvolose
 DS2.1.2.1 Nuvolosità più attenuate sulle rimanenti zone del nord
 DS2.1.2.2 con addensamenti più consistenti sulle zone alpine
 DS2.1.2.3
 DS2.1.2.3.1 specie sui versanti settentrionali
 DS2.1.2.3.2 ove potranno verificarsi delle neviccate.

Cerchiamo la Most nuclear part di questa struttura



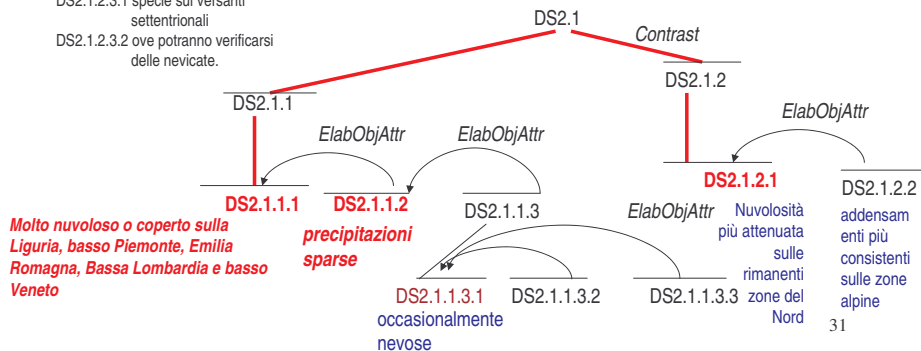
MNP (T, u)
 begin
 if Foglia(u, T) then {leggi il DS associato al nodo}
 else begin
 if RR_Nodo(u, T) = nucleo then begin
 c:= Primogenito (u,T);
 MNP(T, c);
 while not (UltimoFratello(c,T) do begin
 c:= SuccessivoFratello(c,T);
 MNP(T,c)
 end
 end
 end
 end

MNP(T, DS2.1)



DS2.1 Nord
 DS2.1.1 zone molto nuvolose
 DS2.1.1.1 Molto nuvoloso o coperto su Liguria, basso Piemonte, Emilia Romagna, Bassa Lombardia e basso Veneto
 DS2.1.1.2 con precipitazioni sparse
 DS2.1.1.3
 DS2.1.1.3.1 occasionalmente nevole
 DS2.1.1.3.2 anche in pianura
 DS2.1.1.3.1 specie nella seconda parte della giornata
 DS2.1.2 zone meno nuvolose
 DS2.1.2.1 Nuvolosità più attenuate sulle rimanenti zone del nord
 DS2.1.2.2 con addensamenti più consistenti sulle zone alpine
 DS2.1.2.3
 DS2.1.2.3.1 specie sui versanti settentrionali
 DS2.1.2.3.2 ove potranno verificarsi delle neviccate.

**ATTENZIONE:
 Se cambiamo le RR
 cambia anche la most nuclear part**



Frammenti linguistici associati alle RR

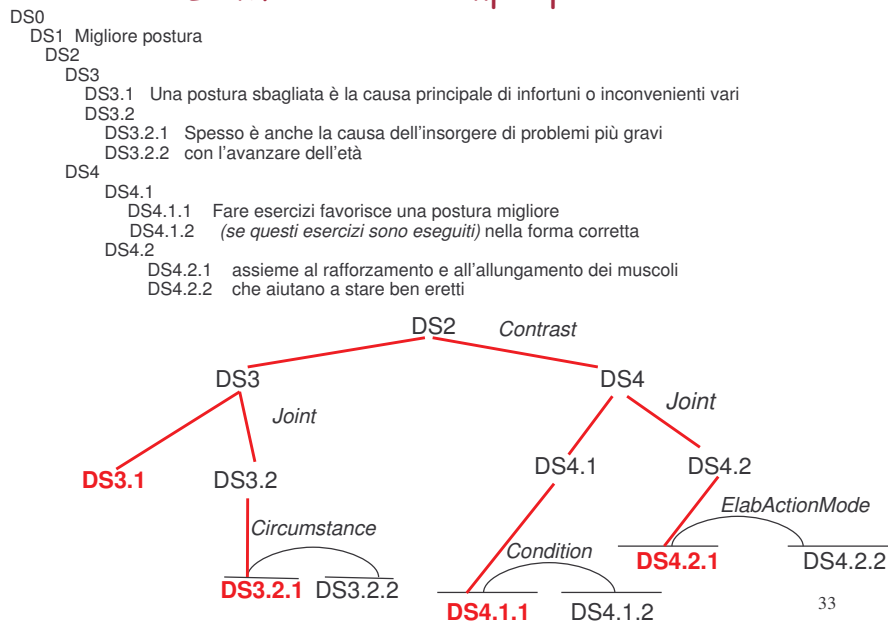
Notare che :

- la RR di Contrast con è stata resa con nessun frammento linguistico. Ma si sarebbe anche potuto dire *'Invece, nuvolosità più attenuata....'*
- il frammento linguistico 'con' può essere interpretato come indicativo di una Elaboration Object Attribute, ma anche di una Joint.

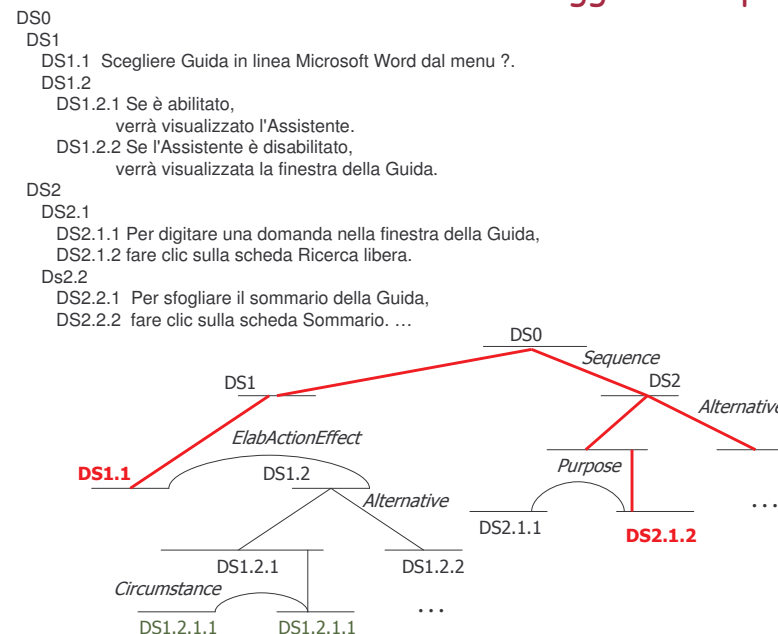
L'interpretazione della struttura retorica di un testo non è univoca.

Non lo è nemmeno la scelta dei frammenti linguistici con cui rendere una RR.

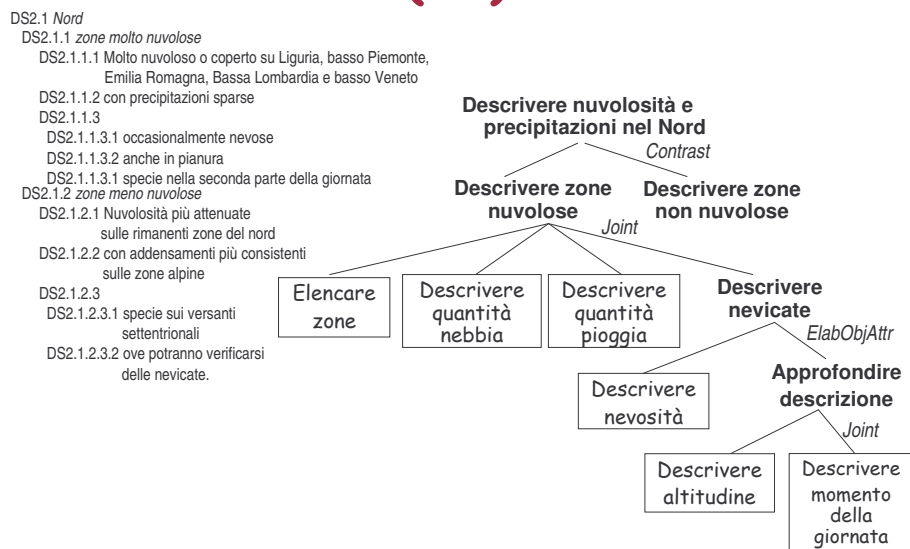
La MNP dell' esempio persuasivo



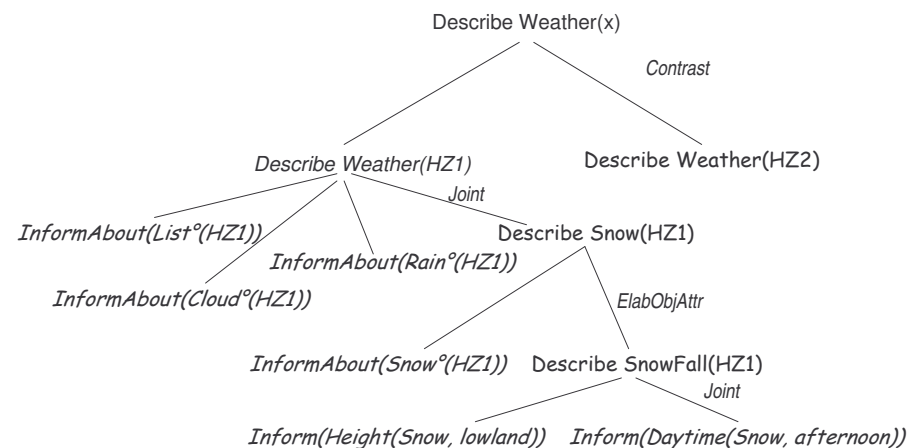
La MNP del messaggio di help



Da RST a Piano del Discorso (DP)



Un DP più generale



Rappresentazione DPML del piano

```

<xml version="1.0">
<DPML>
  <goal name="Describe" term="Weather(x)" RR="Contrast">
    <goal name="Describe" term="Weather(HZ1)" RR="Joint">
      <communicative_act name="InformAbout" term="List°(HZ1)" type="N">
        </communicative_act>
      <communicative_act name="InformAbout" term="Cloud°(HZ1)" type="N">
        </communicative_act>
      <communicative_act name="InformAbout" term="Rain°(HZ1)" type="N">
        </communicative_act>
      <goal name="Describe" term="Snow(HZ1)" RR="ElaborationObjectAttribute">
        <communicative_act name="InformAbout" term="Snow°(HZ1)" type="N">
          </communicative_act>
        <goal name="Describe" term="SnowFall(HZ1)" RR="Joint">
          <communicative_act name="Inform" term="Height(Snow, HZ1)" type="S">
            </communicative_act>
          <communicative_act name="Inform" term="Daytime(Snow, HZ1)" type="S">
            </communicative_act>
        </goal>
      </goal>
    </goal>
  </goal>
  <goal name="Describe" term="Weather(x2)" RR="Joint"> ... .. </goal>
</DPML>

```

37

Pseudo-Algoritmo per la generazione del testo mediante esplorazione del DP rappresentato in DPML

Siano T: RST tree; u, c: nodes;

```

NLG (T, u)
begin
  if Foglia(u, T) then {visualizza la frase preconfezionata associata all'atto
                        linguistico}

  else begin
    {visualizza il frammento linguistico associato la tag RR del nodo u}
    c:= Primogenito (u,T);
    NLG(T, c);
    while not (UltimoFratello(c,T) do begin
      c:= SuccessivoFratello(c,T);
      NLG(T,c)
    end
  end
end
end.

```

38

Realizzazione linguistica: da frase preconfezionata a SimpleNLG

NLG(DP)

se la radice di DP è una foglia allora invoca **SimpleNLG*** e visualizza il risultato

altrimenti

- effettua un table look-up per inserire il frammento linguistico che corrisponde alla RR associata;
- esegui NLG su ogni sottoalbero in ordine da sx a dx.

Esercitazione 4

Piano del Discorso  Generazione superficiale

*Da atti comunicativi a frasi:
Esperienza con SimpleNLG*

39

* class library which does basic NLG lexicalisation and realisation