

# Linguistica Computazionale

## Lezione 9

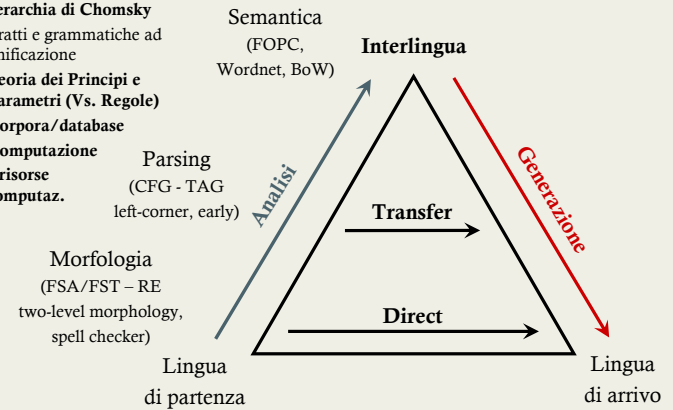
### Generazione

Lunedì 7 Gennaio 2008

Cristiano Chesi, chesi@media.unisi.it

## Riassunto delle cose dette sulla MT

- Regole di riscrittura e gerarchia di Chomsky
- Tratti e grammatiche ad unificazione
- Teoria dei Principi e Parametri (Vs. Regole)
- Corpora/database
- Computazione e risorse computaz.



Lezione 9 - Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Generazione

### • Indice

- Tipologie di generazione
  - Generazione "in scatola", transfer-based diretta e indiretta
  - Generazione con interlingua
  - Reversibilità e flessibilità
- Natural Language Generation
  - Architettura di un sistema di generazione
  - Pianificatore del discorso
  - Realizzatore di superficie

Lezione 9 - Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Lecture, approfondimenti

### • Bibliografia essenziale

- Hutchins & Somers (1992) *Cap. 7*
- Jurafsky & Martin (2000) *Cap. 20*

### • Approfondimenti

- G. Adorni, M. Zock (eds.) (1996) *Trends in Natural Language Generation - An Artificial Intelligence Perspective*. Berlin, Heidelberg: Springer

Lezione 9 - Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Generazione “in scatola” (chiusa)

Tipologie di generazione

- **Hello World!**  
println “Hello World!”;
- **Templates**  
Eliza (Weizenbaum, anno 1966):  
*Utente:* il mio ragazzo dice che sono sempre depressa  
*Eliza:* sono spiacente di sapere che sei depressa  
  
s/ sono [. \* | ](depress[o | a] | triste)/sono spiacente di sapere che sei \1/  
s/ sono tutt[i | e] (.\*) /in che senso sono \1?/  
s/ sempre / potresti far riferimento ad un esempio specifico?
- **Spam** (fresco di inbox!)  
“Buona Giornata, \$send\_to!  
\$from Invia ti eCard con saluti.  
Controlla il tuo attacco ; ) Buon Natale! Bye.”

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Reali necessità in generazione

Tipologie di generazione

- **Flessibilità d’uso**  
un sistema di generazione dovrebbe poter essere usato in molti contesti semantici
- **Riuscire a produrre un’appropriata varietà di forme**  
il sistema non può limitarsi a produrre pochi templates, ma deve padroneggiare varianti linguisticamente rilevanti
- **Scelta della varietà più adeguata**  
il sistema deve saper gestire efficientemente, in base agli intenti comunicativi e al contesto, le forme più appropriate selezionandole dal ventaglio delle (infinite) possibilità generabili

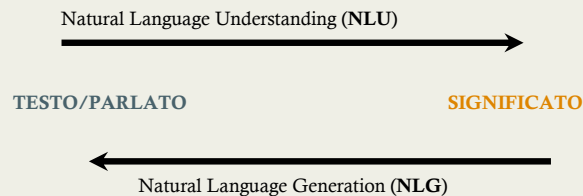
Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Definizione di generazione (NLG)

Tipologie di generazione

- Per **Generazione di Linguaggio Naturale (NLG, Natural Language Generation)** si intende la produzione di espressioni linguistiche grammaticali partendo da input non linguistici



Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Natural Language Understanding (NLU) Vs. Generation (NLG)

Tipologie di generazione

NLU	Vs.	NLG
<b>ambiguità</b> (lessicali, sintattiche, semantiche)		<b>selezione funzione comunicativa</b> (richiesta, dichiarazione)
<b>sottospecificazione tratti</b> (elementi mossi, ellissi)		<b>selezione struttura</b> (frase attiva, passiva...)
<b>input malformato</b> (robustezza errore)		<b>fare “economia” di parole</b> (uso espressioni referenziali, pronomi, ellissi)
		<b>selezione lessicale</b> (sinonimia)

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione transfer-based diretta Vs. indiretta

Tipologie di generazione



nei sistemi diretti  
transfer e generazione non sono  
moduli separati  
(es. SYSTRAN, METEO)

nei sistemi indiretti  
transfer e generazione sono moduli  
distinti (es. METAL)

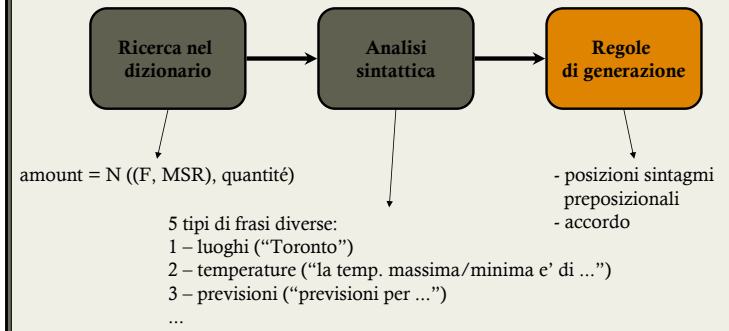
Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione transfer-based diretta Vs. indiretta

Tipologie di generazione

– Architettura sistema di traduzione METEO (lezione 1)



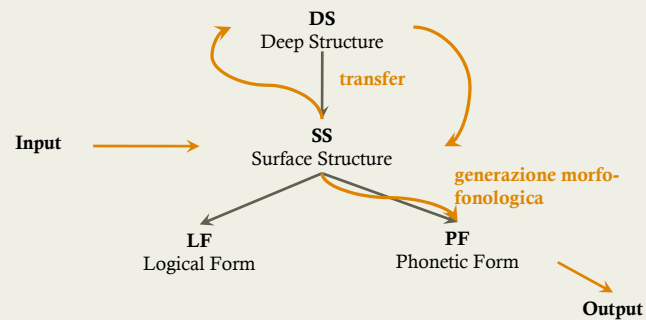
Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione transfer-based indiretta

Tipologie di generazione

– Recupero della Struttura Profonda (DS)



Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione transfer-based indiretta

Tipologie di generazione

Architettura sistema di traduzione ARIANE (GETA)

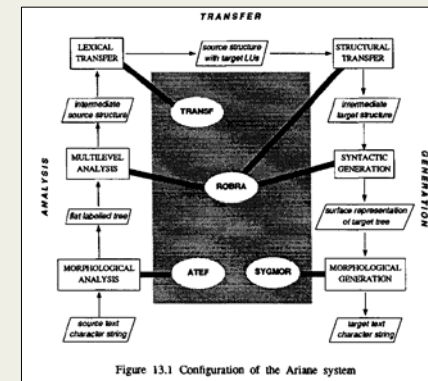


Figure 13.1 Configuration of the Ariane system

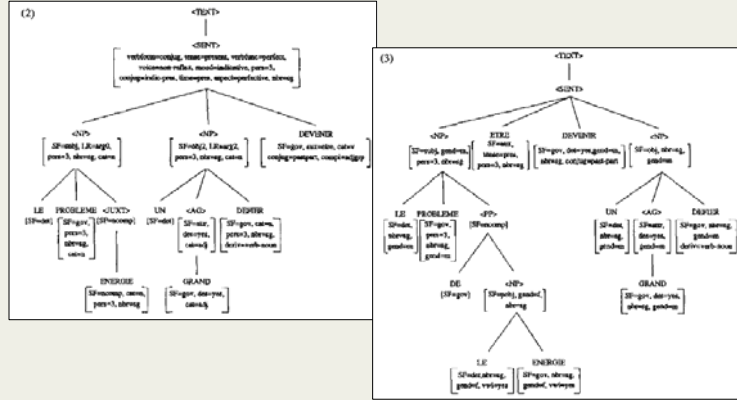
Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione transfer-based indiretta

Tipologie di generazione

- Es. Ariane System (tedesco > francese)



Lezione 9 - Generazione

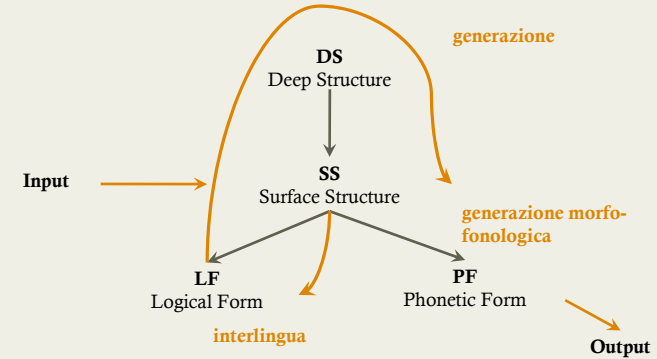
Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione interlingua

(inizio ...)

Tipologie di generazione

- Non si parte dalla Struttura Profonda (DS), ma dalla Forma Logica (LF)



Lezione 9 - Generazione

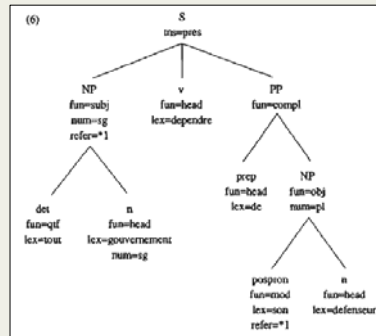
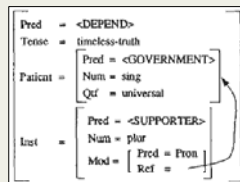
Ling. Comp. A.A.2007/08 - C. Chesì

## Generazione interlingua

(... fine)

Tipologie di generazione

- ENG: Any government is dependent on its supporters  
 FR: Tout gouvernement dépend de ses défenseurs  
 (IT: Ogni governo dipende dai suoi sostenitori)



Lezione 9 - Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Determinismo e complessità

(inizia ...)

Tipologie di generazione

- Tensione tra:
  - preservazione della **struttura** (tendenza sistemi automatici)
  - Vs.
  - preservazione del **significato** (tendenza traduttori umani)
- Il **determinismo** (corrispondenza 1-a-1 tra input e output) presenta vantaggi computazionali:
  - **rapidità** (numero minore di scelte)
  - **reversibilità**
  - **aggiornamento grammatiche**

ma anche svantaggi dal punto di vista "linguistico/cognitivo":

- forme troppo cristallizzate
- trasformazioni "opzionali" venivano ipotizzate anche in varie teorie generative ("inversioni stilistiche")

Lezione 9 - Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Determinismo e complessità

(... continua ...)

Tipologie di generazione

- Visione complessa: il processo di generazione è in realtà composto da una serie di decisioni da prendere

- struttura **topic-comment**

GER: Der Mann jagte den Haifisch  
IT: L'uomo ha cacciato lo squalo

GER: Den Mann jagte der Haifisch  
IT<sub>1</sub>: #Lo squalo ha cacciato l'uomo (così il **TOPIC** è diverso!)

soluzioni possibili:

IT<sub>2</sub>: L'UOMO, lo squalo ha cacciato (**focalizzazione**)  
IT<sub>3</sub>: L'uomo è stato cacciato dallo squalo (**passivizzazione**)

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Determinismo e complessità

(... continua ...)

Tipologie di generazione

- **relativizzazione**

ENG: the trains which are late  
IT: i treni che sono in ritardo

- **modificazione aggettivale**

ENG: the late trains  
IT<sub>1</sub>: \*i ritardi treni  
IT<sub>2</sub>: ?i treni ritardati (tardivi?)  
IT<sub>3</sub>: i treni in ritardo

- **passivizzazione**

ENG: the cake is/gets eaten  
IT<sub>1</sub>: \*la torta è mangiata  
IT<sub>2</sub>: la torta è stata mangiata  
IT<sub>3</sub>: la torta viene mangiata

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Determinismo e complessità

(inizio ...)

Tipologie di generazione

- **modificazione nominale “pesante”**

GER: die in den letzten zehn Jahren entwickelnden Industrien  
*le in gli ultimi dieci anni sviluppani industrie*

ENG: the industries which have been developing in the last ten years

IT: le industrie che si sono sviluppate negli ultimi dieci anni

IT<sub>A</sub>: ?le industrie che si sono sviluppate molto negli ultimi dieci anni in conseguenza del lungo periodo di stagnazione hanno corso i maggiori rischi

IT<sub>B</sub>: hanno corso i maggiori rischi le industrie che si sono sviluppate molto negli ultimi dieci anni in conseguenza del lungo periodo di stagnazione

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Determinismo e complessità

(... continua ...)

Tipologie di generazione

- **Ordine delle parole e portata quantificatori**

GER: Nur Kuchen hat das Kind gegessen

ENG<sub>1</sub>: \*Only cakes has eaten the child  
ENG<sub>2</sub>: The child has eaten only cakes  
ENG<sub>3</sub>: It's only cakes that the child has eaten  
ENG<sub>4</sub>: The child has eaten nothing but cakes

FR<sub>1</sub>: \*Seulement des gâteaux a mangé l'enfant  
FR<sub>2</sub>: Des gâteaux seulement a mangé l'enfant

IT<sub>1</sub>: SOLO DOLCI ha mangiato il bambino  
IT<sub>2</sub>: SOLO DOLCI il bambino ha mangiato  
IT<sub>3</sub>: il bambino ha mangiato solo dolci

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesi

## Determinismo e complessità

(... fine)

Tipologie di generazione

- Regole di preferenza:

- Orientamento all'**azione** (inglese) Vs. **processo** (giapponese)

ENG: I can see Mount Fuji

JP: Fuji-san ga mieru

Monte Fuji NOM può essere visto

- PP (ENG<sub>1</sub>) Vs. double-object construction (ENG<sub>2</sub>)

IT: Il ragazzo ha dato un libro a sua sorella

ENG<sub>1</sub>: The boy gave a book to his sister

ENG<sub>2</sub>: The boy gave his sister a book

ENG: ?The boy gave the man he met yesterday in the park after he'd been for a hamburger a book

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Reversibilità

Tipologie di generazione

- Separazione tra **dati** e **algoritmi**

i dati di un certo tipo (quelli che esprimono relazioni tra simboli) possono essere usati in parsing ed in generazione. (c'è però il rischio di avere sistemi semplicistici, spesso si riesce ad andare poco oltre una Struttura di Superficie)

- **Regole di Riscrittura, Prolog e programmazione logica**

approccio assiomatico, generare significa derivare strutture frasali dagli assiomi (S)

- **Finite States Transducers**

I:O facilmente scambiabili

- **Principi e Parametri**

problemi con inserzioni/cancellazioni, movimenti

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Flessibilità Vs. Reversibilità

Tipologie di generazione

- Gli algoritmi (o parte di questi algoritmi) possono essere applicati ad input diversi sia in **parsing** che in **generazione**

- **Ordinamento**

approccio assiomatico, generare significa derivare strutture frasali dagli assiomi (S), ma proceduralmente non tutte le applicazioni conducono ad uguali risultati

- **Unificazione**

verifica compatibilità unione

- **Aggiunta/cancellazione tratti**

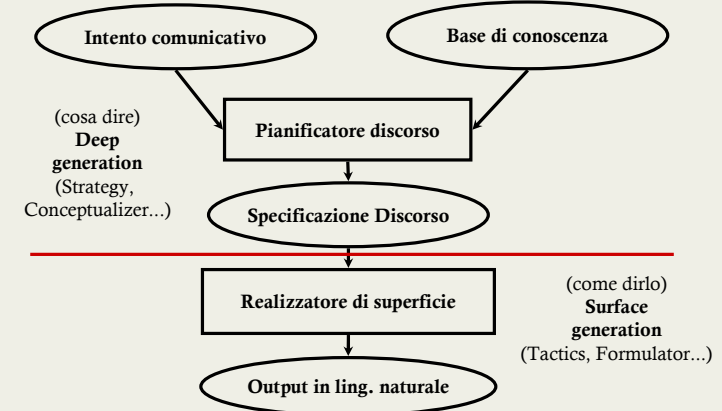
inserzioni/cancellazioni, movimenti (tensione tra economia e informatività)

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Architettura generale per un sistema di NLG

Natural Language Generation

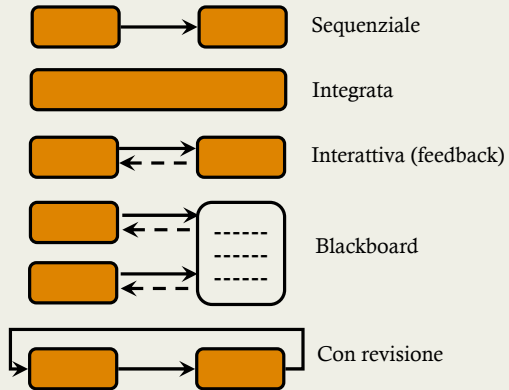


Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Controllo dell'informazione in generazione

Natural Language Generation



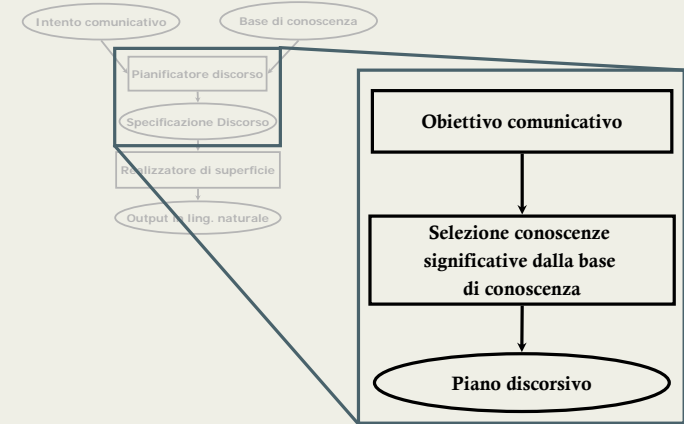
Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Discourse planner

(inizio ...)

Natural Language Generation



Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Discourse planner

(... continua ...)

Natural Language Generation

- Esempio di piano (SPL-constructor, PENMAN system, Mann 1983):

```
(
:process      salva
:actor       sistema
:goal        documento
:speechact   dichiarazione
:tense       futuro
)
```

(ipotizzate che ogni elemento corsivo sia un **synset** come in **Wordnet** o un elemento codificato in un formalismo simile a **KL-ONE**)

Lezione 9 – Generazione

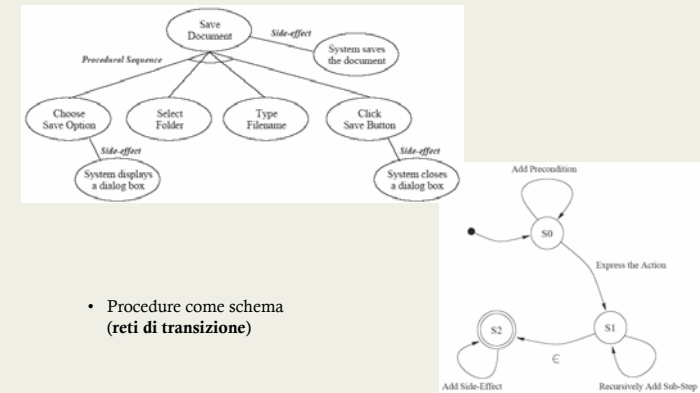
Ling. Comp. A.A.2007/08 - C. Chesì

## Discourse planner

(... fine)

Natural Language Generation

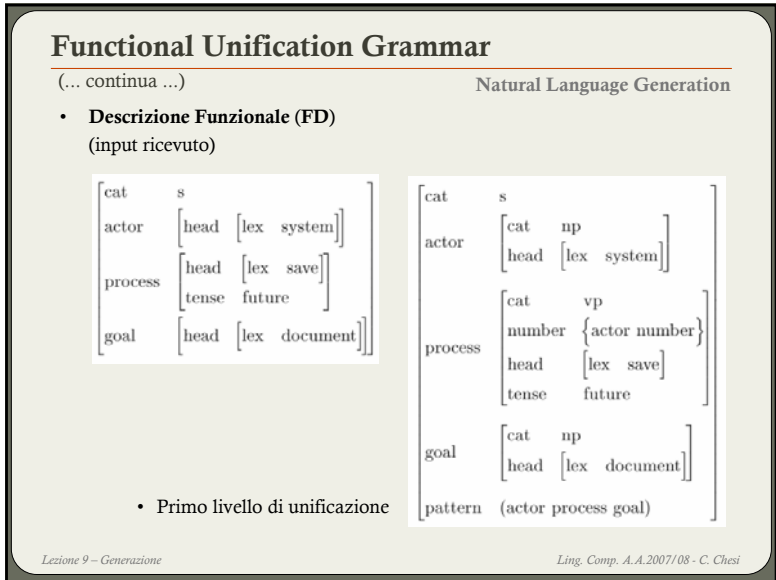
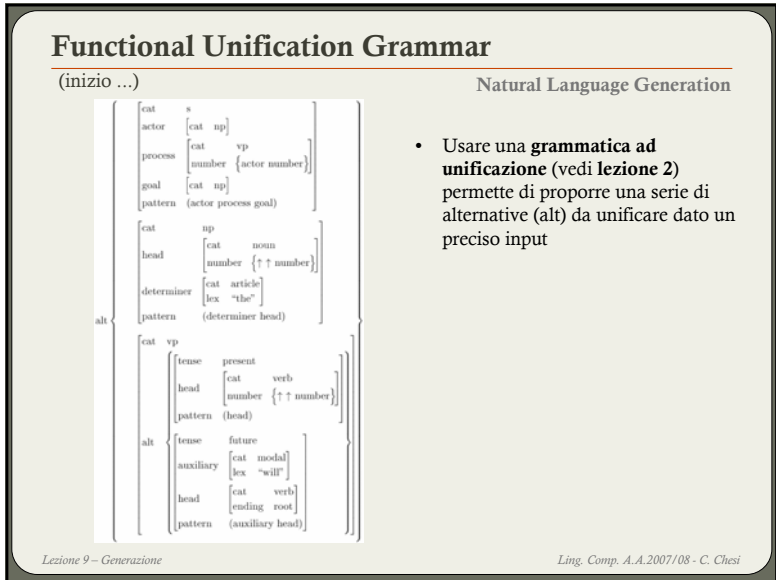
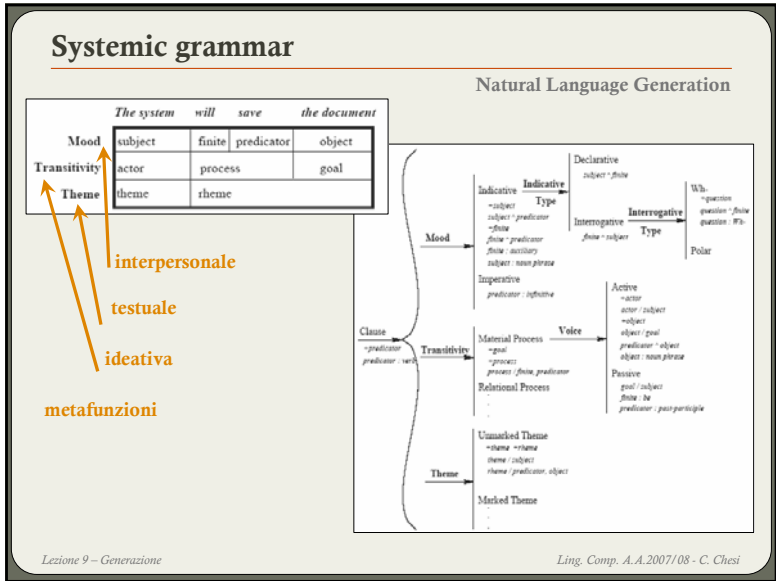
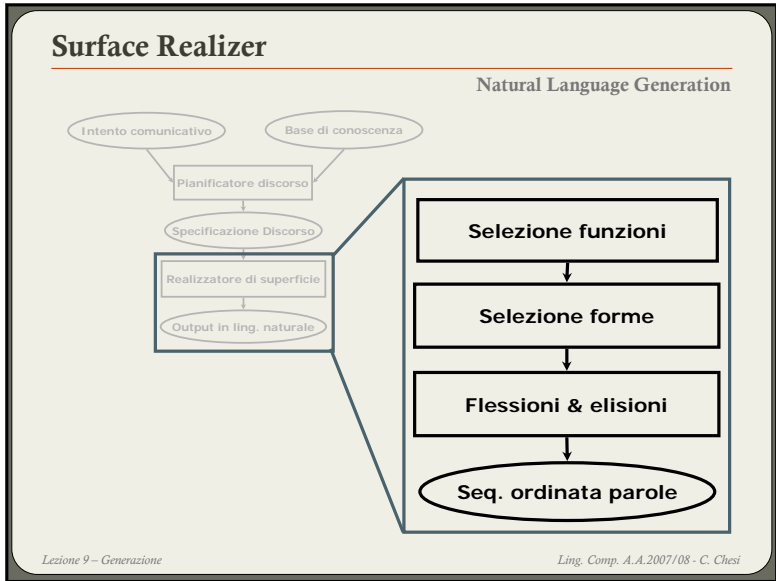
- Rappresentazione della conoscenza (stile **KL-ONE**):



- Procedure come schema (**reti di transizione**)

Lezione 9 – Generazione

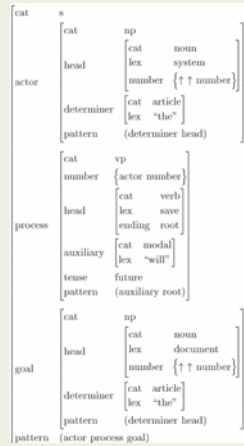
Ling. Comp. A.A.2007/08 - C. Chesì



## Functional Unification Grammar

(... fine)

Natural Language Generation



- Unificazione ricorsiva
- ...
- Output finale

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Aggregation: leggi di minima ridondanza (Dalianis, Hovy 1996)

Natural Language Generation

- **Subject grouping**  
Gianni mangia. Gianni guarda la televisione ⇒ **Gianni** mangia e guarda la televisione
- **Predicate grouping**  
Gianni mangia. Mario mangia. ⇒ Gianni e Mario **mangiano**
- **Clustering**  
Gianni mangia. Mario corre. Gianni guarda la tv. Mario poi si riposa.  
⇒ **Gianni** mangia e guarda la tv. Mario corre e poi si riposa.
- **Casting**  
Gianni mangia una mela. Una canzone è cantata da Mario. ⇒  
Gianni **mangia** una mela. Mario **canta** una canzone.
- **Economy**  
⇒ Gianni mangia una mela Vs. Una mela è stata mangiata da Gianni
- **Repetition**  
Gianni mangia una mela. Mario mangia una mela.  
⇒ Gianni **mangia una mela** e Mario **anche**

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Prossima lezione

(Mercoledì 9 Gennaio, ore 16-19, Aula Informatica A, Via Roma 47)

### • Laboratorio

#### – Valutazione sistemi di MT

- esplorazione sistemi di traduzione disponibili su Internet
- reverse engineering basato su evidenze empiriche e test linguistici

Lezione 9 – Generazione

Ling. Comp. A.A.2007/08 - C. Chesì