
Linguistica Computazionale – Lezione 7

Teoria dell'apprendibilità (learnability theory)

Lunedì 15 Marzo 2004
Cristiano Chesi, chesi@media.unisi.it

Teoria dell'apprendibilità

□ Indice

- Note sulla teoria formale dell'apprendibilità
 - formalizzazione del problema
 - l'apprendibilità nel limite
 - alcuni utili risultati

- L'apprendibilità delle lingue naturali
 - l'acquisizione della prima lingua
 - dagli algoritmi alle euristiche
 - modelli e problematiche

Lecture, approfondimenti

□ Bibliografia essenziale

- Pinker (1979) *Formal models of language learning*
Cognition 7, pp. 217-283

□ Approfondimenti

- Gold (1967) *Language identification in the limit*
Information and Control 10, pp. 447-474

- Stabler (2003) corso UCLA
(<http://ludoviko.linguistics.ucla.edu/stabler/212.pdf>)

Il problema

Teoria dell'apprendibilità

- **Problema di induzione:**
fare valide generalizzazioni sulla base di un numero finito*
di osservazioni (Pinker 1979:217)

* e limitato qualitativamente

Modello (informale) di apprendimento della prima lingua

Teoria dell'apprendibilità

- S_0 (stato iniziale universale, biologicamente determinato)
- **input** (uniforme durante tutto l'apprendimento, solo dati empirici positivi...)
- $S_0 \rightarrow S_1 \rightarrow \dots \rightarrow S_L$ (stadi di "sviluppo linguistico")
- S_L (stadio di sviluppo linguistico adulto, relativamente stabile)

5

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Condizioni che la teoria deve soddisfare

Teoria dell'apprendibilità

1. **Apprendibilità**: le lingue naturali sono effettivamente apprese
2. **Equipotenzialità**: tutte le lingue possono essere apprese
3. **Tempo**: ogni lingua è appresa in un arco di tempo limitato e piuttosto preciso
4. **Input**: tener conto della reale tipologia dell'input linguistico che il bambino riceve
5. **Fasi di sviluppo**: predizioni riguardo agli stadi di sviluppo intermedi effettivamente identificabili
6. **Plausibilità cognitiva**: limitazioni delle risorse cognitive a disposizione (memoria a breve termine, attenzione ecc.)

6

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Grammatiche formali

(riassunto delle puntate precedenti ...)

Teoria dell'apprendibilità

- **Linguaggio** (non finito, Ricorsivamente Enumerabile) = insieme infinito di **espressioni** (stringhe ben formate composte da simboli \in **vocabolario finito**)
- **Grammatica formale** $G = \langle V, V_T, \rightarrow, \{S\} \rangle$ dove:
 - V è il **vocabolario** della lingua
 - V_T è un sottoinsieme di V che racchiude tutti e soli gli **elementi terminali** (il complemento di V_T è V_N)
 - \rightarrow è una relazione binaria, asimmetrica e transitiva definita su V^* detta **relazione di riscrittura (regola di riscrittura)**:
 $\forall A \in V_N: \phi A \psi \rightarrow \phi \tau \psi$ per qualche $\phi, \tau, \psi \in V^*$
 - $\{S\}$ è un sottoinsieme di V_N definito come l'insieme degli assiomi che convenzionalmente contiene il solo simbolo S

7

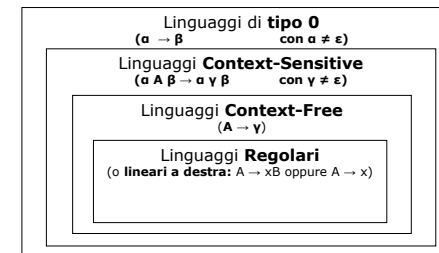
Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Grammatiche formali

(... riassunto delle puntate precedenti)

Teoria dell'apprendibilità

■ Gerarchia di Chomsky



α, γ, β possono essere sequenze di simboli terminali e non terminali (compreso l'elemento nullo ϵ salvo diversa specificazione); A, B sono singoli simboli non terminali; x è una sequenza di simboli terminali

8

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Grammatiche formali

Teoria dell'apprendibilità

- **Decidibilità:** data una stringa S si può dire se $\in L$?



- La classe delle grammatiche primitive ricorsive è **enumerabile**: esiste una **procedura finita** (chiamata grammar-grammar) che lista **esaustivamente** ogni grammatica che appartiene alla classe.

9

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(inizio ...)

Teoria dell'apprendibilità

- **Di cosa c'è bisogno:**
 - definizione di **apprendibilità**
 - un apprendista (**learner**) che identifichi il linguaggio dichiarandone il nome (cioè la grammatica)
 - metodo di **presentazione dell'informazione**
 - almeno una relazione di **identificazione** (che assegna nomi, cioè grammatiche, ai linguaggi)
- un linguaggio si dice **apprendibile** se esiste un algoritmo che l'apprendista può usare per identificare la grammatica che genera/riconosce tale linguaggio
- l'apprendibilità è una **proprietà della classe dei linguaggi** piuttosto che del linguaggio stesso

10

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

- **Identificazione del linguaggio nel limite** (Gold 1967):
 - **tempo**
insieme infinito di istanti discreti t
 - i_t
unità di informazione (una stringa) presentata all'istante t
 - **apprendista**
 $g_t = G(i_1, i_2, \dots, i_t)$
dove:
 g_t = nome della grammatica al tempo t
 G = funzione di apprendimenti
 (i_1, i_2, \dots, i_t) = unità di informazione presentate al tempo t

11

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

- **Procedura di identificazione del linguaggio nel limite:**
 - ad ogni istante t l'**apprendista** riceve un'unità di informazione i_t
 - appena ricevuta l'informazione i_t l'**apprendista** propone una grammatica g_t compatibile con l'informazione ricevuta fino a quel momento
 - Si dice che un linguaggio è **identificato nel limite** sse dopo un tempo finito t l'apprendista proporrà sempre la stessa grammatica g_t , corretta rispetto all'informazione ricevuta, ad ogni istante t_j tale che $t_j > t$
 - I linguaggi sono apprendibili sempre (e solo) rispetto ad una particolare **modalità di presentazione dell'informazione**

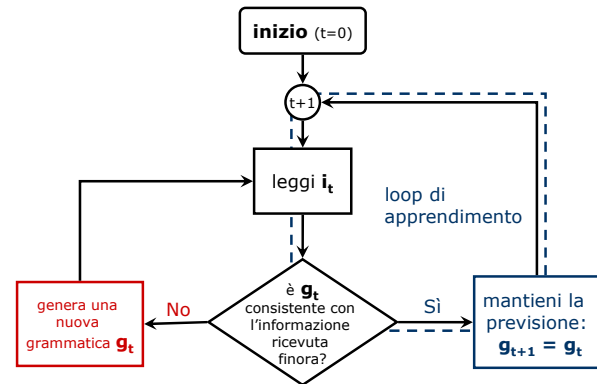
12

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio - schema

(... continua ...)

Teoria dell'apprendibilità



13

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

■ Modalità di presentazione dell'informazione:

- **TESTO (TEXT)**
sequenza informativa del tipo $x_1, x_2, \dots, x_n \in L$ dove ogni stringa che appartiene a L occorre almeno una volta nel testo. Per ogni lingua infiniti testi sono possibili; per esempio:
 - **testo arbitrario**: x_t può essere una qualsiasi funzione di t
 - **testo ricorsivo**: x_t può essere una funzione ricorsiva di t
 - **testo primitivo ricorsivo**: x_t può essere una funzione primitiva ricorsiva di t
- **INFORMATORE (INFORMANT)**
sequenza informativa per cui l'informante specifica se $y_n \in L$ oppure $y_n \notin L$. Per ogni lingua infiniti informatori sono possibili; per esempio:
 - **informatore arbitrario**: y_t può essere una qualsiasi funzione di t a patto che ogni stringa di V_T^* sia presente almeno una volta
 - **informatore metodico**: y_t è il t -esimo elemento di una funzione di enumerazione su V_T^*
 - **informatore a richiesta**: ad ogni istante t l'apprendista sceglie un y_t sulla base dell'informazione ricevuta fino a quel momento

14

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

■ Relazioni di identificazione

algoritmi che determinano l'appartenenza di una stringa s ad un linguaggio L

- **Verificatore (TESTER)**
procedura di decisione per L , cioè funzione $S \rightarrow \{1, 0\}$ (asigna ad ogni frase valore **1** se tale frase $\in L$, **0** se $\notin L$)
- **Generatore (GENERATOR)**
procedura che genera tutte e sole le espressioni in L ; una funzione, cioè $\mathbb{N} \rightarrow S$
(asigna un unico numero intero a ciascuna espressione possibile in L)

15

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

■ Risultati di apprendibilità nel limite:



16

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

■ Problemi di identificazione

- **Esistono infinite grammatiche che generano un linguaggio finito**
es. $L = \text{"La casa brucia"}$
 - G_1 > ogni espressione composta da tre parole
 - G_2 > ogni espressione generabile dal vocabolario $\{la, casa, brucia\}$
 - G_3 > ogni espressione riscrivibile come $\langle \text{Determinante, Nome, Verbo} \rangle$
 - G_4 > ogni espressione che contiene almeno un Nome
 - ...
 - G_L = tutte le frasi dell'italiano

17

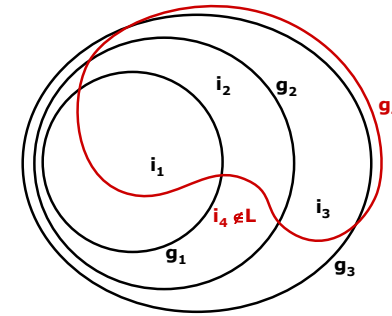
Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

- In che senso un **testo** è meno informativo di un **informatore?**



18

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Modello (formale) di apprendibilità del linguaggio

(... continua ...)

Teoria dell'apprendibilità

- data un'enumerazione delle grammatiche: $g_1, g_2, g_3 \dots g_L$
- **Informatore**
 1. rifiuta g_n se
 - a. $i_n \in L$ e i_n non è generabile/riconoscibile da g_n
 - b. $i_n \notin L$ e i_n è generabile/riconoscibile da g_n
 2. dato che g_n si troverà comunque nell'enumerazione ad un tempo finito t , una volta valutate tale corretta grammatica, l'apprendista non avrà più necessità di cambiare ipotesi
- **Testo**
 1. rifiuta g_n sse $i_n \in L$ e i_n non è generabile/riconoscibile da g_n
 2. è possibile imbrogliare l'apprendista proponendogli un testo basato su un linguaggio infinito costruito in modo da costringere l'apprendista a riformulare un'ipotesi di grammatica ad ogni istante i_n (ad esempio proponendo i_n compatibili con linguaggi finiti sempre più grandi)
 3. se, nel caso di linguaggi finiti, seleziona una grammatica $g_t \supset g_L$, nessun esempio potrà convincerlo a cambiare idea

19

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Implicazioni e proprietà della teoria dell'apprendibilità

(inizio ...)

Teoria dell'apprendibilità

- Proprietà dell'apprendimento nel limite
 - algoritmo di apprendimento è **generale**:
apprendere sequenze di oggetti, numeri, o stringhe di testo finisce per avere la stessa valenza formale
 - l'apprendimento prosegue teoricamente all'**infinito** e l'apprendista può **non sapere** mai che la sua ipotesi di identificazione è quella corretta, visto che potrebbe sempre sussistere il caso:
 $\exists i_t : G(i_1, i_2 \dots i_t) = g_t \neq g_{t-1}$
 - L'algoritmo di apprendimento, anche se esiste, potrebbe **non essere computabile**

20

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Implicazioni e proprietà della teoria dell'apprendibilità

(... continua ...)

Teoria dell'apprendibilità

- Proprietà dell'apprendimento nel limite
 - l'algoritmo di apprendimento deve avere un
 - **limite minimo**: generale sufficientemente da catturare tutte le regole in ogni lingua naturale
 - **limite massimo**: specifico sufficientemente da garantire la distinzione tra qualsiasi g_i , g_j tali che g_i è distinto da g_j
 - ci sono ragionevoli indizi per dubitare che un **testo** non sia sufficiente per apprendere una lingua naturale (teoria della povertà dello stimolo, "problema logico dell'acquisizione del linguaggio" Chomsky 65)
 - le procedure di enumerazione potrebbero essere estremamente **inefficienti**

21

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Implicazioni e proprietà della teoria dell'apprendibilità

(... continua ...)

Teoria dell'apprendibilità

- Ma allora come si può apprendere un linguaggio (naturale) da un **testo**?
 - determinare l'**ordine di presentazione** dell'informazione ad esempio utilizzando una funzione primitiva ricorsiva basata su t (ma sarebbe realistico pensare che ogni frase che il bambino processa è predeterminata esclusivamente dal tempo passato dall'inizio dell'apprendimento?)
 - rivedere la **condizione di successo** non identificare esattamente (nel limite) g_L , ma avvicinarsi ad essa in vari modi:
 - ordinare le grammatiche per complessità (Feldman 1972)
 - metriche di approssimazione (Wharton 1974)

22

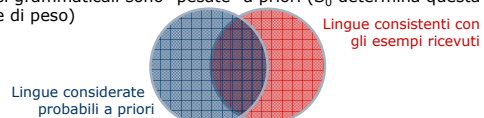
Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Implicazioni e proprietà della teoria dell'apprendibilità

(... fine)

Teoria dell'apprendibilità

- Ma allora come si può apprendere un linguaggio (naturale) da un **testo**?
 - **distribuzione statistica** sul campione di testo (Horning 1969)
 - data una grammatica probabilistica G_p , la probabilità che un'espressione S sia generata da tale grammatica è uguale al prodotto delle regole di riscrittura necessarie per generarla
 - ipotizzando che ogni G_p può essere enumerata seguendo un criterio di probabilità assoluta stipulato a priori, per ogni testo esiste un apprendista che determina la grammatica della lingua in questione
 - **vincoli sulle ipotesi grammaticali** (Chomsky 1965)
 - le ipotesi grammaticali sono "pesate" a priori (S_0 determina questa funzione di peso)



23

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Alcuni risultati rilevanti

Apprendibilità delle lingue naturali

- Fin qui la teoria presentata sembra adeguata per tre delle sei condizioni presentate all'inizio:
 - **apprendibilità** ci sono lingue effettivamente apprendibili (nel limite)
 - **equipotenzialità** (indirettamente) ogni lingua inclusa nella classe dei linguaggi primitivi ricorsivi è apprendibile
 - **input** è possibile dimostrare che:
 1. le lingue naturali non sono incluse nella classe dei **linguaggi regolari finiti**
 2. è comunque possibile apprendere (nel limite) linguaggi inclusi in classi gerarchicamente più ampie ponendo limitazioni sulla tipologia del testo presentato o sulla natura dell'enumerazione delle grammatiche

24

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Alcuni riflessioni sulle condizioni non soddisfatte

Apprendibilità delle lingue naturali

- Cosa manca ancora al modello di apprendimento?
- 1. **Tempo**: ogni lingua è appresa in un arco di tempo limitato e piuttosto preciso
- 2. **Fasi di sviluppo**: predizioni riguardo agli stadi di sviluppo intermedi effettivamente identificabili
- 3. **Plausibilità cognitiva**: limitazioni delle risorse cognitive a disposizione (memoria a breve termine, attenzione ecc.)

25

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Dagli algoritmi alle euristiche per problemi di TEMPO

Apprendibilità delle lingue naturali

- Soluzioni **algoritmiche**
 - **enumerazione e/o generazione cieca** di grammatiche inefficiente (problemi con il **tempo** di apprendimento) perché:
 - non considera relazioni di **equivalenza** (simboli non terminali con nomi diversi, ma usati allo stesso modo)
 - soluzioni **inefficienti** (loop inutili che non modificano l'input, regole che si annullano a vicenda, elementi nulli o non utilizzati)
 - **Quality control** (Wharton 1977) + **grammatical covering** (eliminare intere classi di grammatiche sulla base della valutazione di inefficienze di una singola grammatica)
- Soluzioni **euristiche**
 - non **valutano/rifutano grammatiche tutte insieme**, ma le costruiscono pezzo per pezzo (questo garantisce l'apprendibilità almeno di parte della grammatica g_L ma non di tutta la grammatica!)
 - ogni i_t non viene valutata nei termini $i_t \in / \notin g_t$, ma come fonte di **indizi** da usare incrementalmente per costruire le regole della grammatica (approccio incrementale, Brown 1973)

26

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Esempi di apprendimento euristico

Apprendibilità delle lingue naturali

- **Solomonoff (1964)**
 - apprendere **regole ricorsive context-free** ($A \rightarrow xAy$) con un informatore
 - 1. l'apprendista riceve un'informazione $i_t = Z$
 - 2. l'apprendista chiede all'informatore se la stringa $xZy \in L$
 - a. se $xZy \in L$, chiedi se anche $xxZyy$ e $xxxZyyy \in L$ vai al passo 3
 - b. se $xZy \notin L$, torna al passo 1
 - 3. aggiungi alla grammatica le regole
 - $A \rightarrow xAy$
 - $A \rightarrow Z$
 - evidenti i problemi con le limitazioni, ad esempio, sul numero massimo di incassamenti

27

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Apprendimento euristico... per FASI

(inizio ...)

Apprendibilità delle lingue naturali

- **Kelley (1967)**
 - primo sistema che simula l'apprendimento della prima lingua
 - obiettivo: tener conto delle **fasi di sviluppo** linguistico
 - euristica: **word-class position learning** (identificazione di categorie grammaticali sulla base di regolarità di distribuzione, Pivot Grammar, Braine 1963)
 - posizione assoluta (es. 1ª posizione nella frase)
 - posizione relativa (es. prima del Verbo)
- Marco
Maria
Marco
Maria

mangia
mangia
mangia
morde

una mela
una mela
una banana
una mela

}

$S \rightarrow A B C$
- A****B****C**

28

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Apprendimento euristico... per FASI

(... fine)

Apprendibilità delle lingue naturali

■ Kelley (1967)

- Tre insiemi di ipotesi che corrispondono alle fasi in cui il bambino produce frasi con 1, 2 o 3 parole:

Fasi	Euristica
1 parola	Usa solo parole contenute molto frequenti: S → C (per ipotesi taggate dall'adulto)
2 parole	Cose + azioni: S → Cose Azioni ; S → Cose Cose ... (l'algoritmo fa ipotesi e il sistema impara per rinforzo)
3 parole	Associa alle regole formate precedentemente un nuovo layer: S → Cose-Azioni Cose ...

- problemi:
 - la **frequenza** delle parole sembra irrilevante (Brown 1973)
 - trovare un **informatore** che dice se la classificazione è corretta sembra non realistico
 - difficile andare oltre ai **tre stadi** con l'apprendimento (d'altra parte ogni teoria ad hoc riuscirebbe a render conto di 3 stadi di questo tipo)

29

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

PLAUSIBILITA' COGNITIVA e gli approcci semantici

Apprendibilità delle lingue naturali

■ Cognitive Theory (Macnamara 1972), presupposti:

- apprendimento = associare **<espressione, significato>**
(<PF, LF> nelle recenti teorie)
- le espressioni sono sempre inserite in un **contesto**
- il **contesto** (semantico-pragmatico) fornisce indizi simili a quelli di un **informatore**
- le **informazioni semantiche** sembrano effettivamente utilizzate quando si apprende una lingua
- la **struttura semantica** e quella sintattica corrispondono in molti sostanziali aspetti

30

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

PLAUSIBILITA' COGNITIVA e gli approcci semantici

Apprendibilità delle lingue naturali

■ Formalizzazione della **Cognitive Theory** (Anderson 1974-76)

- significato delle espressioni = formalizzabile quindi enumerabile (Gödelizzazione)
- linguaggio naturale = f : espressioni → significati
- l'apprendista (il bambino) accede a coppie
<espressioni, significati>
(dove i significati vengono inferiti dal contesto)

31

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

PLAUSIBILITA' COGNITIVA e gli approcci semantici

Apprendibilità delle lingue naturali

■ **Language Acquisition System** (LAS, Anderson 1974-76)

- **Human Associative Memory System** (HAM) modello di rappresentazione dell'informazione
formalismo usato:
S = Soggetto
P = Predicato
R = Relazione
O = Oggetto
...
- **Augmented Transition Network** (ATN)
l'algoritmo di apprendimento cerca di parseggiare una frase da sinistra a destra (costruendo archi), se fallisce altera la grammatica

32

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

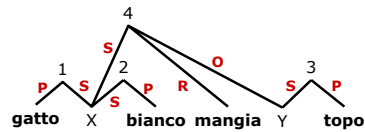
PLAUSIBILITA' COGNITIVA e gli approcci semantici

Apprendibilità delle lingue naturali

■ Language Acquisition System

(LAS, Anderson 1974-76)

- **prototype structure**: considera solo le parole contenuto
- **tree-fitting heuristic**
i bambini usano le informazioni semantiche per collegare le parole contenuto tra di loro:
Il gatto bianco mangia un topo
{1: X è un gatto, 2: X è bianco, 3: Y è un topo, 4: X mangia Y}



33

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

PLAUSIBILITA' COGNITIVA e gli approcci semantici

Apprendibilità delle lingue naturali

■ Language Acquisition System

(LAS, Anderson 1974-76)

- **Semantic-induced generalization**
unione strutture semantiche = generalizzazione
es. il gatto bianco mangia il topo
l'uomo mangia il pollo
il gatto bianco mangia *con gusto* il topo
- **Problemi**
 - mancate generalizzazioni ("sopra" ≠ "sotto")
 - ipergeneralizzazioni ("versare" = "riempire")
 - difficoltà nell'integrare nella struttura "morfemi grammaticali" (elementi funzionali come articoli, preposizioni ecc.)

34

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Difficoltà nell'apprendere grammatiche con movimento

Apprendibilità delle lingue naturali

- Gli elementi si muovono... non è facile capire da dove vengono
 - What_i did John buy t_i?
 - ... ma a questo punto dovremmo saperne un po' più di parsing

35

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì

Prossima lezione (domani ore 16-18 - Aula 6 Virtus)

- Indice
 - Problemi che si prestano ad un approccio subsimbolico
 - in cosa consiste
 - quando e perché lo si preferisce ad un approccio simbolico
 - Reti neurali
 - neurofisiologia
 - reti di neuroni artificiali
 - NLP con reti neurali: Simple Recurrent Networks
 - memoria a breve termine
 - apprendere incrementalmente il linguaggio
 - acquisire proprietà ricorsive con le SRN

36

Linguistica Computazionale AA 2003-04 - L. Rizzi, C. Chesì