

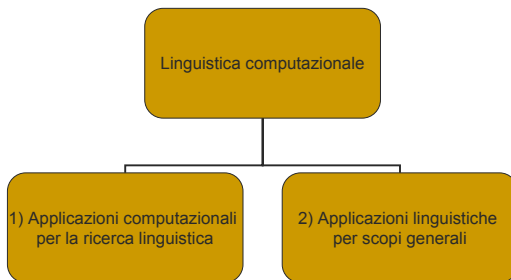
Informatica e lingue naturali

Teorie e applicazioni computazionali per la ricerca sulle lingue

La linguistica computazionale

- Luogo d'incontro tra linguistica teorica (e applicata) e tecnologie informatiche
- Area interdisciplinare e multidisciplinare
- Area legata a teorie linguistiche anche molto distanti

Linguistica computazionale



1) Applicazioni computazionali per la linguistica

- *Linguistica dei corpora*
 - Paradigma comportamentista
 - Approccio legato allo studio dei TESTI
- *Natural Language Processing*
 - Paradigma chomskyano
 - Approccio legato allo studio delle REGOLE e della competenza linguistica

2) Applicazioni linguistiche per scopi generali

- Approccio inizialmente a-teorico improntato a scopi tecnico-pratici applicativi
 - Traduzione automatica (primi stadi)
 - Dizionari elettronici
 - Correttori ortografici

Contributi teorici collaterali

Teorie che hanno fortemente influenzato gli sviluppi recenti della linguistica computazionale

Contributi principali

- Chomsky: modelli e linguaggi formali
- La teoria dell'informazione (Shannon)
- La statistica linguistica (Zipf, Guiraud, Herdan)
- L'intelligenza artificiale

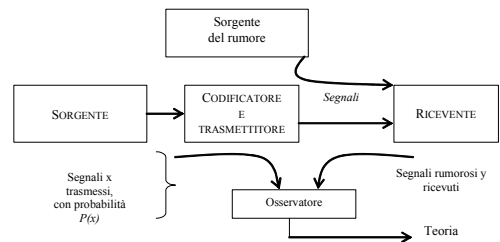
Il paradigma di Chomsky

- Teoria generativo-trasformativale
- Formalizzazione delle lingue
- Centralità del sistema a REGOLE
- Centralità della SINTASSI
- Natural Language Processing
 - Natural language generation
 - Natural language understanding
- Il Parsing
- Statistical Natural Language Processing

La teoria dell'informazione

- Modello di **teoria della trasmissione** (fisica del segnale) – Shannon
- Informazione come scelta (a-semantica)
- *Teorema Fondamentale per un canale discreto con disturbo*
- «vi sono, comunque, dei modi di trasmettere le informazioni i quali sono ottimali nel combattere il disturbo» (1949: 72)
- La ridondanza
- Centralità della dimensione statistica del messaggio

La teoria dell'informazione



La statistica linguistica

- Primi lavori statistici degli **stenografi** (Kaeding, Estoup)
- **G. K. Zipf** e il principio del minimo sforzo (tendenze del lessico e dei sistemi fonologici)
- La distribuzione statistica del lessico (**Guiraud**)

La statistica linguistica

- La didattica delle lingue: **wordbooks** (Thorndike, Knease)
- I **dizionari fondamentali** (*Basic English*, Vocabolario di base, ecc.)
- I **lessici di frequenza** (LIF, LIP)