

# Linguistica Computazionale

## Lezione 7

Rappresentazione della conoscenza e disambiguazione

Lunedì 15 Dicembre 2007  
Cristiano Chesi, chesi@media.unisi.it

## Il problema dell'apprendibilità

- **Indice**
  - Rappresentazione del significato
    - First Order Predicate Calculus
    - Eventi, tempo, aspetto e modo
    - Metodi complementari per la rappresentazione della conoscenza
  - Analisi semantica
    - Come si associano predicati del primo ordine a strutture sintattiche
    - Fenomeni semanticamente interessanti (e problematici)
  - Strategie di disambiguazione
    - Bag of Words
    - Reti semantiche

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

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

## Lecture, approfondimenti

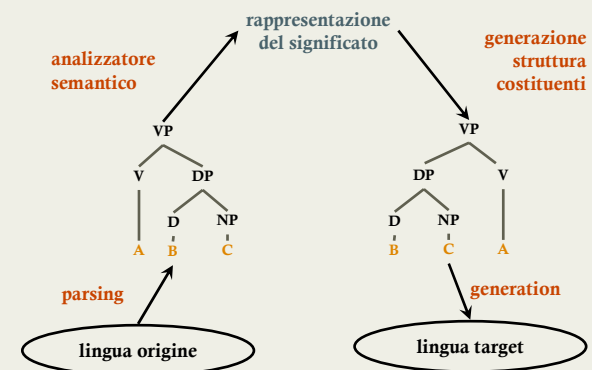
- **Bibliografia essenziale**
  - Hutchins & Somers (1992) *Cap. 6*
  - Jurafsky & Martin (2000) *Cap. 14, 15(, 16), 17*
- **Approfondimenti**
  - Ide & Véronis (1998) *Word Sense Disambiguation: The State of the Art. Tutorial*
  - Brachman & Schmolze (1985) *An overview of the KL-ONE knowledge representation system. Cognitive Science, 9(2), pp. 171-216*
  - Miller G. (1993) *Five papers on wordnet.*
  - Agirre & Edmonds (2006) *Word Sense Disambiguation Algorithms and Applications. Springer*
  - Heim & Kratzer (1998) *Semantics in generative grammar. Malden, MA., Blackwell Publishers*

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

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

## Modello dell'Interlingua (lezione 1)

Rappresentazione del significato



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

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

## Proprietà di una rappresentazione semantica

Rappresentazione del significato

- **Verificabilità**  
decidibilità della relazione tra rappresentazione del significato e stato del mondo
- **Non ambiguità**  
rappresentazione non riferibile a due distinti stati del mondo contemporaneamente (Vs. vaghezza)
- **Forma canonica (canonical form)**  
date due rappresentazioni superficiali distinte (es. forma passiva e attiva) deve esistere uno standard per rappresentarne univocamente il significato
- **Inferenza**  
Trarre valide conclusioni dalla rappresentazione data (uso di variabili)
- **Espressività**  
Rappresentare adeguatamente un certo numero di stati del mondo

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(inizio ...)

Rappresentazione del significato

- **Termini (oggetti)**
  - **costanti** {A, B, C... Gianni...}
  - **variabili** {x, y, z...}
  - **funzioni** f(termini)
- **Connettivi**
  - {¬, ∧, ∨, ⇒ ...}
- **Quantificatori**
  - {∀, ∃}

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Predicati** (e strutture argomentali):

Mario beve birra  
*beve*(Mario, Birra)

↑            ↙            ↘  
**predicato**   **argomenti**

(funzione) (ruoli tematici, arità = num. di argomenti)

- **Nomi e preposizioni/avverbi** pure possono proiettare una struttura tematica:

*sopra*(Torta, Tavolo)

- Predicato (Termini)
- Formule:
  - **Predicati** ∪ **Connettivi** ∪ **Quantificatori**

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Ragionamento**
  - **Modus Ponens**  
$$\frac{P \quad P \Rightarrow Q}{Q}$$
$$\frac{\text{uomo}(\text{Socrate}) \quad \forall x \text{ uomo}(x) \Rightarrow \text{mortale}(x)}{\text{mortale}(\text{Socrate})}$$
  - **Forward chaining**  
applicazione di una serie di regole note (da sinistra a destra)
  - **Backward chaining**  
inversa al forward chaining (applicazione regole da destra a sinistra)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Categorie**  
predicati unari:  
uomo(Socrate) (ma così le categorie sarebbero semplicemente relazioni e non “oggetti”)  
\*più\_famoso(Socrate, uomo)  
(un argomento di un predicato deve essere un termine, non un altro predicato)
- **Reificazione**
  - **ISA** (“is a”, “è un”)  
*ISA*(Socrate, Filosofo)
  - **AKO** (“a kind of”, “è un tipo di”)  
*AKO*(Filosofo, Uomo)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Eventi**
  - mangiare(Panino, Gianni)
  - mangiare(Panino, Gianni, Oggi)
  - mangiare(Panino, Gianni, Oggi, 13:30, Bar, P.ta Romana)
  - ...
  - **meaning postulates**  
 $\forall x, y, z \text{ mangiare}_3(x,y,z) \Rightarrow \text{mangiare}_2(x,y)$
- usando la reificazione (Davidson '67, Parson '90) si risolvono i problemi legati alla necessità di specificare meaning postulates

$\exists x \text{ ISA}(x, \text{mangiare})$   
 $\wedge \text{mangiante}(x, \text{Parlante}) \wedge \text{mangiato}(x, \text{Panino})$   
 $\wedge \text{luogo\_pasto}(x, \text{Bar}) \dots$

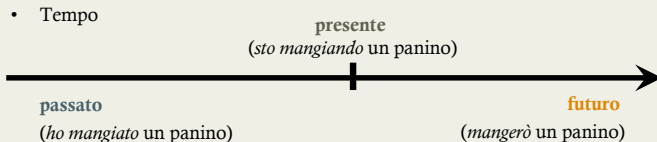
Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato



$\exists x \text{ ISA}(x, \text{mangiare})$   
 $\wedge \text{mangiante}(x, \text{parlante}) \wedge \text{mangiato}(x, \text{panino})$

- Come prima approssimazione possiamo dire che:
    - un evento si svolge in un intervallo di tempo (i)
    - ha un punto di terminazione (e)
    - è collocato nel tempo rispetto al presente (now)
- presente:** ...  $\wedge \text{intervallo\_di}(x, i) \wedge \text{membro\_di}(i, \text{now})$   
**passato:** ...  $\wedge \text{intervallo\_di}(x, i) \wedge \text{termine}(i, e) \wedge \text{precede}(e, \text{now})$   
**futuro:** ...  $\wedge \text{intervallo\_di}(x, i) \wedge \text{termine}(i, e) \wedge \text{precede}(\text{now}, e)$

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

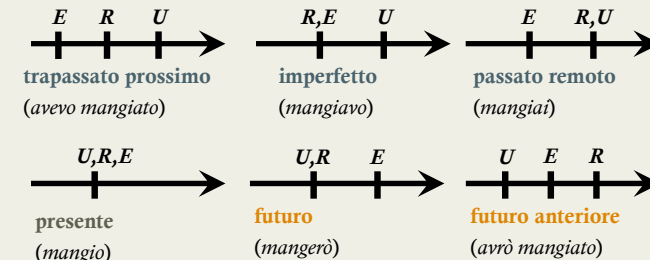
Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- Introduzione del **Reference Point** (Reichenbach '47)
  - *E* tempo dell'evento
  - *R* tempo del riferimento
  - *U* tempo dell'enunciazione



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Aspetto**

si definisce aspetto sia un tratto *lessicale*, inerente al predicato, che uno più propriamente *grammaticale* che definisce il modo con cui il parlante presenta l'evento

le **classi aspettuuali** (Vendler 1967) che determinano l'aspetto lessicale sono solitamente identificate con le seguenti:

*durativi / atelici*

- *stato* (Socrate è mortale)
- *attività* (Socrate camminava per ore)

*terminativi / telici*

- *completamento* (*accomplishment*, Socrate bevve il vino)
- *compimento* (*achievement*, Socrate raggiunse la vetta)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Classi aspettuuali** e tratti associati (Vendler 67, Dowty 79)

Classe aspettuuale	Telicità	Statività	Duratività	Esempi
Stati	-	+	+	sapere, possedere
Attività	-	-	+	camminare, cantare
Completamenti	+	-	+	costruire una casa, cantare una canzone
Compimenti	+	-	-	raggiungere la vetta, vincere una corsa
Puntuali	-	-	-	tossire, schiacciare la frusta

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- Per determinare la (a)telicità di un evento si può ricorrere al test in-X-time (predicati telici), for-X-time (predicati atelici). Tale proprietà è determinata sia *inerentemente* che *composizionalmente*:

- **assegnazione di caso** e telicità dell'evento (es. partitivo/accusativo in Finlandese)
- **transitività/intransitività**:  
Gianni ha bevuto *per un'ora* / *?in un'ora*  
Gianni ha bevuta una birra *?per un'ora* / *in un'ora*
- la **definitezza** dell'oggetto diretto (Specific Quantity Property, Verkuyl 93):  
Gianni beve birra *?in un'ora* / *per un'ora*  
Gianni beve due birre *in un'ora* / *?per un'ora*
- **conoscenza contestuale**:  
(nessun contesto) *?Gianni correva in un'ora*  
(ogni giorno Gianni corre per 10 Km) Gianni correva in un'ora (10 km).

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **L'aspetto grammaticale** determina la prospettiva (o punto di vista) del parlante sull'evento e serve a focalizzarne una porzione precisa (Smith 91, Klein 94, Olsen 94):

- **interna** (aspetto imperfettivo, viene focalizzata la parte intermedia o finale dell'evento, escludendo però le periferie estreme, cioè i punti di inizio e di fine; es. sto mangiando una mela)
- **esterna** (aspetto perfettivo, viene focalizzata la fine dell'evento; es. ho appena finito di mangiare una mela)

intuitivamente sembra sussistere una correlazione naturale tra:

- perfettività ⇔ telicità
- imperfettività ⇔ atelicità

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... continua ...)

Rappresentazione del significato

- **Modalità (e credenze)**  
Permette di riferirsi a “mondi possibili” (Lewis 73) ad esempio usando verbi speciali come *credere*, *volere*, *immaginare*, *sapere*...  
Le stesse categorie “modali” possono essere espresse in certe lingue attraverso il **modo** in altre attraverso la **modalità**:
  - il **modo** esprime l’opinione del parlante (la sua posizione nei confronti di una proposizione) ed è integrato nella morfologia verbale
  - la **modalità** è espressa attraverso una morfologia distinta da quella verbale (verbi, ausiliari o altre particelle come gli avverbi)
- Alcune delle distinzioni modali che si possono fare sono le seguenti:
  - reale/irreale (indicativo/congiuntivo);
  - deontico/epistemico (dovere/potere);

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## First Order Predicate Calculus (FOPC)

(... fine)

Rappresentazione del significato

- Rappresentare la modalità con FOPC
- “Credo che Mario abbia mangiato un panino”**
- $$\exists x \text{ ISA}(x, \text{mangiare})$$
- $$\wedge \text{mangiante}(x, \text{Mario}) \wedge \text{mangiato}(x, \text{panino})$$
- ...
- ma “mangiato(x, panino)” non è necessariamente vero!
- \*credere(io, mangiare(Mario, panino)) (non è una formula del First Order!)
- credere(io, < $\exists e \text{ ISA}(e, \text{mangiare})$ >  
 $\wedge \text{mangiante}(e, \text{Mario}) \wedge \text{mangiato}(e, \text{panino})$ )

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

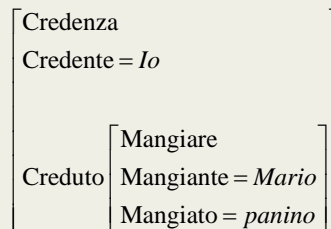
Ling. Comp. A.A.2007/08 - C. Chesì

## Formalismi per rappresentare il significato

(inizio ...)

Rappresentazione del significato

- **Semantic Networks (o frames)**  
“Credo che Mario abbia mangiato un panino”



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Formalismi per rappresentare il significato

(... fine)

Rappresentazione del significato

- **Semantic Networks (o frames)**  
un esempio: **KL-One**  
Brachman & Schmolze 1985

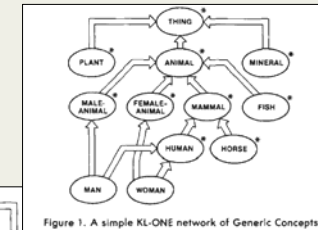


Figure 1. A simple KL-ONE network of Generic Concepts.

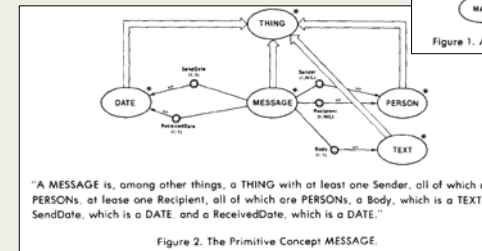


Figure 2. The Primitive Concept MESSAGE.

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

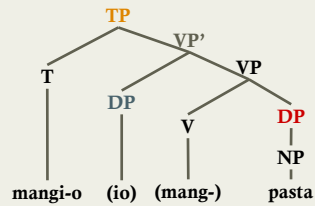
## Syntax-driven semantic analysis

Analisi semantica

- **Obiettivo:** associare rappresentazioni semanticamente significative alle espressioni ben formate, sulla base del solo lessico e della grammatica

$\exists x \text{ ISA}(x, \text{mangiare})$

$\wedge \text{mangiante}(x, \text{io}) \wedge \text{mangiato}(x, \text{pasta})$   
 $\wedge \text{intervallo\_di}(x, i) \wedge \text{membro\_di}(i, \text{now})$



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Ipotesi Rule-to-Rule (Bach 76)

Analisi semantica

- **Associare ad ogni regola di riscrittura un'adeguata rappresentazione semantica da combinare composizionalmente**

- A	→ $\alpha_1 \alpha_2 \dots \alpha_n$	$\{f(\alpha_j.\text{sem} \dots \alpha_k.\text{sem})\}$
- N	→ panino	$\{\text{Panino}\}$
- N <sub>proprio</sub>	→ Gianni	$\{\text{Gianni}\}$
- V	→ mang-	$\{\exists e, x, y \text{ ISA}(e, \text{mangiare})$ $\wedge \text{mangia}(e, x) \wedge \text{mangiato}(e, y)\}$
- T	→ -o	$\{\exists i, e \text{ intervallo\_di}(i, e) \wedge$ $\text{membro\_di}(i, \text{now}) \}$
- D	→ un	$\{\exists\}$
- DP	→ N <sub>proprio</sub>	$\{\text{N}_{\text{proprio}}\}$
- DP	→ D N	$\{\langle \text{D.sem} \times \text{N.sem}(x) \rangle\}$
- VP	→ V DP	$\{\text{V.sem}(\text{DP.sem})\}$

(dopo vediamo perché queste ultime 2 regole sono così “complesse”)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... continua ...)

Analisi semantica

### • Frasi

le diverse tipologie di frasi possono essere espresse da operatori che si applicano alle formule FOPC

- **Dichiarative** (informazioni fattive, es. “mangio un panino”)
   
S → VP DP      {DCL(VP.sem(DP.sem))}
- **Imperative** (richiesta, azione, es. “mangia!”)
   
S → VP      {IMP(VP.sem(tu))}
- **Domande sì/no** (es. “hai mangiato un panino?”)
   
S → aux V DP?      {YNQ(VP.sem(DP.sem))}
- **Domande wh-** (es. “cosa hai mangiato?”)
   
S → wh- VP DP      {WHQ(DP.sem.var, VP.sem(DP.sem))}

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... continua ...)

Analisi semantica

### • Sintagmi Nominali

gli SN (o DP) possono essere sia semplici espressioni FOPC che termini complessi, il problema è che spesso la sintassi non è totalmente trasparente (pochi indizi per rappresentare univocamente il significato), ad esempio:

- **genitivi**
  
“La rocca di David”

DP → NP di NP      {< $\exists x \text{Nominal.sem}(x) \wedge \text{GN}(x, \text{di\_NP.sem})$ >}  
 di NP → di NP      {NP.sem}

{< $\exists x \text{ ISA}(x, \text{rocca}) \wedge \text{GN}(x, \text{David})$ >}

problema: questa regola analizzerebbe (erroneamente) anche “la rocca di pietra” come genitivo

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... continua ...)

Analisi semantica

### • Sintagmi aggettivali

– **nominali** Vs. **predicativi**

“(una) palla rossa”

“una palla è rossa”

Nominale  $\rightarrow$  N Agg  $\{\exists x \text{ Nominale.sem} \wedge \text{ISA}(x, \text{Agg.sem})\}$

Agg  $\rightarrow$  rossa {rosso}

N  $\rightarrow$  palla {palla}

$\exists x \text{ ISA}(x, \text{palla}) \wedge \text{ISA}(x, \text{rossa})$

- problema: questo tipo di rappresentazione funziona bene solo con classi (nominali e aggettivali) **intersettive**, ma non con cose del tipo “pistola finta” o “ex fidanzata”

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... continua ...)

Analisi semantica

### • Sintagmi verbali

- infinitivi e **verbi a controllo** del soggetto/oggetto

“Mario **chiede** a Gianni di partire”

$\exists e, f \text{ ISA}(e, \text{chiedere}) \wedge \text{ISA}(f, \text{partire})$

$\wedge \text{chiede}(e, \text{Mario}) \wedge \text{chiede\_a}(e, \text{Gianni}) \wedge \text{chiesto}(e, f)$

$\wedge \text{partito}(f, \text{Gianni})$

“Mario **promette** a Gianni di partire”

$\exists e, f \text{ ISA}(e, \text{promettere}) \wedge \text{ISA}(f, \text{partire})$

$\wedge \text{promette}(e, \text{Mario}) \wedge \text{promette\_a}(e, \text{Gianni}) \wedge \text{promesso}(e, f)$

$\wedge \text{partito}(f, \text{Mario})$

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... continua ...)

Analisi semantica

### • Sintagmi preposizionali

- **modificatori nominali**

“un numero sulla busta”

$\{\exists x \text{ ISA}(x, \text{numero}) \wedge \text{su}(x, \text{busta})\}$

DP  $\rightarrow$  D Nominale  $\{<D.sem \ z \ N.sem(z)>\}$

Nominale  $\rightarrow$  N PP  $\{\exists z \ N.sem(z) \wedge \text{PP.sem}\}$

PP  $\rightarrow$  P DP  $\{P.sem(DP.sem)\}$

P  $\rightarrow$  su  $\{\exists x, y \ \text{su}(x, y)\}$

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... continua ...)

Analisi semantica

### • Sintagmi preposizionali

- **modificatori verbali**

– “ricorda **con rabbia**”

VP  $\rightarrow$  VP PP  $\{\exists y \ \text{VP.sem}(y) \wedge \text{PP.sem}(\text{VP.sem.variabile})\}$

ricorda  $\{\exists x, e \ \text{ISA}(e, \text{ricordare}) \wedge \text{ricorda}(e, \text{lui})\}$

con rabbia  $\{\exists x \ \text{con}(x, < \exists h \ \text{rabbia}(h) >)\}$

- **argomenti verbali**

“dare un bacio **a Maria**”

PP  $\rightarrow$  P NP  $\{\text{NP.sem}\}$

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Fenomeni semanticamente interessanti

(... fine)

Analisi semantica

- **Espressioni idiomatiche**

“la punta dell’Iceberg”

NP → la punta dell’Iceberg {inizio}

“tirare le cuoia”

VP → tirare le cuoia {morire}

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Ambiguità semantica: prospettiva (cross-)linguistica

English wall	German	Wand (inside a building) Mauer (outside)
English river	French	rivière (general term) fleuve (unique river, flowing into sea)
English leg	Spanish	pierda (human) pata (animal, table) pata (chair)
	French	étape (of a journey) jambe (human) patare (animal, insect) piéde (table, chair)
English blue	Russian	goluboi (pale blue) sinii (dark blue)
French laser	English	laser or laser
French colombe	English	pigeon or dove
German Leihen	English	borrow or lend
English rice	Malay	padi (unharvested grain) beras (mooched) nasi (cooked) emping (mashed) pulut (glutinous) bahar (cooked as a gnef)
English wear	Japanese	kimu (generic) Asone (coat or jacket) haku (shoes or trousers) kaburu (hat) Asunaru (ring or gloves) shimera (belt or tie or scarf) tsukeru (brooch or clip) kakuru (glasses or necktie)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Ambiguità semantica: prospettiva (cross-)linguistica

- **Possibili ambiguità:**

- **polisemia**

Mario lava i *piatti*

Mario serve i *piatti* che ha preparato

I Red Hot Chili Pepper hanno *lanciato* un nuovo disco

Mario ha *lanciato* il bastone al cane

Mario ha *lanciato* i *piatti* dalla finestra

- **uso metaforico**

Mario ha *mangiato* la *foglia*

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (rule-to-rule)

- l’approccio **semantico compositazionale** ignora il problema dell’ambiguità
- in un **approccio integrato (rule-to-rule)** le metodologie di disambiguazione fanno uso principalmente di due strumenti:
  - **restrizioni selettive** (Katz e Fodor 1963, Hirst 1987)  
bloccare le violazioni di restrizione del tipo  
piatto<sub>portata</sub> > edibile, cucinabile ...  
piatto<sub>scodella</sub> > lavabile ...
  - **gerarchie di tipi**  
sfruttare dei vincoli derivanti dalla rete semantica formata dai concetti (es. Wordnet)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (rule-to-rule)

- **restrizioni sulla selezione verbale**
  - servire<sub>di portata</sub> > oggetto<sub>[edibile]</sub>
  - lanciare<sub>di evento</sub> > oggetto<sub>[eventivo]</sub>
- **prerequisiti**

analisi sintattica (almeno gruppi verbali e sottocategorizzazione) + specificazione di tratti semantici
- **problemi**
  - contesti troppo generali per poter applicare questo tipo di selezioni: es. che tipo di *piatto* ci raccomanda?
  - usi metaforici: violazioni produttive delle restrizioni semantiche (es. mangiare > oggetto<sub>[edibile]</sub> in *mangiare la foglia*; oppure bere un bicchier d'acqua)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

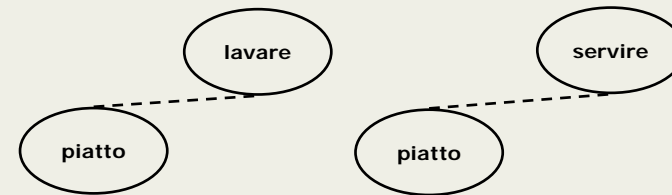
Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (rule-to-rule)

- **vincoli derivanti dalla gerarchia semantica**

associazione probabilistica tra il predicato e la classe che domina tale predicato
- **selectional association (Resnik 1997)**

sfrutta la relazione di iperonimia in WordNet combinata con le informazioni recuperate da un corpus taggato



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (stand alone)

- **approccio statistico**

l'idea è quella di acquisire informazioni semplicemente dalle parole (bag of words) piuttosto che da pre-analisi linguistiche
- **feature vectors**
  - input: parole target (parole da disambiguare) + contesto (porzione di testo intorno ai target)
  - es. Marco lava i piatti con il detersivo
  - rappresentazione tratti in un vettore:
    - **collocazione** (finestra ad es. di 3 parole prima e 3 parole dopo il target)
    - [Marco, N, lava, V, i, D, con, P, il, D, detersivo, N]
    - **co-occorrenza** (presenza di parole altamente frequenti in testi campione usati come training per la disambiguazione)
    - guanti, lavare, acqua, detersivo > [0, 1, 0, 1]

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (stand alone)

- massima probabilità di un significato  $s$  (appartenente ad un insieme  $S$  di significati possibili), dato un vettore di tratti  $V$ :

$$\hat{S} = \operatorname{argmax}_{s \in S} P(s|V)$$

- secondo il teorema di Bayes:

$$\hat{S} = \operatorname{argmax}_{s \in S} \frac{P(V|s) \cdot P(s)}{P(V)}$$

- Naive Bayes classifier:

$$P(V|s) \approx \prod_{j=1}^n P(v_j|s) \quad \hat{s} = \operatorname{argmax}_{s \in S} P(s) \prod_{j=1}^n P(v_j|s)$$

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (stand alone)

- **Decision list classifiers**

parole chiave collegate da operatori booleani costituiscono regole di classificazione:

regole	significato
detersivo, ¬pasta ⇒	piatto <sub>scodella</sub>
lavare, lavastoviglie ⇒	piatto <sub>scodella</sub>
pasta, carne ⇒	piatto <sub>portata</sub>
servire, gustoso ⇒	piatto <sub>portata</sub>
...	

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (stand alone)

- **bootstrapping by seeds**

l'idea è quella di iniziare l'apprendimento riducendo il training set ad un insieme limitato di esempi prototipici rispetto ad un determinato senso per poi cercare di generalizzare su un corpus più grande

testo prototipico	significato
can che abbaia non morde ⇒	cane <sub>animale</sub>
caricò il cane prima di premere il grilletto della pistola ⇒	cane <sub>parte_pistola</sub>
...	

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione (stand alone)

- Apprendimento non supervisionato per **agglomerative clustering**
- Approcci basati su **dizionari di macchina**
  - sfruttare le definizioni delle parole come vettori di significato (Lesk 1986); accuratezza del 50-70%
  - considerare nel vettore anche le parole nella cui definizione si ritrova la parola target (Morris 1985)
  - utilizzare, ove presenti, anche i codici tematici (subject codes, Guthrie e al. 1991); accuratezza del 47-72%

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Modello spazio vettoriale

- ogni vettore è rappresentativo dei termini  $t$  presenti nella query

$$\vec{q}_k = (t_{1,k}, t_{2,k} \dots t_{n,k})$$

- e nei documenti

$$\vec{d}_j = (t_{1,j}, t_{2,j} \dots t_{n,j})$$

( $n$  è il numero totale di termini nella collezione,  $t$  può essere presente, 1, o assente, 0 nel vettore; Salton 1971)

- la rilevanza di un documento rispetto ad una query è data dal numero di termini  $t$  in comune

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Modello spazio vettoriale

- un vettore può avere **valori pesati** anziché binari, che esprimono l'importanza di ciascun termine

→

$$q_k = (w_{1,k}, w_{2,k} \dots w_{n,k})$$

→

$$d_j = (w_{1,j}, w_{2,j} \dots w_{n,j})$$

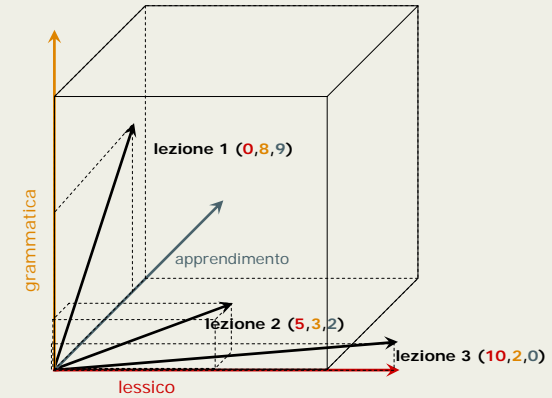
- matrice di termini per documento (term-by-document matrix)**

	$d_1$	$d_2$	...	$d_j$
$t_1$	$w_{1,1}$	$w_{1,2}$		$w_{1,j}$
$t_2$	$w_{2,1}$	$w_{2,2}$		$w_{2,j}$
...				
$t_i$	$w_{i,1}$			$w_{i,j}$

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Modello spazio vettoriale



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Modello spazio vettoriale

- Normalizzazione**  
conversione di tutti i vettori ad una lunghezza standard: ad esempio dividendo ogni dimensione per la lunghezza totale del vettore

$$A = \begin{pmatrix} 0 & .81 & .98 \\ .67 & .49 & .20 \\ .75 & .32 & 0 \end{pmatrix} \begin{matrix} \text{lessico} \\ \text{grammatica} \\ \text{apprendimento} \end{matrix}$$

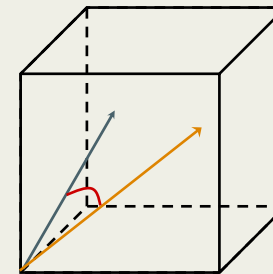
lezione 1  
lezione 2  
lezione 3

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Distanza tra vettori

- Prodotto scalare**  
la misura della similarità tra due vettori è data dal prodotto scalare tra i due vettori normalizzati



$$\text{sim}(\vec{q}_k, \vec{d}_j) = \vec{q}_k \bullet \vec{d}_j = \sum_{i=1}^n w_{i,k} \times w_{i,j}$$

coseno: da 0 (doc. ortogonali)  
a 1 (doc. identici)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Selezione del peso specifico dei relativi termini

- **frequenza**  
quante volte un termine appare in un testo (lista delle occorrenze)
- **importanza**  
meno volte un termine appare nell'intera collezione, più significativa è la sua presenza all'interno di un testo
- **stop list**  
lista di parole ad altissima frequenza (parole funzionali tipo articoli, povere di contenuto semantico) che vengono escluse dai vettori ("to be or not to be" > "not", Frakes and Baeza-Yates 1992)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

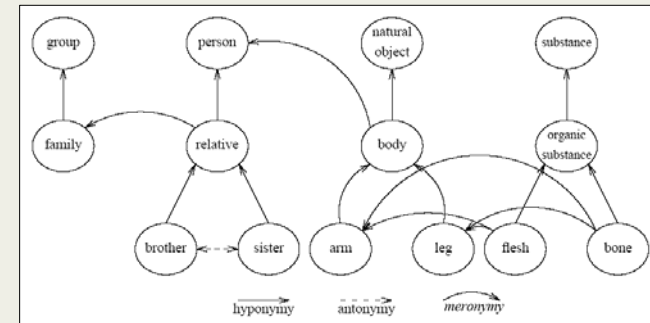
Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione con reti semantiche (lezione 3)

(... fine)

Lessico e analisi morfologica

- Esempio di relazioni semantiche (Miller 1993)

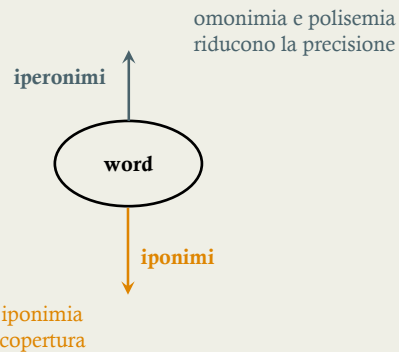


Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Disambiguazione con reti semantiche

- usare WordNet per la disambiguazione:



Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì

## Prossima lezione

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

- **Laboratorio su ontologie e disambiguazione**
  - esplorazione di wordnet
    - word sense disambiguation
    - sottocategorizzazione verbale (Levin 93)

Lezione 7 – Rappresentazione della conoscenza e disambiguazione

Ling. Comp. A.A.2007/08 - C. Chesì