
ANNO ACCADEMICO 2024/2025

Tecnologie del Linguaggio Naturale

Teoria - Radicioni

Altair's Notes



UNIVERSITÀ
DI TORINO



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1	SEMANTICA LESSICALE	PAGINA 5
1.1	Introduzione Programma della Seconda Parte del Corso — 5 • Che cos'è la semantica lessicale? — 5 • Alcune Definizioni di Ontologia — 6	5
1.2	Design di Ontologie DOLCE — 8	8
CAPITOLO 2	KNOWLEDGE REPRESENTATION	PAGINA 11
2.1	Knowledge Bases Rappresentazioni Strutturate — 11	11
2.2	Rappresentazioni Gerarchiche Eredità delle Proprietà — 14	13
2.3	Frame I Frame per Rappresentare la Conoscenza — 15 • Strutture dei Frame — 16	15
2.4	Teorie del Significato Il Significato delle Parole — 17 • Calcolo Sintagmatico del Significato — 19	16
CAPITOLO 3	WORDNET E FRAME.NET	PAGINA 21
CAPITOLO 4	NATURAL LANGUAGE TOOLKIT (NLTK)	PAGINA 23
CAPITOLO 5	NGRAMS	PAGINA 25
CAPITOLO 6	UN'INTRODUZIONE AI TRANSFORMERS	PAGINA 27
CAPITOLO 7	SEMINARI	PAGINA 29
7.1	Eye Tracking	29
7.2	Dementia Detection	29

Premessa

Licenza

Questi appunti sono rilasciati sotto licenza Creative Commons Attribuzione 4.0 Internazionale (per maggiori informazioni consultare il link: <https://creativecommons.org/version4/>).



Formato utilizzato

Box di "Concetto sbagliato":

Concetto sbagliato 0.1: Testo del concetto sbagliato

Testo contenente il concetto giusto.

Box di "Corollario":

Corollario 0.0.1 Nome del corollario

Testo del corollario. Per corollario si intende una definizione minore, legata a un'altra definizione.

Box di "Definizione":

Definizione 0.0.1: Nome delle definizioni

Testo della definizione.

Box di "Domanda":

Domanda 0.1

Testo della domanda. Le domande sono spesso utilizzate per far riflettere sulle definizioni o sui concetti.

Box di "Esempio":

Esempio 0.0.1 (Nome dell'esempio)

Testo dell'esempio. Gli esempi sono tratti dalle slides del corso.

Box di "Note":

Note:-

Testo della nota. Le note sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive.

Box di "Osservazioni":

Osservazioni 0.0.1

Testo delle osservazioni. Le osservazioni sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive. A differenza delle note le osservazioni sono più specifiche.

1

Semantica Lessicale

1.1 Introduzione

1.1.1 Programma della Seconda Parte del Corso

Per quanto riguarda questa parte del corso ci si concentrerà sulla semantica lessicale, in particolare sui seguenti punti:

- Problema della Knowledge Representation.
- WordNet e BabelNet.
- FrameNet.
- N-grams.
- Word/Sense embeddings.
- Encoder/Decoder.

1.1.2 Che cos'è la semantica lessicale?

Definizione 1.1.1: Semantica Lessicale

La semantica lessicale è lo studio del significato delle parole e delle loro relazioni. Lo studio di cosa i singoli oggetti lessicali significano, cosa fanno, come possono essere rappresentati e combinati.

Sono scelte le ontologie per:

- *Parte metodologica*: adottare un piano molto interdisciplinare a un alto livello di generalità.
- *Parte architetturale*: centralità del ruolo che un'ontologia può giocare in un sistema informativo.

Le parole sono flessibili:

- Capacità di ripetere le parole nei *sensi* già noti, combinarle in enunciati anche nuovi, capacità di estendere una parola o una frase per esprimere nuovi sensi:
 - Queste capacità sono ereditate dagli esseri umani.
 - Non si sviluppano in ogni circostanza.
- La cooperazione tra le capacità del punto precedente è importante per sviluppare la *metalinguistica*:
 - Che cosa significa questa parola? Che cosa vuol dire? Si può dire così? Si scrive gocce o gocce?
 - Le parole hanno funzione riflessiva: si usa la *lingua per riflettere sulla lingua*.

Rigidità vs. Deformabilità:

- La chimica e la matematica parlano di determinati aspetti dell'esperienza (sono *rigide*).
- Le lingue invece sono *deformabili*: se ne può alterare o dilatare il significato.

1.1.3 Alcune Definizioni di Ontologia

Domanda 1.1

Cosa significa il termine ontologia?

Definizione 1.1.2: Ontologia Filosofica

In filosofia il termine ontologia indica lo studio dell'essere e delle sue categorie fondamentali. Un'ontologia definisce un insieme di primitive rappresentazionali con il quale modellare un dominio di conoscenza o di discorso. È un sistema organizzato in categorie e relazioni. Le loro definizioni includono informazioni sul loro significato e vincoli su come applicarle in maniera consistente.

Note:-

Per esempio una categoria rappresenta tutti i tipi di entità che possono fungere da soggetto in un predicato.

Problema: questa definizione ammette che quasi tutto può essere considerato un'ontologia.

Definizione 1.1.3: Ontologia come Concettualizzazione

Un'ontologia è una specifica esplicità di una concettualizzazione. Una struttura formale di un pezzo della realtà come percepito e organizzato da un agente indipendentemente da:

- Vocabolario utilizzato.
- L'occorrenza in una situazione specifica.

Note:-

Situazioni diverse coinvolgenti lo stesso oggetto descritte da un vocabolario diverso possono avere la stessa concettualizzazione.

Ontologia e semantica:

- Un'ontologia riguarda ciò che c'è, la semantica si riferisce a ciò che c'è.
- Differenti aspetti del linguaggio hanno differenti ruoli nell'ontologia.

Ontologie e Knowledge Bases:

- Componente *terminologica* (ontologia):
 - Indipendente da un particolare stato.
 - Pensata per supportare servizi terminologici.
- Componente *asserzionale*:
 - Riflette specifici stati.
 - Pensata per problem solving.

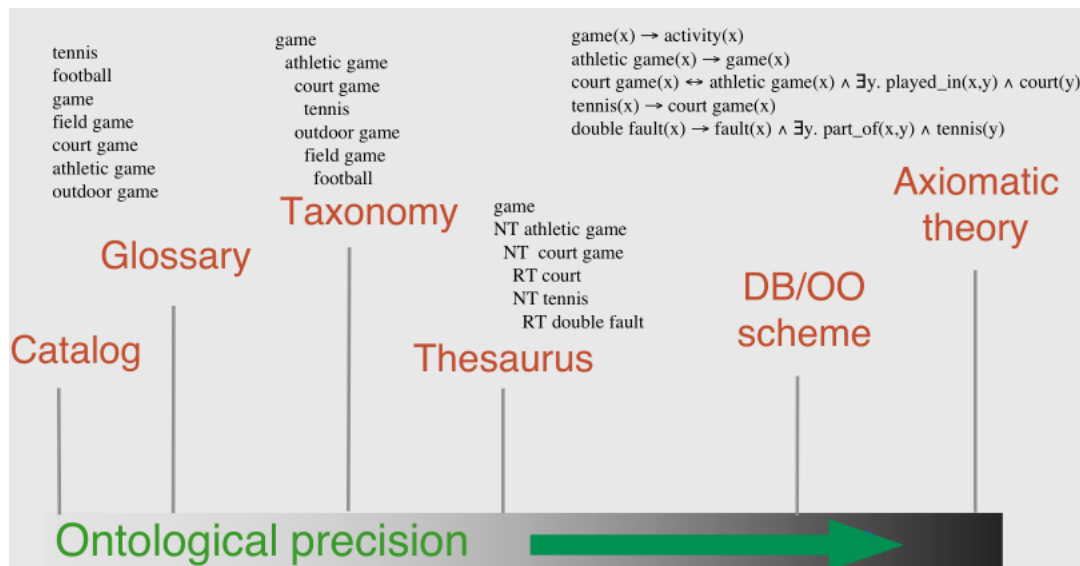


Figure 1.1: Livelli di precisione delle ontologie.

Ontologia e lessico:

- Livello lessicale: il *lessico* è un elenco di parole in una lingua (vocabolario + conoscenza sull'utilizzo delle parole).
- Relazioni lessicali: sinonimia (termini con lo stesso significato), iponimia (relazione di sottoclasse), iperonimia (relazione di superclasse), meronimia (costituente di qualcosa), olonimia (intero), antonimia (termini di significato opposto).

Caratteristiche del lessico:

- *Overlapping Word Senses:*
 - Nelle ontologie le sottocategorie di una categoria generale sono *mutualmente esclusive*.
 - Nel lessico esistono *sovrapposizioni di significato* (e.g. quasi sinonimi).
- *Buchi nel lessico:*
 - Il lessico di un linguaggio omette riferimenti a *categorie ontologiche* che non sono lessicalizzate in quel linguaggio.
 - Sono categorie che richiedono perifrasi (giri di parole) per essere definite.
- *Lessici tecnici:* nei contesti tecnici il linguaggio è molto vicino all'ontologia del dominio.

Domanda 1.2

Perché costruire ontologie?

- *Condivisione* della comprensione delle entità di un certo dominio.
- *Riutilizzo* dei dati e dell'informazione.
- Creare comunità di ricercatori.

1.2 Design di Ontologie

Entità ed Eventi:

- **Entità:** oggetti che continuano per un periodo di tempo mantenendo la propria identità.
- **Eventi:** oggetti che accadono, si svolgono o si sviluppano nel tempo.

Definizione 1.2.1: Ontologie Fondazionali

Un'ontologia fondazionale cattura un insieme di distinzioni base valide in vari domini.

Note:-

Alcune celebri sono DOLCE, SUMO e CYC.

1.2.1 DOLCE

Definizione 1.2.2: DOLCE

DOLCE è un'ontologia fondazionale con lo scopo di rappresentare le strutture concettuali di base che emergono dal linguaggio naturale e dalla cognizione umana.

Scelte alla base di DOLCE:

- **Endurant:** entità che sono completamente presenti in ogni momento della loro esistenza.
- **Perdurant:** eventi che si estendono nel tempo e sono parzialmente presenti in ogni istante.
- **Quality:** proprietà specifiche di un'entità, dipendenti da essa.
- **Approccio moltiplicativo:** differenti oggetti ed eventi possono essere co-localizzati nello spazio-tempo.

Definizione 1.2.3: Criteri di Identità

I criteri di identità sono utilizzati per valutare se due entità sono in relazione di sottoclasse e sono proprietà necessarie delle entità confrontate.

Note:-

Per esempio "intervallo di tempo" non può essere sottoclasse di "durata di tempo" perché utilizzano criteri diversi.

Endurants vs. Perdurants in DOLCE:

- Gli endurants possono cambiare nel tempo: lo stesso endurant può avere proprietà incompatibili in tempi diversi.
- I perdurants non cambiano, hanno una locazione spaziale ben definita dagli endurants che vi partecipano.
- La relazione tra endurants e perdurants è la **partecipazione**: un endurant vive partecipando in qualche perdurant.

Qualities e Quality regions in DOLCE:

- Le qualities possono essere viste come entità base che possono essere percepite e misurate (e.g. forma, colore, suono, etc.).
- Le qualities sono caratteristiche per specifici individui.
- Si distingue tra una qualità e il suo valore (e.g. il colore di una specifica rosa e quale sia la particolare sfumatura di rosso).

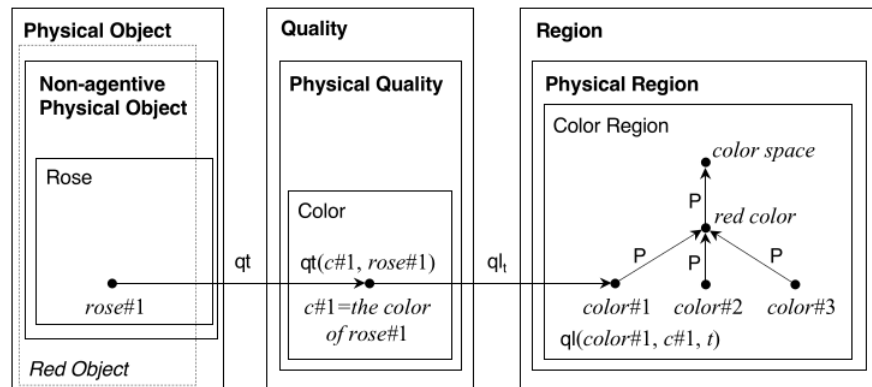


Figure 1.2: Rappresentazione del colore di una rosa.

2

Knowledge Representation

2.1 Knowledge Bases

2.1.1 Rappresentazioni Strutturate

Durante la prima fase iniziale dell'IA ci si è concentrati sullo sviluppo di *metodi generali* per la risoluzione di problemi¹. Successivamente, dalla seconda metà degli anni '60, si vollero abbandonare i domini astratti e semplificati per passare a problemi reali in cui era richiesta una *conoscenza sul mondo in cui il sistema opera*.

Linguaggio e Operazioni:

- Vengono studiati i formalismi adatti a rappresentare le conoscenze necessarie.
- Un sistema di conoscenza deve consistere di:
 - Un *linguaggio di rappresentazione*: un insieme di strutture sintattiche adatte a codificare le informazioni da rappresentare.
 - Un *insieme di regole o di operazioni*: per manipolare il linguaggio.
- L'implementazione delle regole deve portare a *inferenze desiderate* e le regole devono poter essere espresse come *procedure*.

Strutturazione dell'informazione:

- Per rappresentare conoscenze è possibile utilizzare *formule logiche* rappresentanti proposizioni indipendenti.
 - Con le formule logiche non è possibile collegare le varie formule, organizzandole in blocchi omogenei.
 - La logica necessita di ulteriori meccanismi computazionali.
- Esistono *formalismi* per *aggregare conoscenze elementari* in strutture più complesse.
- SI deve poter accedere alla struttura in cui le conoscenze relative all'oggetto in questione sono direttamente disponibili.

Definizione 2.1.1: Reti Semantiche

Le reti semantiche sono un formalismo nato dai primi progetti di traduzione automatica. Hanno la caratteristica comune di utilizzare una struttura a grafo (rete) in cui i nodi rappresentano dei concetti e gli archi rappresentano relazioni tra concetti o proprietà dei concetti.

¹Tali metodi erano indipendenti da specifici domini.

Corollario 2.1.1 Grafi Relazionali

I grafi relazionali sono le reti semantiche più semplici. Sono costituiti da grafi relazionali che permettono di descrivere le relazioni tra le diverse entità del grafo stesso.

Note:-

Per esempio lo si può usare per descrivere uno stato nel mondo dei blocchi.

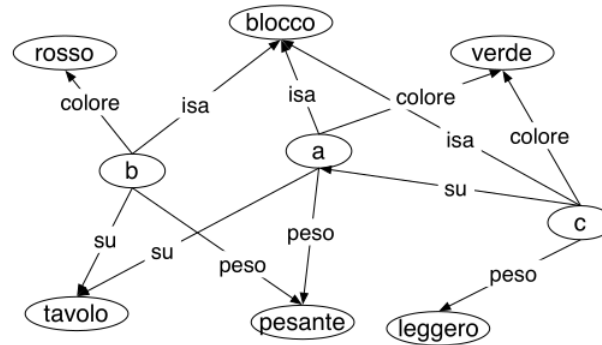


Figure 2.1: Rappresentazione del mondo dei blocchi mediante grafo relazionale.

Osservazioni 2.1.1

- Il grafo è composto da diversi *nodi* ognuno rappresentante un'entità.
- Da ciascun nodo si dipartono *archi* che lo collegano ad altri, sono *etichettati* in modo da esplicitare la relazione che intercorre tra i nodi collegati.
- Una relazione importante è *isa*: il tipo di concetto che un nodo rappresenta.

Espressività di un grafo relazionale:

- È implementato un *sottoinsieme del calcolo dei predicati del primordine*: gli archi sono i predicati e i nodi sono i termini.
- *Limitazioni di efficacia espressiva*: i grafi rappresentano la congiunzione in maniera implicita.
- È difficile rappresentare disgiunzione o implicazione.
- Inoltre è complicato esprimere la quantificazione universale.
- Le relazioni espresse dagli archi sono per loro natura binarie ma i predicati logici possono avere qualsiasi arietà.
- Una possibile soluzione è quella di tradurre tutte le relazioni in relazioni binarie:
 - Accresce la granularità (e quindi l'espressività) e richiede l'introduzione di nodi per rappresentare oggetti e insiemi di oggetti e situazioni e azioni.
 - I predicati con arietà superiore a 2 esplodono in una serie di relazioni binarie: una che chiarisce il tipo di predicato, altre che esplicitano ruolo e funzione degli argomenti.

Note:-

Anche se una relazione ad alta granularità e una come formula del primordine formano un isomorfismo mettono in luce aspetti differenti.

Corollario 2.1.2 Reti Proposizionali

Le reti proposizionali sono reti semantiche i cui nodi possono rappresentare non solo entità semplici, ma intere proposizioni.

Note:-

Ammettendo la possibilità di avere nodi proposizionali si accresce l'espressività del linguaggio. Sono state proposte reti fortemente legate alla logica del primordine.

La negazione:

- Può essere rappresentata mediante un arco che collega il risultato della negazione con la proposizione che viene negata.
- Si possono rappresentare idee piuttosto articolate e distinguere tra:
 - *Negazione di un'intera proposizione.*
 - *Negazione di una proposizione incassata* all'interno di un'altra proposizione.

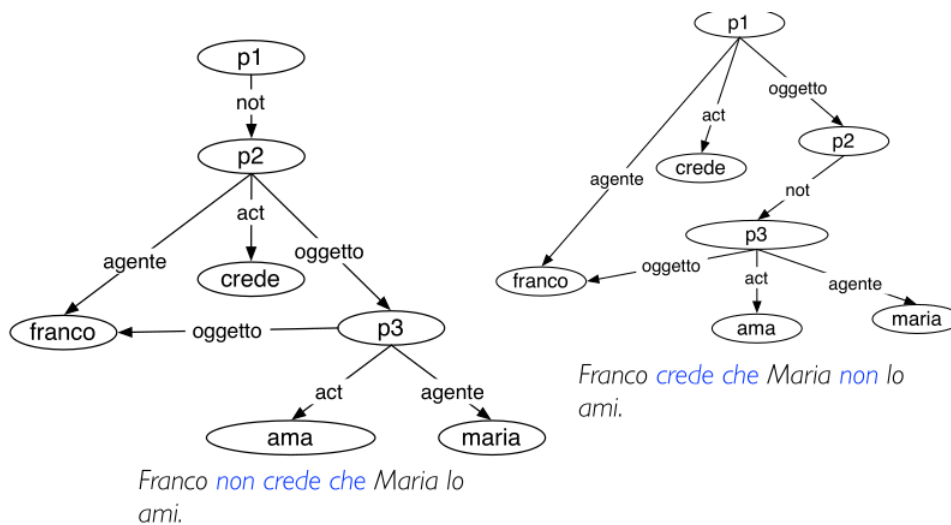


Figure 2.2: Introduzione della negazione.

Disgiunzione:

- Viene introdotta dopo la negazione.
- Questo perché la congiunzione è implicita e si può rappresentare la disgiunzione in sua funzione mediante le *leggi di De Morgan*.

Note:-

La scelta di quale rete utilizzare va effettuata in base a leggibilità, flessibilità, efficienza, facilità di espressione, etc.

2.2 Rappresentazioni Gerarchiche

Molte delle conoscenze sono organizzate in *gerarchie*. Varie entità sono raggruppate in *classi* che a loro volta sono raggruppate in *sottoclassi*. Queste gerarchie non sono limitate a oggetti, ma si possono estendere ad azioni (e.g. camminare, marciare, etc.), eventi, stati, proprietà, etc. Per esempio, nelle reti semantiche se si vuole esprimere l'idea "gli elefanti sono mammiferi" è sufficiente un nodo per gli elefanti e uno per i mammiferi connessi da un arco etichettato isA. Così facendo si può fare inferenza, sfruttando proprietà delle relazioni (la relazione isA è transitiva).

2.2.1 Eredità delle Proprietà

Definizione 2.2.1: Eredità delle Proprietà

Il meccanismo di eredità delle proprietà afferma che le proprietà asserite per un nodo valgono anche per i nodi che si trovano al livello inferiore della gerarchia.

Se la rete che rappresenta le diverse proprietà isA è un **albero** è facile stabilire se un concetto X gode della proprietà p :

- Per capire se $p(X)$ è vero basta considerare gli antenati di X e vedere se qualcuno gode di p .
- Questa ricerca è più efficiente dei processi di inferenza.

Vantaggi:

- **Economia di rappresentazione:** invece di replicare una proprietà per tutti i nodi essa viene asserita solo al livello più alto in cui si applica.
- **Semplifica la manutenzione:** una modifica a una proprietà richiede una sola operazione.

Trattamento delle eccezioni:

- Si può risolvere con un approccio procedurale.
- Esempi: gli uccelli solitamente volano, ma alcuni no. I mammiferi partoriscono i figli, ma l'ornitorinco no.

Definizione 2.2.2: Validità per Default

Vengono rappresentate conoscenze che valgono fino a prova contraria. Le eccezioni vengono memorizzate in corrispondenza dei nodi a cui si riferiscono.

Note:-

In questo caso si ottiene l'eccezione prima di raggiungere il nodo di Default.

Definizione 2.2.3: Ereditarietà Multipla

Nel caso una classe abbia più di una classe superordinata ($X \text{ isA } Y, X \text{ isA } Z$ con $Y \neg = Z$) la rete semantica passa da albero a grafo.

Note:-

Reti Semantiche 1, Java 0. *If you know, you know.*

Ammettendo l'ereditarietà multipla:

- Il tempo di ricerca passa da lineare a esponenziale, non esistono euristiche.
- Nel caso di un **conflitto di valori** non è possibile stabilire un criterio risolutivo che abbia carattere generale.
- Diverse strategie:
 - Ricerca in profondità.
 - Ricerca in ampiezza.

Problemi:

- ✗ Le tecniche di ricerca non permettono di giungere a risultati intuitivamente corretti.
- ✗ I risultati sono inconsistenti. Esempio con Nixon Diamond: Nixon è sia pacifista che guerrafondaio.

Definizione 2.2.4: Dissonanza Cognitiva

Si può considerare la rete stessa come ambigua e in grado di esprimere più interpretazioni: ogni interpretazione è consistente con sé stessa, ma diventa inconsistente quando viene considerata con le altre.

Appartenenza e Inclusione:

- Non esiste una distinzione tra nodi che rappresentano individui e nodi che rappresentano classi o insiemi di individui.
- Il legame isA viene usato per denotare sia l'*appartenenza* (elemento in insieme) che l'*inclusione* (insieme in insieme).
- Non si può distinguere tra proprietà vere per tutti gli individui di una classe e proprietà vere della classe in quanto tale.
- È impossibile separare la semantica di una rete dal suo utilizzo.

2.3 Frame

Le persone utilizzano un insieme strutturato di conoscenze per interpretare le diverse situazioni che si trovano a dover affrontare. Quando ci si imbatte in una situazione si recupera una *rappresentazione a carattere generale* che viene poi raffinata e modificata.

Definizione 2.3.1: Frame

Un frame è una struttura che rappresenta le conoscenze di carattere generale che un individuo ha riguardo situazioni, luoghi, oggetti, etc. Fornisce una cornice concettuale all'interno della quale vengono interpretati nuovi dati alla luce delle conoscenze derivate dall'esperienza precedente.

Caratteristiche:

- Un sistema che usa i frame può *formulare previsioni* e *avere aspettative*.
- Aiuta il processo di interpretazione di situazioni ambigue.
- Vengono facilitati il recupero di informazioni e i processi inferenziali.

Corollario 2.3.1 Script

Lo script è una struttura di conoscenza di alto livello che integra e organizza le frasi.

2.3.1 I Frame per Rappresentare la Conoscenza

Mancanza di una semantica formale:

- I frame rappresentano la conoscenza in maniera dichiarativa, ma senza una semantica formale.
- Si presuppone l'esistenza di procedure in grado di utilizzare le informazioni contenute nei frame.

Concetti nei frame:

- Livello di base: costituisce il modo naturale di categorizzare gli oggetti e le entità.
- Livello superordinato/subordinato: rispettivamente generalizzazioni e specializzazioni di concetti di base.
- **Prototipi**: membri tipici di una categoria. L'appartenenza viene categorizzata in maniera di maggiore o minore somiglianza rispetto ai prototipi (e.g. passero è migliore rappresentante di uccello rispetto ad airone che è migliore rispetto a struzzo).

2.3.2 Strutture dei Frame

Strutture gerarchiche:

- Gli elementi di un frame sono collegati da relazioni isA o **ako** (a kind of) che consentono l'ereditarietà.
- Le proprietà ad alto livello sono fisse, mentre ai livelli più bassi (sottoclassi o istanze individuali) possono essere specializzate.

Domanda 2.1

Com'è fatto un frame?

- Ha un nome univoco.
- Le caratteristiche sono rappresentate da un insieme di slot (caselle in cui viene inserito un determinato tipo di informazione).
- Valori di default per gli slot.
- Conflitti tra valori ereditati in caso di ereditarietà multipla.
- Uno slot può essere una struttura complessa.
- Procedure per rendere la computazione efficiente:
 - If needed: calcolo del valore di uno slot.
 - If added: solo quando si tenta di riempire uno slot.

2.4 Teorie del Significato

Alcune teorie tipiche:

- Prototipi: come accennato in precedenza, si effettuano approssimazioni per rappresentare una categoria.

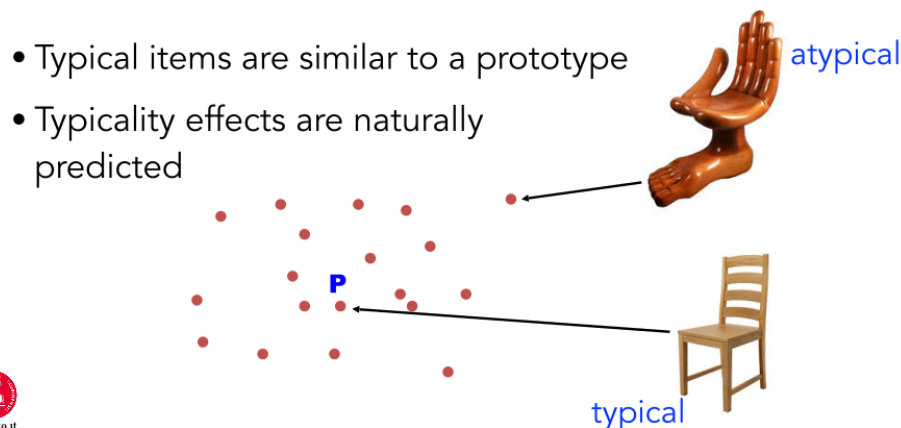


Figure 2.3: Teoria dei prototipi.

- Esempi: la rappresentazione mentale di un concetto è l'insieme di alcuni esempi di quella categoria.
- Teoria: i concetti sono parte della comprensione del significato, collegati ad altri concetti.

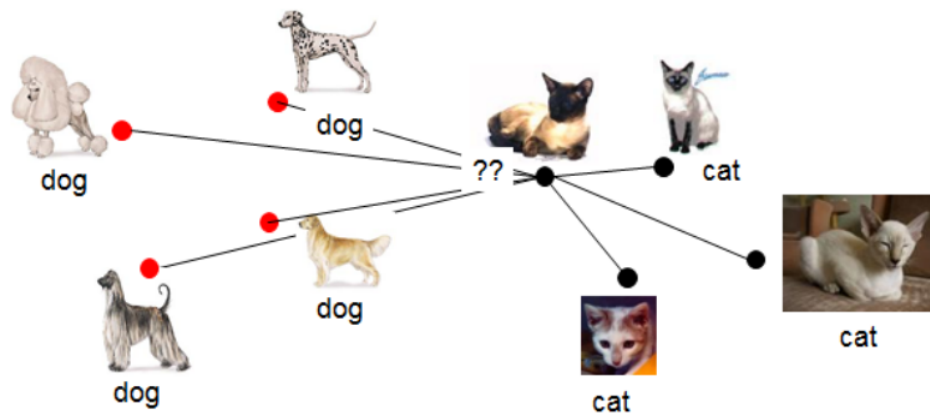


Figure 2.4: Teoria degli esempi.

Note:-

Quelle elencate qui sopra sono proprio tre teorie (probabilmente non avevano voglia di trovare nomi migliori).

Approccio duale:

- *Sistema 1* (implicito): categorizzazione non monotona, rappresentazione continua.
- *Sistema 2* (esplicito): categorizzazione monotona, rappresentazione dei dati proposizionale.

2.4.1 Il Significato delle Parole**Due problemi:**

- Polisemia.
- Semantica frasale.

Definizione 2.4.1: Contesto

Il contesto è l'insieme degli elementi adiacenti a una parola. Può essere:

- Sintattico: elementi adiacenti a una parola dal punto di vista delle loro proprietà sintattiche.
- Semantico: elementi adiacenti a una parola dal punto di vista delle loro proprietà semantiche.
- Linguistico.
- Situazionale (pragmatico): richiede una conoscenza del mondo esterna alla lingua.

Ambiguità e Polisemia:

- L'*ambiguità* è la proprietà di una forma lessicale di avere più di un significato:
 - *Contrastativa* (o omonimia): due significati contraddittori.
 - *Complementare* (o polisemia): stessi significati in contesti diversi.
- *Principio di economicità linguistica*: si usano le stesse parole per esprimere più di un significato, per contenere le dimensioni del lessico, etc.

Note:-

Solitamente i verbi tendono a essere più polisemici dei sostantivi.

Definizione 2.4.2: Teoria Referenziale del Significato

Le parole sono uno strumento attraverso il quale facciamo riferimento a ciò che esiste. Il significato delle parole consiste nella loro capacità di stabilire una relazione con elementi della realtà al di fuori della lingua.

Il riferimento può realizzarsi attraverso due procedimenti:

- *Denotazione*: si indica la classe di elementi.
- *Designazione*: si indica un particolare elemento della classe.

Interpretazioni:

- Ampia: ciascun elemento della lingua istituisce un riferimento con la realtà extralinguistica.
- Restrittiva: distingue tra atto di riferimento e atto di predicazione.

Definizione 2.4.3: Teoria Mentalista

La teoria mentalista (o concettuale) arricchisce la teoria referenziale con l'ipotesi che il riferimento tra parole e realtà non sia diretto, ma mediato dall'immagine mentale dei concetti. Le parole non fanno riferimento diretto alla realtà extralinguistica, ma soltanto al modo in cui tale realtà è concettualizzata e categorizzata nella mente del parlante.

Mediazione concettuale:

- Si può parlare non soltanto di entità esistenti o di eventi che accadono, ma anche di cose astratte, immaginarie o ipotetiche.
- L'accento è sugli aspetti psicologici e sul legame tra concettualizzazione ed esperienza psico-percettiva.

Concetti e lessicalizzazione:

- I concetti cognitivi sono entità instabili: possono differire individualmente e culturalmente:
 - Appartengono alla struttura mentale.
 - Possono essere considerati come degli universali.
- I concetti lessicalizzati sono più stabili: individualmente e socialmente condivisi:
 - Appartengono alla struttura linguistica.
 - Variano da lingua a lingua.

Definizione 2.4.4: Teoria Strutturale

Il significato dei termini non è il suo riferirsi a un oggetto, ma nel valore che la parola assume in relazione alle altre parole presenti nella lingua che fanno parte dello stesso campo semantico.

Note:-

Il valore semantico di un termine è il suo contenuto informativo.

Definizione 2.4.5: Teoria Distribuzionale

Il significato delle parole è determinato in larga misura dall'insieme di altre parole con cui queste co-occorrono. Questa teoria è ritornata di moda di recente grazie all'enorme disponibilità di corpora.

Corollario 2.4.1 Metafora Geometrica del Significato

I significati delle parole corrispondono a punti in uno spazio multidimensionale, in maniera tale che a punti vicini corrispondono parole con significato prossimo.

Note:-

Questo è alla base delle rappresentazioni vettoriali e della cosine similarity.

2.4.2 Calcolo Sintagmatico del Significato

Problemi fondamentali delle teorie del significato:

- Contestualità del significato.
- Polisemia.

Definizione 2.4.6: Principio di Composizionalità del Significato

Spiega come il significato degli enunciati si formi a partire dal significato degli elementi lessicali che li compongono.

Osservazioni 2.4.1

In alcuni casi questo principio viene meno:

- Espressioni idiomatiche.
- Usi metaforici.
- Polisemia.

Possibili approcci ai problemi:

- *Enumerazione dei sensi*: i diversi sensi sono elencati nella parola, specificano i contesti in cui i diversi significati possono attivarsi.
- *Concezione dinamica*: le parole vengono concepite come entità permeabili e il significato di ciascuna interagisce con il significato delle parole adiacenti.

Principi di interazione semantica:

- Co-composizione: il significato di un verbo può essere determinato dai suoi argomenti. Ciascun verbo ha una parte di base che non cambia, ma viene ridefinita e specializzata dal suo complemento.
- Forzatura o conversione di tipo: un verbo in combinazione con un nome specifico lo spinge a significare determinate cose anche variandone il tipo semantico.
- Legamento selettivo: l'aggettivo può selezionare una specifica porzione del significato del nome.

3

WordNet e FrameNet

4

Natural Language ToolKit (NLTK)

5

Ngrams

6

Un'Introduzione ai Transformers

7

Seminari

7.1 Eye Tracking

7.2 Dementia Detection

