





Non è un
dizionario

Il primo punto che dobbiamo capire è perché un dizionario classico non può funzionare per svolgere il nostro task di ricerca di significati.

In un dizionario tipico i termini sono organizzati in ordine lessicografico mentre noi vogliamo organizzare il lessico in gruppi di significati così da poter ottenere da un certo termine il suo significato (e fin qui tutto bene) e dal suo significato tutti i significati in qualche modo correlati (cosa assai complessa da recuperare da un dizionario classico).

Un concetto dispone sia di una veste lessicale che di un significato vero e proprio.

Il nostro scopo è quello di ottenere, per ogni significato, la lista di vesti lessicali che quel significato può assumere (*quindi la lista di parole della lingua inglese che hanno quel significato*).

Il tutto è facilmente rappresentabile mediante una **lexical matrix**

Word Meanings	Word Forms				
	F ₁	F ₂	F ₃	. . .	F _n
M ₁	E _{1,1}	E _{1,2}			
M ₂		E _{2,2}			
M ₃			E _{3,3}		
⋮				⋮	
M _m					E _{m,n}

Es. di lexical matrix. Un'entry (es. E_{1,1}) in una cella implica che la forma sulla colonna della cella sia utilizzabile per descrivere il significato sulla riga della cella.

Ad esempio E_{1,1} indica che F₁ è una forma (*un termine*) che ha come significato M₁

Se ne deduce che se due entry hanno la stessa colonna ma riga diversa la loro forma è **polisemica**

Se due entry hanno la stessa riga ma una colonna diversa le loro forme sono **sinonimi**

Lexical matrix

Synonym sets **synset**

I **synonym sets** (detti **synset** / *significati di una parola*) sono immediatamente definibili a partire da una *lexical matrix*.

Infatti, il synset di un certo significato M_i sarà l'insieme delle forme delle entry che appaiono sulla riga relativa ad M_i della *lexical matrix*.

I synset possono essere utilizzati in diverse situazioni, ad esempio, come **disambiguatori** (*word sense disambiguation*).

Consideriamo il termine **board** che può significare “*tavola di legno*” ma anche “*insieme di persone riunite per uno scopo*”. Tale termine apparirà in almeno due synset:

{board, plank, . . . } per il primo significato.

{board, committee, . . . } per il secondo significato.

Se disponiamo dunque di un contesto, possiamo confrontare le parole del contesto con quelle dei due synset per disambiguare il significato

Nota:

In realtà il task di Word Sense Disambiguation è leggermente più complicato, lo studieremo nel dettaglio in seguito.

Osserviamo dunque che i synset non spiegano che cosa sia un certo concetto ma ne identificano soltanto i sinonimi:

WordNet è uno strumento ideato per nativi anglofoni e dunque dà per scontato che il significato delle parole (sia in input che nel synset) sia già conosciuto.

Giacché l'inglese è ricco di sinonimi, i synset sono spesso ben popolati e dunque sono sufficienti per adempiere a funzionalità di differenziazione come quella appena descritta.

In alcuni casi, però, non è possibile trovare un sinonimo appropriato per alcuni termini. In quei casi la polisemia è risolta da WordNet includendo anche una gloss (in inglese gloss/ *un commento*) insieme al synset. Per il nostro esempio della board potremmo allora avere qualcosa del genere:

{board, (a committee having supervisory powers; "the board has seven members"), ...}

Come vediamo, la glossa comprende una perifrasi (*giro di parole*) che descrive il concetto ed un esempio d'uso.

Gloss

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- **S: (n) house** (a dwelling that serves as living quarters for one or more families) "he has a house on Cape Cod"; "she felt she had to get out of the house"
 - [direct hyponym](#) / [full hyponym](#)
 - [part meronym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [derivationally related form](#)
- **S: (n) firm, house, business firm** (the members of a business organization that owns or operates one or more establishments) "he worked for a brokerage house"
- **S: (n) house** (the members of a religious community living together)
- **S: (n) house** (the audience gathered together in a theatre or cinema) "the house applauded"; "he counted the house"
- **S: (n) house** (an official assembly having legislative powers) "a bicameral legislature has two houses"
- **S: (n) house** (aristocratic family line) "the House of York"
- **S: (n) house** (play in which children take the roles of father or mother or children and pretend to interact like adults) "the children were playing house"

Es. synset per la parola house

Relazioni semantiche

WordNet è organizzato mediante **relazioni semantiche**.

Poiché una relazione semantica è una relazione fra significati e i significati sono rappresentati da synsets è naturale pensare ad una relazione semantica come ad un puntatore fra synsets.

Una caratteristica di queste relazioni semantiche è il fatto che siano **simmetriche**:

se c'è una relazione R fra i significati $\{x, x', \dots\}$ e i significati $\{y, y', \dots\}$ allora c'è anche una relazione R' fra i significati $\{y, y', \dots\}$ ed i significati $\{x, x', \dots\}$.

Nota:

Si noti che non stiamo parlando di una simmetria nel senso matematico del termine ma stiamo solo dicendo che ogni relazione ha sempre una controparte: R ed R' possono essere relazioni diverse (pensiamo ad esempio iperonimia/iponimia e meronimia/olonimia).

Sinonimia

È ormai evidente che la relazione più importante all'interno di WordNet sia la sinonimia, rappresentata mediante i synset.

Secondo una definizione formale, una parola x è sinonimo di una parola y se in una frase ove appare x è possibile sostituire y senza modificare il valore di verità della frase.

Con una definizione di questo tipo però avremmo pochissimi sinonimi (e quindi synset poco popolati), dato che spesso due parole sono sinonime soltanto all'interno di un certo contesto (*in quanto hanno più significati*).

Per questa ragione si adotta una definizione più debole:

Definizione

Una parola x è sinonimo di una parola y se, **all'interno di un contesto C** , in una frase ove appare x è possibile sostituire y senza modificare il valore di verità della frase.

Es. nel contesto della carpenteria sostituire plank al posto di board in una frase non ne altera il valore di verità.

Nota:

La relazione di sinonimia è simmetrica nel senso matematico del termine.

Se A è sinonimo di B allora B è sinonimo di A

Antonimia

La parola *ricco* è un antonimo della parola *povero*, in pratica, un opposto.

Sebbene possa essere banale, l'antonimia è parecchio difficile da definire.

Intuitivamente potremmo pensare che l'antonimo della parola *x* sia semplicemente *non x*, ma si tratta di una definizione fuorviante. Infatti, essere *non ricchi* è ben diverso da essere *poveri*.

Si ponga particolare attenzione al fatto che **l'antonimia è una relazione lessicale fra forme di parole e non fra significati di parole**. Per chiarificare questo fatto consideriamo i due synset:

$S:\{\text{rise, ascend, ...}\}$

$S:\{\text{fall, descend, ...}\}$

Evidentemente *rise* e *fall* sono antonimi così come lo sono *ascend* e *descend* mentre è difficile sostenere che *rise* e *descend* lo siano.

Quindi l'antonimia non è definita fra il significato di *rise/ascend* e *fall/descend* ma proprio fra alcune delle forme che lo rappresentano.

Capiamo dunque che la relazione di antonimia deve essere trattata separatamente, ovvero come una relazione fra forme di parole e non fra significati.

L'osservazione che abbiamo fatto vale anche per la sinonimia, infatti, i synset sono costituiti da termini e non da significati (che è unico per ogni synset).

Anche la sinonimia è fra termini e non fra synset (significati)

Nota:

Infine osserviamo che la relazione di antonimia è **simmetrica** nel senso matematico del termine.

Iponimia

A differenza della sinonimia e della antonimia l'iponimia/iperonimia è una **relazione semantica fra significati** e non fra termini.

Es. maple (acero) è un iponimo del significato di tree (ovvero non del termine tree in particolare ma del concetto che tale parola rappresenta) che è a sua volta un iponimo del significato di plant.

Definizione

Un concetto rappresentato dal synset {x, x', ...} è un iponimo del concetto rappresentato dal synset {y, y', ...} se un nativo anglosassone può sostenere che:

Un x è un tipo particolare di y, per un qualsiasi x ed un qualsiasi y presi dai corrispettivi synsets.

Es. Un acero è un tipo particolare di albero

Nota:

Se x è un iponimo di y, y è detto iperonimo di x.

L'iponimia è una relazione **transitiva** ed **asimmetrica**

Transitiva:

Se A è iponimo di B e B è iponimo di C allora A è anche iponimo di C

Asimmetrica

Se A è iponimo di B, B **non è** iponimo di A (ma iperonimo)

Poichè solitamente un certo concetto ha un solo iponimo superordinato (*acero come iponimo diretto ha solo albero*) è possibile generare una struttura semantica gerarchica nella quale un iponimo è detto essere sotto il suo supeordinato.

Un iponimo eredita tutte le caratteristiche dal suo concetto supeordinato (*più generico*) e vi aggiunge almeno un elemento distintivo.

Meronymia

La meronymia, come l'iponimia, è una relazione semantica fra significati e non fra termini.

Possiamo immaginare la meronymia come una relazione part-of

Definizione

Un concetto rappresentato dal synset $\{x, x', \dots\}$ è un meronimo del concetto rappresentato dal synset $\{y, y', \dots\}$ se un nativo anglosassone può sostenere che:

Un y ha un x (come parte) oppure Un x è una parte di y , per un qualsiasi x ed un qualsiasi y presi dai corrispettivi synsets.

Es. dito è meronimo di mano

Nota:

Se x è un meronimo di y , y è detto olonimo di x .

La meronymia è una relazione **transitiva** ed **asimmetrica**

Può anch'essa essere utilizzata per costruire una gerarchia, anche se il problema è un po' più complesso rispetto a quando trattiamo iponimi poiché un meronimo può avere più olonimi.

Un complessa rete
di relazioni

Tutte queste, e altre, relazioni possono essere rappresentate in WordNet da puntatori (archi etichettati) da un synset ad un altro

Queste relazioni rappresentano delle associazioni che formano una rete complessa.

Sapere dove è situata una parola in questa rete è una parte molto importante per comprenderne il suo significato