

I nomi in WordNet

La definizione di un nome presa da un dizionario classico esplicita solitamente il termine superordinato a quello preso in analisi ed alcune delle caratteristiche distintive del nome considerato.

Questa informazione di base è ottima per la costruzione di WordNet, anche se come stiamo per imparare vorremmo anche aggiungere altre informazioni (abbiamo già detto che un dizionario classico non basta per risolvere il task al quale WordNet adempie (**WSD**)).

Nota:

Ogni parola eredita tutte le feature di tutti i propri superordinati

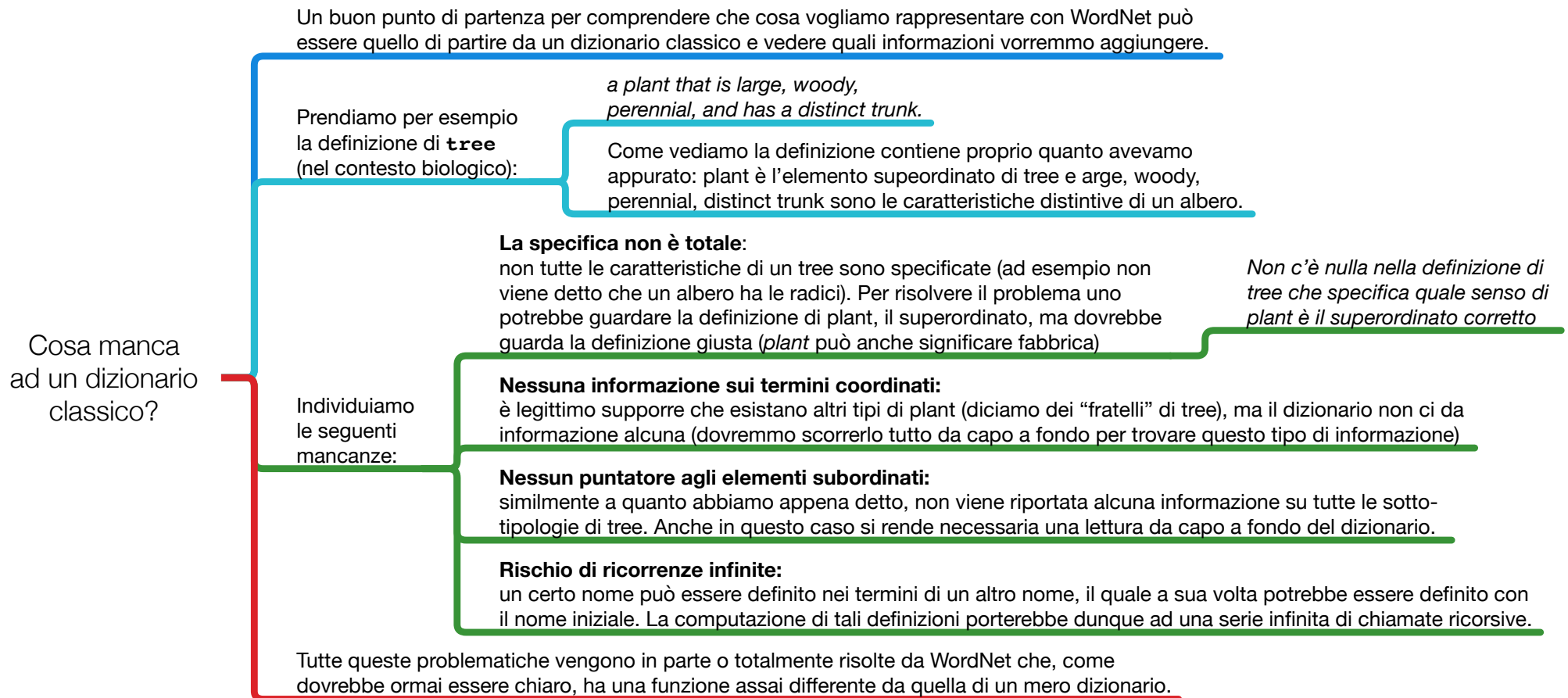
Possiamo in realtà formalizzare un poco di più quali siano le *caratteristiche distintive* di un determinato nome a cui siamo interessati. Ci concentriamo infatti su tre features distintive:

Attributes

Parts (e quindi meronimie).

Functions

Nelle sezioni a seguire andremo a vedere come definire per ogni concetto queste caratteristiche e andremo finalmente toccare con mano la struttura di WordNet.



Assunzioni psicolinguistiche

Poiché WordNet è un dizionario ispirato da teorie psicolinguistiche, i **nomi** sono stati organizzati seguendo un *sistema gerarchico* (modello lessicale adottato dalla mente umana). Le evidenze psicolinguistiche che ci indicano la presenza di tale struttura usano come prova il modo col quale noi siamo in grado di gestire le **anafore**.

Anafora:

Figura retorica che consiste nel ripetere una o più parole all'inizio di più frasi

Consideriamo ad esempio la frase *He owned a rifle (fucile), but the gun had not been fired*: risulta per noi immediato comprendere che gun si riferisca a rifle, ovvero, abbiamo trovato immediatamente l'iponimo supeordinato alla parola gun.

Un altro esempio è portato dalla frase *A rifle is safer than a gun*: in questo caso abbiamo tentato di confrontare in maniera semanticamente *insensata* un superordinato con il suo iponimo. È la nostra struttura mentale gerarchica a permetterci di riconoscere l'anomalia semantica che quel periodo esprime.

Altri esempi di anafora:

«È qui che piantai il **pioppo**. **L'albero** è però ormai morto.»

«Hanno preso **l'operaio**. **L'uomo** sembra abbia ucciso la moglie.»

Evidenze da esperimenti

In riferimento agli esperimenti e teorie di Collins e Quillian

Secondo alcuni autori è possibile **esaminare i tempi di risposta** a determinate domande per indicare la distanza in termini di livelli gerarchici fra due termini.

È stato ad esempio osservato che i tempi di risposta (il tempo impiegato dall'umano per dire che la frase è vera) sono minori per la frase *a canary can sing* rispetto a *a canary can fly*. Ancora più tempo è necessario per rispondere alla frase *a canary has skin*.

Grazie a questo esempio possiamo assumere che all'interno della mente dell'umano:

sing sia memorizzato come feature di canary (precisamente come function)

can fly come feature di bird

has skin come feature di animal

Chiaramente se tutte le informazioni relative a canary fossero storate con il nome stesso i tempi di reazione sarebbero gli stessi per tutte e tre le fasi. Ma poiché le evidenze mostrano che i tempi sono diversi, possiamo immaginare la presenza di una struttura gerarchica che recupera informazioni aggiuntive su un concetto soltanto quando necessario.

Evidenze da
esperimenti
problem

Anche se il fatto che i nomi comuni siano organizzati nella mente con una struttura semantica gerarchica è ormai dato per assodato, le modalità con le quali le informazioni generiche sono ereditate o memorizzate in maniera ridondante non sono affatto chiare.

Es. secondo una teoria di Quillian robin (*pettirosso*) e ostrich (*struzzo*) condividono lo stesso link semantico al loro iponimo bird; eppure la frase *A robin is a bird* è confermata in tempi minori rispetto alla frase *An ostrich is a bird*.

Questo tipo di risultato ha portato alcuni scienziati a sostenere che non sia vero che la mente rappresenti gerarchicamente i nomi comuni, mentre altri (e fra questi i progettisti di WordNet) hanno invece dedotto che l'ipotesi dell'ereditarietà sia corretta ma che i tempi di risposta non misurino una distanza semantica ma piuttosto una **distanza pragmatica e quindi non una vera e propria distanza fra significati ma soltanto fra modi di utilizzare le parole.**

Organizzazione dei nomi all'interno di WordNet

Al top della struttura di WordNet sono presenti 25 categorie molto generali, dette *primitive semantiche* (**supersensi**).

Ognuna di queste categorie corrisponde a campi semantici quanto più possibile differenti tra loro.

I 25 top level sono stati definiti al fine di avere per ogni nome della lingua inglese un iperonimo che lo rappresenti, perciò, ogni supersenso è iperonimo di tutti i nomi che si trovano nella gerarchia semantica di cui è la radice

ACT	corsa	MOTIVATION	ragione, motivo
ANIMAL	cane, gorilla	NATURAL OBJ	mare, salma
ARTIFACT	edificio, bomba	NATURAL PHENOM	vento, maltempo
ATTRIBUTE	ricchezza, eleganza	PERSON	ricercato, padre
BOBYPART	braccia, gambe	PLANT	acero, zucchini
COMMUNICATION	libro, discussione	POSSESSION	feudo, piantagione
EVENT	incidente	PROCESS	parto, erosione
FEELING	paura, felicità	QUANTITY	minuto, miliardo
FOOD	pianino, pasta	SHAPE	curva, cilindro
GROUP	pubblico, assemblea	STATE	crisi, coma
KNOWLEDGE	cultura, senno	SUBSTANCE	marmo, pesce
LOCATION	vetta, polo nord	TIME	oggi, compleanno

Le gerarchie che si dipartono da questi 25 supersensi sono tendenzialmente mutualmente esclusive (anche se a volte ci sono cross-references) e sono espanse in maniera eterogenea (certe categorie sono più dettagliate di altre)

I 25 synset che rappresentano i supersensi

{*living thing, organism*}

{*thing, entity*}

{*non-living thing, object*}

{*plant, flora*}

{*animal, fauna*}

{*person, human being*}

{*natural object*}

{*artifact*}

{*substance*}

{*food*}

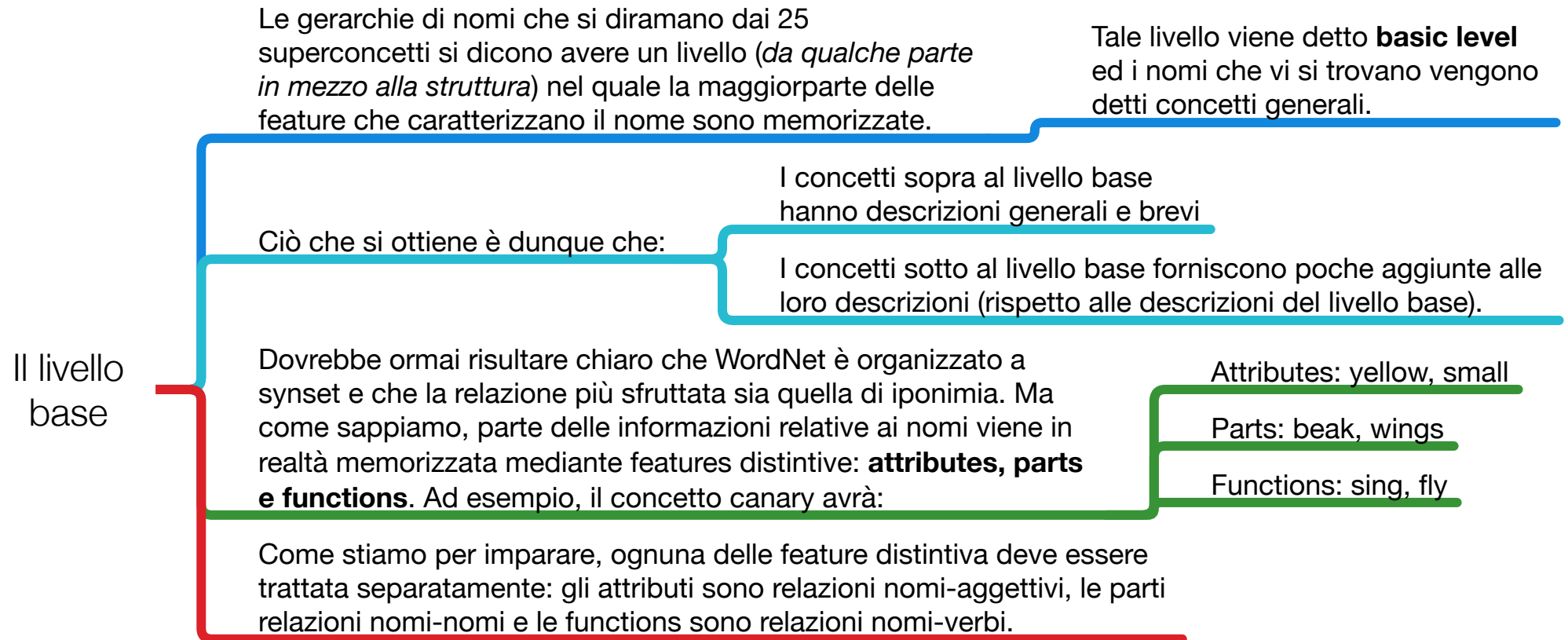
Es. 7 dei 25 superconcetti fungono da specificazione (iponimi) per gli esseri viventi e gli esseri non viventi

Componenti semantiche

Le Feature distintive

Ricapitoliamo quanto detto fin'ora: WordNet è una struttura suddivisa in sezioni, ognuna delle quali tratta un tipo di part of speech diverso.

Al momento ci stiamo concentrando sui nomi ed abbiamo appreso che i nomi vengono categorizzati, seguendo assunzioni psicolinguistiche, come gerarchie basate sulla relazione di iponimia. In questa sezione andremo a capire come le features distintive (attributes, parts e functions) vengano storate all'interno della gerarchia e sui singoli nomi.



Attributes (modifiers)

I valori degli attributes di un nome sono espressi mediante gli aggettivi.

Es. gli attributi `size` e `color` di `canary` sono espressi da aggettivi quali *small* e *yellow*.

In questo senso gli aggettivi sono detti modificatori di un nome.

Nota:
si noti che alcuni aggettivi non possono essere intesi come modificatori effettivi ma soltanto come modificatori metaforici:

Es. Un canary può essere definito **affamato**, in quanto avere fame è una feature degli animali e un canary è un animale. Un canary **generoso** invece può essere solamente interpretato metaforicamente in quanto la generosità non è una feature dei canary

Parts (meronimie)

Le parts di un certo nome sono rappresentate mediante relazioni di meronimia.

Sappiamo già cosa siano i meronimi: poniamo ora l'attenzione sul fatto che la relazione di meronimia venga espressa in WordNet mediante link fra synset poiché è una relazione semantica, e soprattutto sul fatto che le meronimie sono feature distintive che gli iponimi possono ereditare.

Es. se beak (becco) e wing (ala) sono feature distintive di bird (*sono parti di bird*), allora canary essendo iponimo di bird avrà come parti anche beak e wing.

Come sappiamo la meronimia è asimmetrica e transitiva anche se talvolta la transitività può risultare un po' forzata: è vero che una porta ha maniglia ed una casa ha una porta, ma è piuttosto strano sostenere che una casa abbia una maniglia, no?

Problema

Tale problematica nasce dal fatto che in realtà vi sono sei tipi diversi di meronimie:

Component-object (branch/tree)

Member-collection (tree/forest)

Portion-mass (slice/cake)

Stuff-object (aluminum/airplane)

Feature-activity (paying/shopping)

Place-area (Princeton/New Jersey)

Functions (predicati)

La feature distintiva che chiamiamo function descrive ciò che un certo nome normalmente fa o farebbe.

Talvolta questa definizione può sembrare un po' strana, infatti, è naturale dire che la funzionalità di un coltello è quella di tagliare, mentre è meno naturale sostenere che le funzionalità di un canarino siano quelle di cantare e volare.

La relazione di function viene quindi istituita fra nomi e verbi e può essere di uno di questi tipi:

Instruments (knife/cut)

Materials (wool/knit) (*lana/lavorare la maglia*)

Products (picture/paint) (*disegno/dipingere*)

Containers (box/hold) (*scatola/contenere*).

Si noti che spesso (ma non sempre) il ruolo semantico (*il significato*) di un nome è indipendente dal ruolo sintattico (es. *soggetto, complemento oggetto, etc.*) che ricopre.

Es. le frasi *John hit the mugger with a hammer* e *the hammer hit him on the head*: nel primo caso hammer era complemento oggetto mentre nel secondo era soggetto, eppure, dal punto di vista semantico hammer è sempre un instrument.

Vi sono anche ragioni linguistiche per le quali pare essere corretto assumere che ciò che fa una certa cosa è funzione del suo significato.

Es. la frase *good pencil*: stiamo intendendo che una certa matita è una matita che scrive bene. A seconda del termine preposto a pencil staremo perciò dicendo che una matita scrive bene/male/ecc.

Un *good knife* è un coltello che taglia bene, etc.. Mettere, ad es, la parola *good* davanti ad un nome ne va dunque a modificare il significato

Se dovessimo considerare l'inserimento di tutti questi significati all'interno di WordNet ci troveremmo in enorme difficoltà, ma ci viene in aiuto un risultato sorprendente: un oggetto che non è un x, se viene definito come un "buon x", svolge bene le funzionalità che x normalmente svolge.

Es. Una persona che si siede su una scatola può dire: *questa scatola è una buona sedia* indicando che la scatola svolge bene la funzione che ci si aspetta che svolgano le sedie