

Dictionaries as Mirror of the Mind

Stevan Harnad¹

*joint work with A. Blondin Massé, G. Chicoisne,
Y. Gargouri, O. Marcotte et O. Picard*

¹ISC, Institut des sciences cognitives

February 28th, 2008

The Defining Vocabulary

Definition

The **defining vocabulary** of a dictionary D is the **unique maximal** subdictionary M of D such that every defined word of M appears in **at least one** definition of D , i.e.

$$\text{DEFINIENDA}(M) = \text{DEFINIENTES}(M).$$

It is **computed** as follows :

- 1 If **$\text{DEFINIENDA}(D) = \text{DEFINIENTES}(D)$** then stop.
Otherwise, go to the next step.
- 2 **Remove** from D every word that does **not appear** in any definition of D .
- 3 **Repeat** steps 1 and 2.

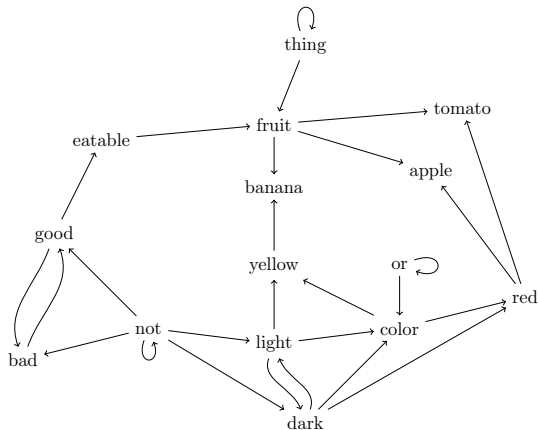
Example

Let D be this (very) simple **dictionary** :

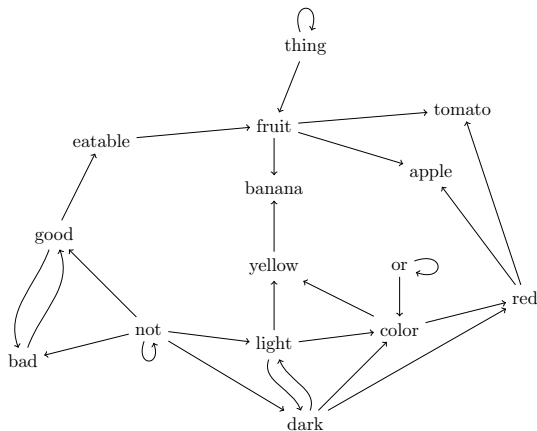
apple	:	red fruit,	bad	:	not good,
banana	:	yellow fruit,	color	:	dark or light,
dark	:	not light,	eatable	:	good,
fruit	:	eatable thing,	good	:	not bad,
light	:	not dark,	not	:	not,
or	:	or,	red	:	dark color,
thing	:	thing,	tomato	:	red fruit,
yellow	:	light color			

D can be represented as a **graph**.

Computing the Defining Vocabulary (Example)

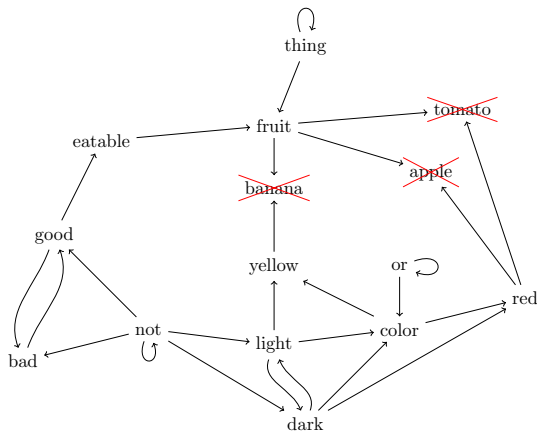


Computing the Defining Vocabulary (Example)



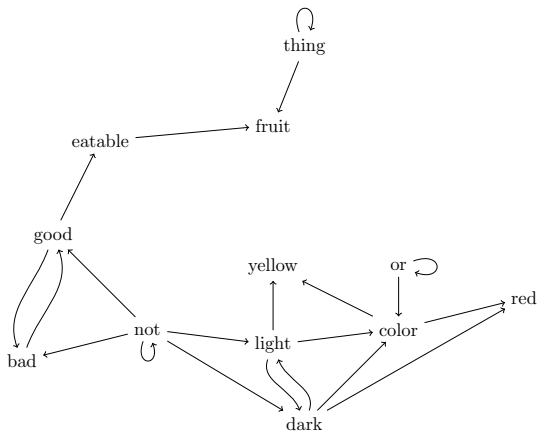
Step 1 : the words **apple**, **banana** and **tomato** are removed

Computing the Defining Vocabulary (Example)

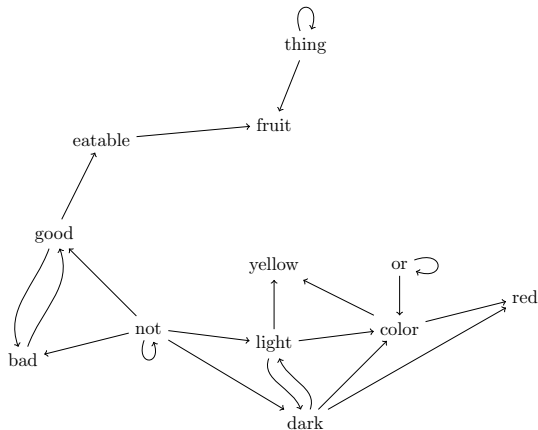


Step 1 : the words **apple**, **banana** and **tomato** are removed

Computing the Defining Vocabulary (Example)

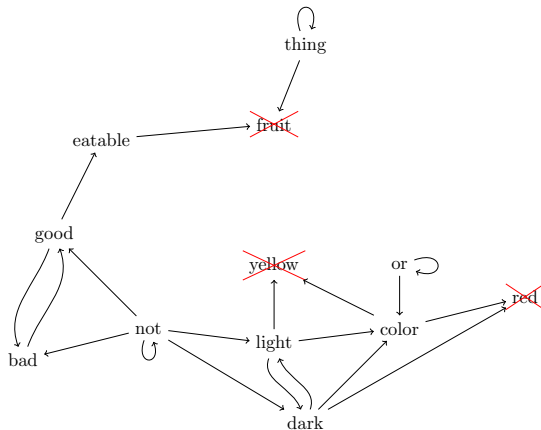


Computing the Defining Vocabulary (Example)



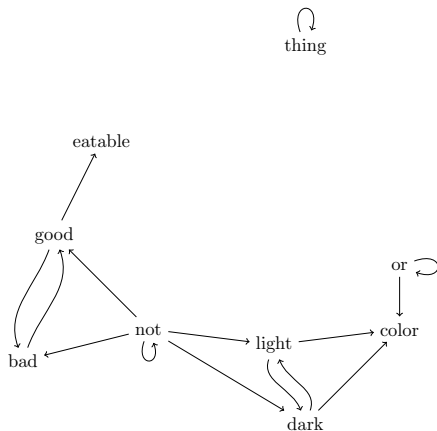
Step 2 : the words **fruit**, **red** and **yellow** are removed

Computing the Defining Vocabulary (Example)

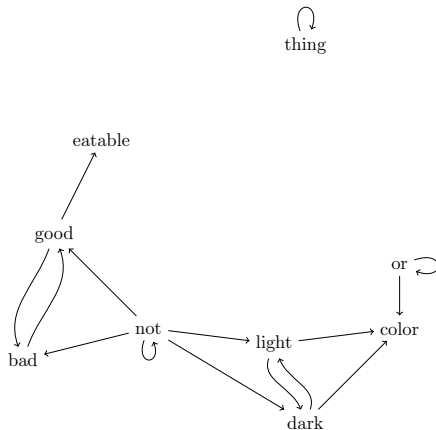


Step 2 : the words **fruit**, **red** and **yellow** are removed

Computing the Defining Vocabulary (Example)

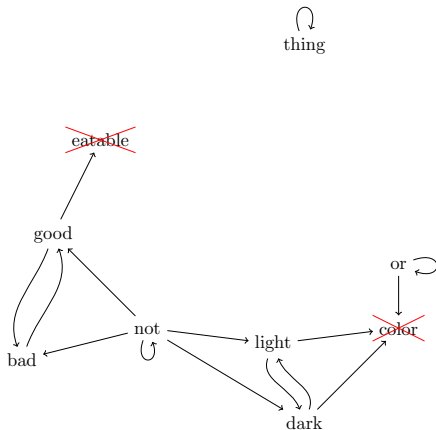


Computing the Defining Vocabulary (Example)



Step 3 : the words **color** and **eatable** are removed

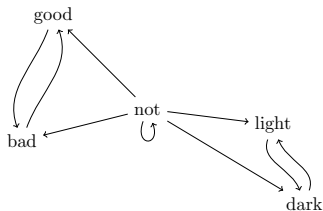
Computing the Defining Vocabulary (Example)



Step 3 : the words **color** and **eatable** are removed

Computing the Defining Vocabulary (Example)

thing

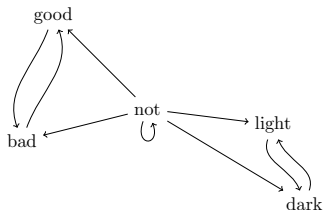


or



Computing the Defining Vocabulary (Example)

thing

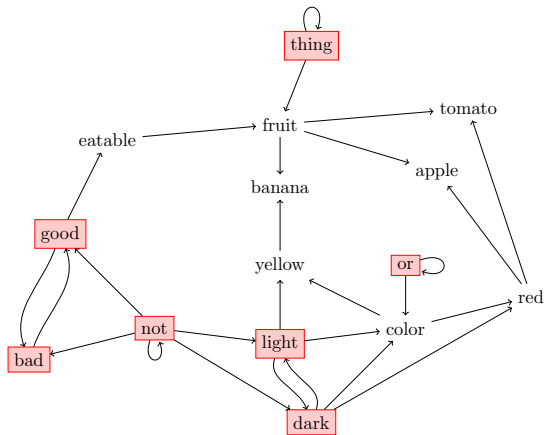


or



Stop : this is the **defining vocabulary**

Computing the Defining Vocabulary (Example)



Stop : this is the **defining vocabulary**

Definition

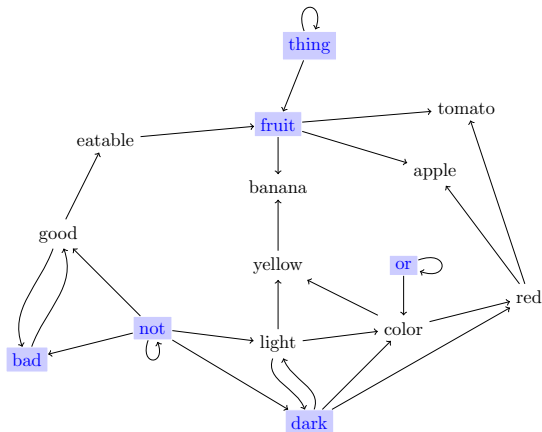
Let D be a dictionary and V be some vocabulary of D (i.e. any finite set of words defined in D). The **learnable vocabulary** from V is the set $L(V)$ defined as follows

- 1 If $w \in V$, then $w \in L(V)$.
- 2 If $w \in D$ and $d_w \subseteq L(V)$, then $w \in L(V)$.

Remark

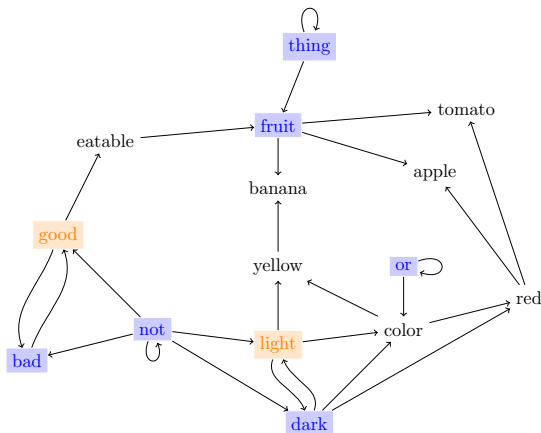
- **Rule 1** means that every word in V can be (trivially) learned from V .
- **Rule 2** means that it suffices to have learned **all words in the definition** of a word to be able to learn it.

Computing the learnable vocabulary (Example)



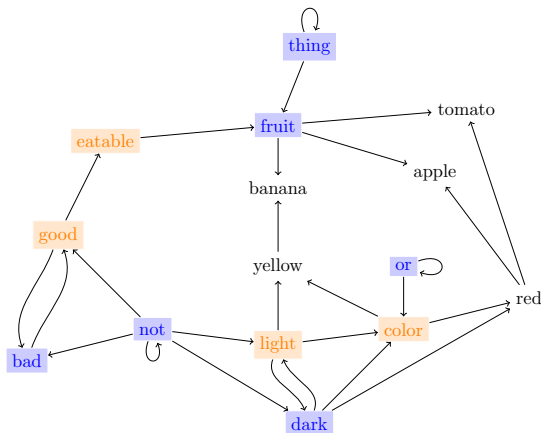
Initially : **six** words are **grounded**

Computing the learnable vocabulary (Example)



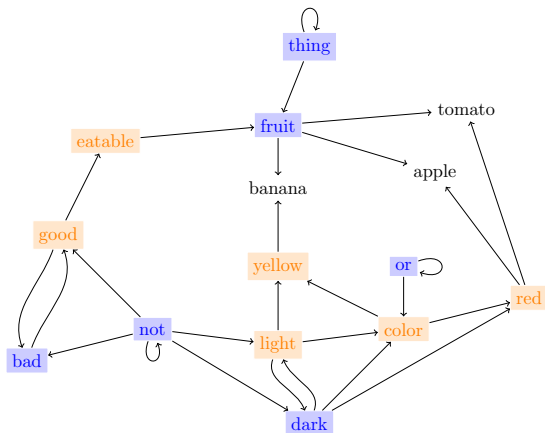
Step 1 : the words **good** and **light** are learned

Computing the learnable vocabulary (Example)



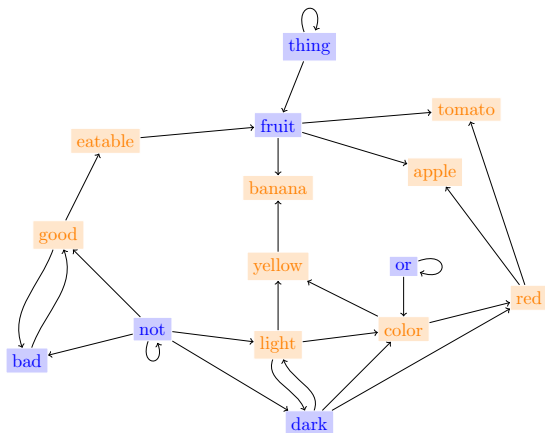
Step 2 : the words **eatable** and **color** are learned

Computing the learnable vocabulary (Example)



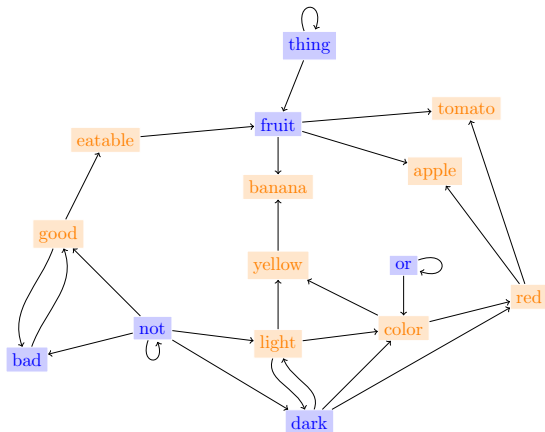
Step 3 : the words **red** and **yellow** are learned

Computing the learnable vocabulary (Example)



Step 4 : the words **apple**, **banana** and **tomato** are learned

Computing the learnable vocabulary (Example)



The initial six words **suffice** to understand all other words

Here are some **statistics** about the **LDOCE** :

Entries (with polysemy)	80,294
Entries (without polysemy)	24,023
Defining vocabulary	7,136
Maximal definitional distance	6
Average definitional distance	5

Note that the **definitional distance** between words w_1 and w_2 is the length of the smallest path from w_1 to w_2 in the **non-directed** graph of the **definitional relationship**.

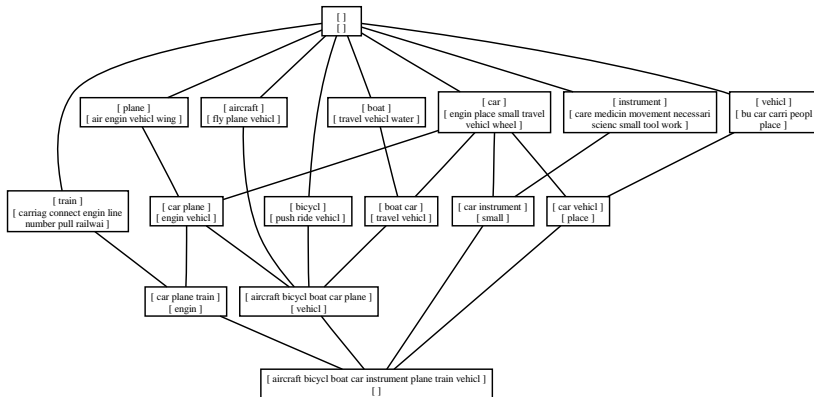
Extracting Meaning Using FCA

Consider the following **formal context** extracted from **Longman's** dictionary, after applying **Porter's** algorithm and removing **functional words** :

vehicle	:	carri car bu place peopl
bicycle	:	push vehicl ride
car	:	wheel vehicl engin travel place small
plane	:	vehicl engin wing air
instrument	:	medicin tool work necessari small scienc care movement
aircraft	:	fly vehicl plane
train	:	pull engin railwai number carriag connect line
boat	:	water vehicl travel

Objects are definienda, **attributes** are definientes

Extracting Meaning Using FCA (continued)



- 1 What **attributes** would be the best ones to characterize **semantic relationships** between words ? Is the **grammatical category** important ? What about extracting attributes from the **dictionary** itself ?
- 2 The words **train** and **car** are both **vehicles**. Does there exist a measure on **lattices** to distinguish the relation **train-vehicle** from the relation **train-car** ?