KDI
An Example of Linguistic Resource: WordNet

Fausto Giunchiglia and Mattia Fumagallli

University of Trento

- A lexical database:
  - psycholinguistic grounding
  - just for supporting humans in browsing vocabularies

Version1.6

	Total	Noun	Verb	Adj	Adv
Word	**129.625**	94.503	12.156	20.199	4.575
Synset	**99.758**	66.054	10.348	17.944	3.604

	Total	Noun	Verb	Adj	Adv
Word	**155,327**	117,097	11,488	22,141	4,601
Synset	**117,597**	81,426	13,650	18,877	3,644
Monosemic	128,321	101,321	6,261	16,889	3,850

▶ **Synset = synonym set**
  ▶ set of synonyms as lexicalized concepts, e.g.,
  {vehicle, car, automobile}

▶ **relations**
  ▶ *lexical*: between words composing synsets
    ▶ synonym, antonym, …
  ▶ *semantics*: between synsets
    ▶ hypernym, meronym, implication, …

SYNSET

Vehicle
Car
Automobile

Railway
Car
Railway Car

SYNSET

Vehicle with 4 wheels

Railway car

▶ A term can be replaced in at least one context

WordNet Synonym: Two words W1 and W2 are synonyms if replacing W1 with W2 in at least one (linguistic) context, the meaning of the given sentence does not change

Synonym text:
If X is Noun1, then X it is Noun2, and vice
    versa
It is a *fiddle*, therefore it is a *violin*
   It is a *violin*, therefore it is a *fiddle*

Relations

Category	Relation	*Type*	*Example*
Noun	Hypernym/hypo	Sem	dog IS A KIND OF animal
	Meronym	Sem	arm IS A PART OF body
Verb	Implication: Cause Precondition Troponym Inclusion	Sem	to kill CAUSES to die to succeed ENTAILS DOING to try to limp IS ONE WAY TO walk snore ENTAILS DOING to sleep
	Opposition	Lex	to die ANTONYM to be born
Adj	Antonym	Lex	hot ANTONYM cold
Adv	Derived adj	Lex	quickly DERIVED FROM quick
	Antonym	Lex	quickly ANTONYM slowly
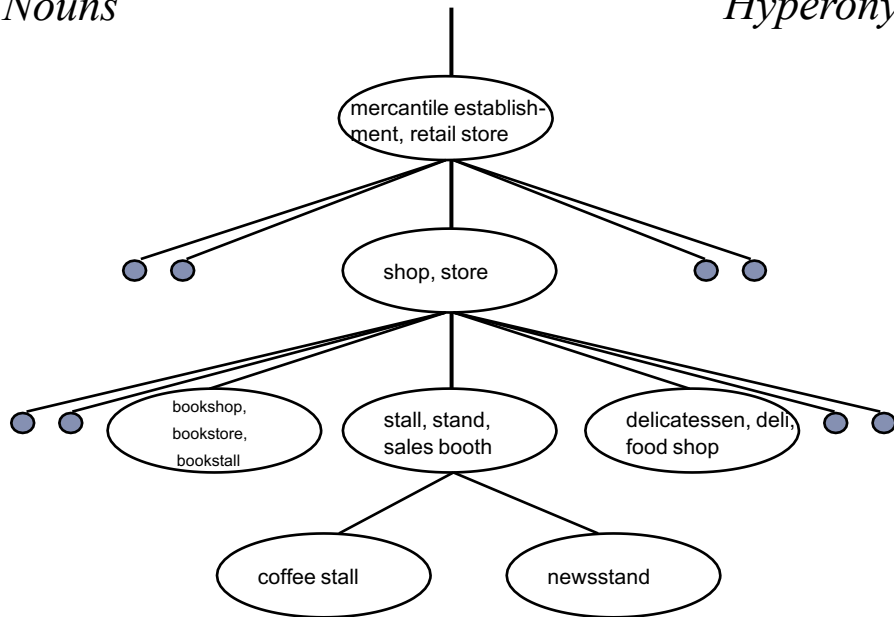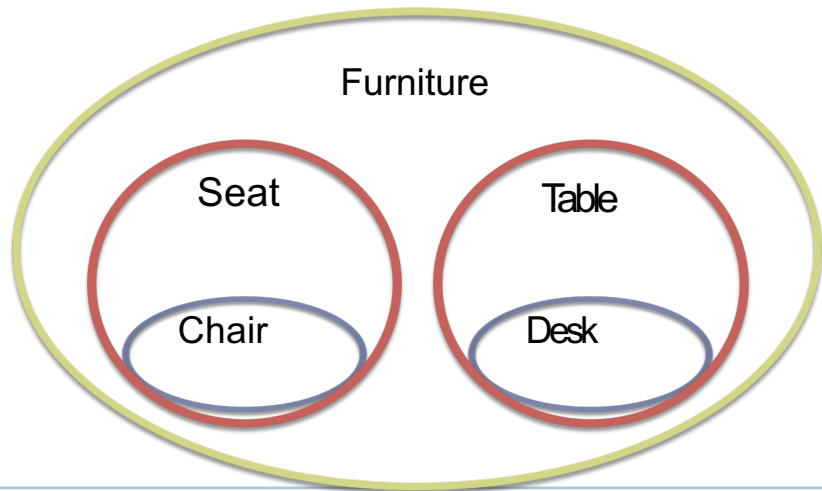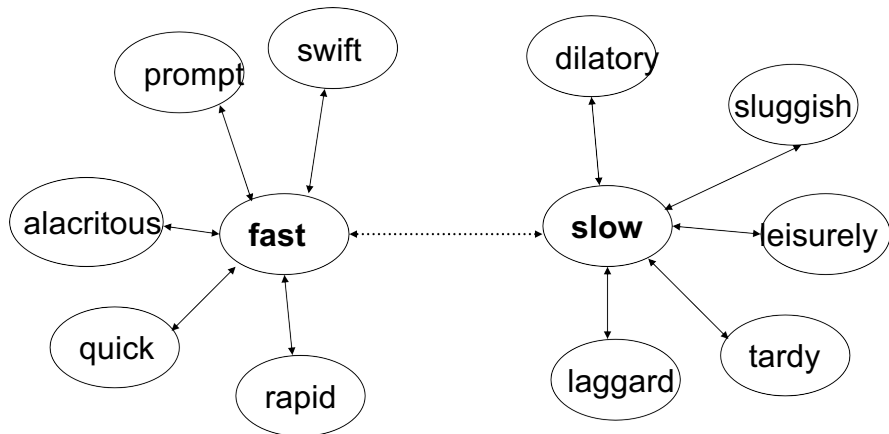
» single words {*palace*, *castle*}
» compound words {*blueberry*}
» collocations {*one way*}
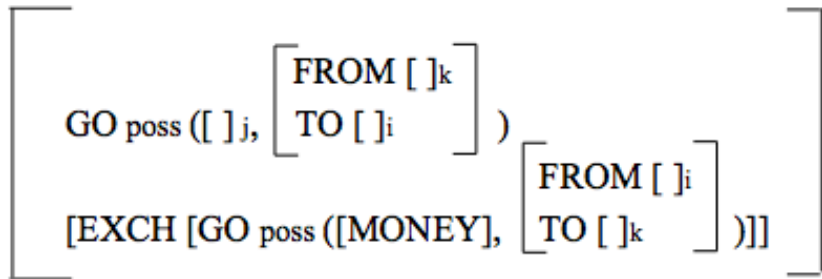» idiomatic expressions {*kick the bucket*, *buy the farm*, *snuff it*}

» artificial nodes: the do not represent lexical concepts
  {*create by mental act*, *create mentally*}

▶ Synset:
  ▶ it does not provide a full specification of the word meaning
  ▶ it points to a lexical concept and represent its (partial) meaning by means of its lexical and semantics relations with other lexical concepts
▶ The core approach:
  ▶ allowing the distinction between two lexicalized concepts is enough

▶Meaning composition
▶Meaning postulate
▶Prototypes
▶Semantic networks
▶…

▶ word meaning = set of atomic concepts
  ▶ E.g.: to **buy** (Jackendoff 1983)

$$\left[\begin{array}{l} \text{GO poss} \left([\ ]_j, \begin{bmatrix} \text{FROM } [\ ]_k \\ \text{TO } [\ ]_i \end{bmatrix} \right) \\ [\text{EXCH } [\text{GO poss} ([\text{MONEY}], \begin{bmatrix} \text{FROM } [\ ]_i \\ \text{TO } [\ ]_k \end{bmatrix})]] \end{array} \right]$$

▶ Meaning postulates: representation of word meaning by representing meaning relations between words

▶ E.g.: to buy

$buy(x,y,z) \longrightarrow get(x,y,z)$

$buy(x,y,z) \longrightarrow pay(x,y,z)$

$buy(x,y,z) \longrightarrow choose(x,y)$

$buy(x,y,z) \longrightarrow sell(z,y,x)$

▶ E.g.: bachelor

$bachelor(x) \longrightarrow man(x) \wedge \neg married(x)$

⟩ Word meaning = information that is true about the most typical exemplars related to that concept

⟩ e.g. tiger

property	Possible values	Default values
being	feline	
Weight	< 180 Kg.	120 Kg
Height	< 106 cm.	80 cm.
Length	< 250 cm.	180 cm.
Color	yellow with black stripes White Black Yellow	Yellow with black stripes
Habitat	jungle rivers	jungle

➤ Meaning of a word = relations with other words
  ➤ e.g.: to buy
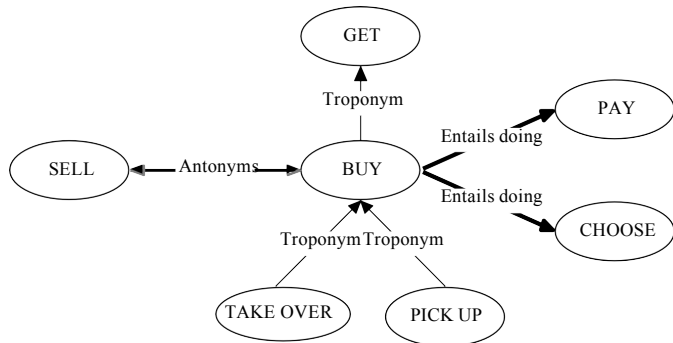
A closer look on the word "get"

….
» 17. {catch, get}
» 18. {catch, arrest, get}
» 19. {get, catch}
» 20. {get}
» 21. {get}
» 22. {get}
» 23. {catch, get}
» 24. {catch, get}
» …

"get" senses

….

» 17. {catch, get} →→{understand}
» 18. {catch, arrest, get} →→{attract, pull, pull in, draw, draw in}
» 19. {get, catch} →→{hit}
» 20. {get} →→{}
» 21. {get} →→{get, acquire}
» 22. {get} →→{buy, purchase}
» 23. {catch, get} →→{hear}
» 24. {catch, get} →→{hurt, ache, suffer}
» …

"get" glosses

» 17. {catch, get} -- (grasp with the mind or develop an understanding of) "did you catch that allusion?"; "We caught something of his theory in the lecture"; "don't catch your meaning"; "did you get it?"; "She didn't get the joke"; "I just don't get him"

» 18. {catch, arrest, get} -- (attract and fix) "His look caught her"; "She caught his eye"; "Catch the attention of the waiter"

» 19. {get, catch} -- (reach with a blow or hit in a particular spot) "the rock caught her in the back of the head"; "The blow got him in the back"; "The punch caught him in the stomach"

» 20. {get} -- (reach by calculation) "What do you get when you add up these numbers?"

» 21. {get} -- (acquire as a result of some effort or action) "You cannot get water out of a stone"; "Where did she get these news?"

» 22. {get} -- (purchase) "What did you get at the toy store?"

» 23. {catch, get} -- (perceive by hearing) "I didn't catch your name"; "She didn't get his name when they met the first time"

» 24. {catch, get} -- (suffer from the receipt of) "She will catch hell for this behavior!"

snake, serpent, ophidian – *(limbless scaly elongate reptile; some are venomous)*

snake, snake in the grass – *(a deceitful or treacherous person)*

Snake, Snake River – *(a tributary of the Columbia River)*

Hydra, Snake – *(a long faint constellation near the equator stretching between Virgo and Cancer)*

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: [   ] Search WordNet

Display Options: [(Select option to change) ▼] Change

http://wordnetweb.princeton.edu/perl/webwn

WordNet: Let's try it

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: [machine]   [Search WordNet]

Display Options: [(Select option to change) ▾]   [Change]
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **machine** (any mechanical or electrical device that transmits or modifies energy to perform or assist in the performance of human tasks)
- S: (n) **machine** (an efficient person) *"the boxer was a magnificent fighting machine"*
- S: (n) **machine** (an intricate organization that accomplishes its goals efficiently) *"the war machine"*
- S: (n) **machine**, simple machine (a device for overcoming resistance at one point by applying force at some other point)
- S: (n) **machine**, political machine (a group that controls the activities of a political party) *"he was endorsed by the Democratic machine"*
- S: (n) car, auto, automobile, **machine**, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*

Verb

- S: (v) **machine** (turn, shape, mold, or otherwise finish by machinery)
- S: (v) **machine** (make by machinery) *"The Americans were machining while others still hand-made cars"*

▸ *EuroWordNet*

  ▸ Create synsets, create relations for every language

  ▸ Then map sysnets
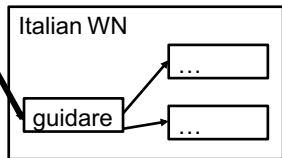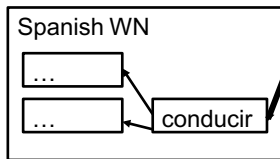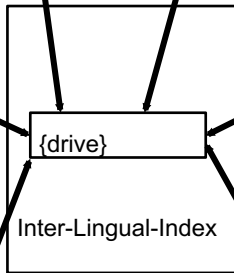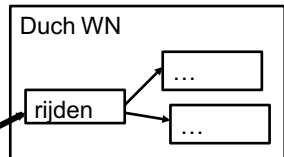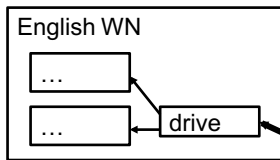
▸ *MultiWordNet*

  ▸ Create synsets for a new WordNet mapped to the English wordnet synsets (Princeton WordNet, PWN)

  ▸ Importing the semantic relations the new wordnet

- Dutch, Italian, Spanish, English (30,000 ss)
- German, French, Estonia, Czech (10,000 ss)
- Relation set extended with
  relations between languages
  (*near_synonym*, *xpos_...*)
- Language Index (*ILI*) for relations between
  languages (*eq_...*)
- Ontology of core shared concepts
- Hierarchy of labels for each domain

- Una An unstructured list of ILI indexes
- Where every ILI index is composed by:
  - a synset
  - an English gloss

- ILI codes are linked to:
  - Specific synsets meaning for the given language
  - One or more higher general terms
  - Possible domains

- High level concepts and domains can be linked with equivalence relations between ILI indexes and meanings of a specific language

EWN Structure

Ontologia di dominio
Road Traffic
...

Ontologia di alto livello
location
...

English WN
drive
...
...

Duch WN
rijden
...
...

Inter-Lingual-Index
{drive}

Spanish WN
conducir
...
...

Italian WN
guidare
...
...
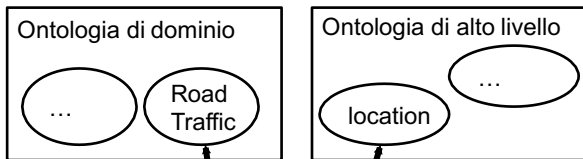
- The starting list is grounded on WordNet 1.5
- The list can be extended into two ways:
  - Adding concepts that are present in WordNet with other languages (not present in WN 1.5)
  - Adding Global Senses fro grouping more specific meanings

WordNet
- {dog} HAS_PART {tail}
- {wood} HAS_MEMBER {tree}
- {ice} HAS_SUBSTANCE {water}

In EuroWordNet some relations have been changed

EuroWordNet
- {hand} HAS_MERO_PART {finger}
- {fleet} HAS_MERO_MEMBER {ship}
- {book} HAS_MERO_MADEOF {paper}
- {bread} HAS_MERO_PORTION {slice}
- {desert} HAS_MERO_LOCATION {oasis}

MULTI WordNet

Human Language Technology Group

language [Italian ▼] word [scarpa] [Lookup]

MultiWordNet is a multilingual lexical database in which the Italian WordNet is strictly aligned with Princeton WordNet 1.6.
The Italian synsets are created in correspondence with the Princeton WordNet synsets, whenever possible, and semantic relations are imported from the corresponding English synsets; i.e., we assume that if there are two synsets in PWN and a relation holding between them, the same relation holds between the corresponding synsets in Italian. While the project stresses the usefulness of a strict alignment between wordnets of different languages, the multilingual hierarchy implemented is able to represent true lexical idiosyncrasies between languages, such as lexical gaps and denotation differences.

The information contained in the database can be browsed through the MultiWordNet browser, which facilitates the comparison of the lexica of the aligned languages. The MultiWordNet browser also allows for the access to the **Spanish**, **Portuguese**, **Hebrew**, **Romanian** and **Latin WordNets**, made available by courtesy of the **TALP Group** at the Universitat Politecnica de Catalunya (Spain), the **NLX-Group** at the University of Lisbon (Portugal), the **Computational Linguistic Group** at the University of Haifa (Israel), the "Alexandru Ioan Cuza" **University of Iasi** (Romania), and **University of Verona** (Italy) respectively. Although the Spanish, Portuguese, Hebrew, Romanian and Latin WordNets are compatible with the MultiWordNet model, these wordnets are not part of the MultiWordNet distribution.

News!
28.05.2008
Portuguese WordNet is online! Try 'lugar'
22.10.2007
A new version of **MultiWordNet (1.4.2)** is available!
22.10.2007
Latin WordNet is online! Try 'domus'

top

32

| Italian ▼ | Word ▼ | macchina | Search |

Noun

Overview ▼

The word "**macchina**" has 6 senses:

Italian WordNet created by HLT group - FBK-irst (Italy)

Noun
▶ 1. auto, automobile, autovettura, vettura, *macchina*   (Tourism) veicolo a quattro ruote, con motore generalmente a scoppio, per il trasporto su strada di un numero limitato di persone; "alle dieci riusciamo finalmente a partire con una comoda macchina giapponese ed un timido autista zoppo"; "é partito in macchina mezz'ora fa"
▶ 2. *macchina*   (Buildings) qualsiasi meccanismo meccanico od elettrico che esegue od assiste lo svolgimento di compiti umani
▶ 3. locomotiva, *macchina*, motrice   (Railway) veicolo ferroviario usato per il traino di convogli e vagoni ferroviari; "il treno ha fatto una sosta per il cambio della macchina"
▶ 4. *macchina*   (Person) si usa come oggetto di identificazione o termine di confronto per una persona che lavora molto, senza stancarsi mai; "é da stamattina che non fa una pausa, quel ragazzo é una macchina"
▶ 5. *macchina*   (Factotum) complesso di organi, poteri, funzioni che agiscono in stretta interdipendenza per il raggiungimento di uno scopo comune o lo svolgimento di un'unica attività; "noi dobbiamo essere certi che una volta attuato il federalismo la macchina dello Stato sarà più efficiente"; "soddisfazione negli ambienti di centro-destra che sin dal dopo elezioni avevano messo in moto la macchina della legge con alcuni ricorsi"
▶ 6. *macchina*   (Sport) nel ciclismo, automobilismo e motociclismo, la lunghezza del mezzo, adottata nelle gare come unità di misura per indicare il distacco tra un concorrente e l'altro; "ha vinto per mezza macchina"

Elaboration time: 0 sec

Visitors since 09-04-2008 0 2 7 4 2 9 5 3

As for EuroWordNet, MultiWordNet was created for addressing the most used languages:

Spanish, Portuguese, Italian, English, Rumanian, Latin, Jewish.

The main difference is the strategy followed for creating the interlingua index

In MultiWordNet the different languages graphs are built upon the English Wordnet graph.

Pros:

- Less manual work

- High compatibility between different languages graphs

- *Automatic procedures for building new resources*

Cons:

- *Highly dependent from English WordNet structure*

Italian WordNet (version 1.4)

	Nouns	Verbs	Adjectives	Adverbs	*Total*
Word senses	46,086	8,894	5,430	1,955	*62,365*
Lemmas	33,418	4,814	4,686	1,521	*44,439*
Synsets	26,747	4,532	3,101	1,097	*35,477*

- Assignment procedure
  - Efficient construction of synsets starting from the English reference
  - Given an Italian sense for a word it provides a weighted list of similar English synsets

- Lexical Gaps
- Individuation of lexical gaps

- Collins Dictionary
- Princeton WordNet (PWN)
- WordNet Domains
- Italian dictionary (DISC)

wood [wUd] **1.** n **a**. *(material)* legno; *(timber)*
b. *(forest)* bosco **c.** *(Golf)* mazza di legno; *(Bowls)*
2. adj **a.** *(made of wood)* di legno **b**. *(living etc. in a wood)*
di bosco, silvestre.

▶ *Translation groups* (TGR):
　▶ Different senses translated in both languages
　▶ English part: 40.959 words, <u>60</u>.901 TGRs
　▶ Italian part: 32.602 words, 46.545 TGRs

It helps Lexicographer to focus on PWN synsets that are more similar to the one they need to create

» The procedure finds a restricted set of synsets

» The lexicographer selects the right synset and discard the others

1) Find *synsets for every sense*
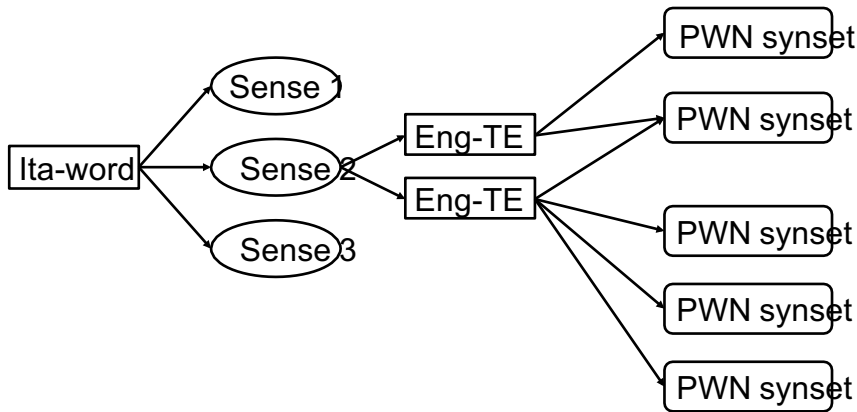
2) List synsets according to the following major criteria:
- ▶▶ Generic probability
- ▶▶ Translation
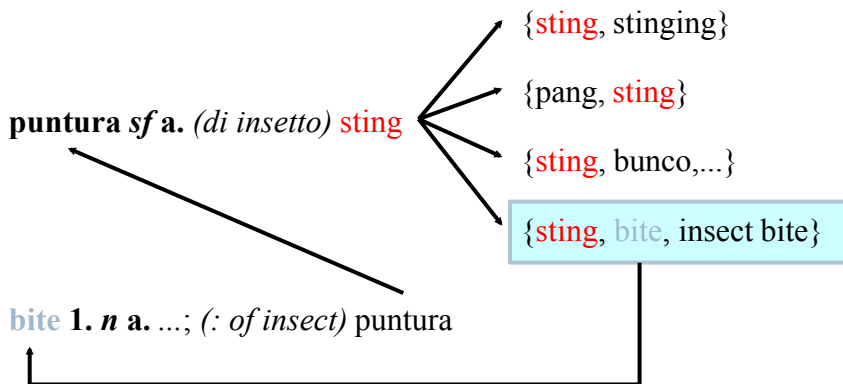- ▶▶ Glosses similarity
- ▶▶ Intersection between synsets

3) Select "best" synsets

dagherrotipo *sm* daguerreotype $\rightarrow$ {daguerreotype} (Atsererias et al. 97)

cane 1. *sm* *(Zool)* dog; …

{dog, domestic dog,…}
{frump, dog}
{dog}
{cad, blackguard, dog,...}
{pawl, detent, click, dog}
{andiron, firedog, dog,...}

▸ Semantic field:
   sclerosis *n* *(Med)* sclerosi

▸ Synonym, hyperonym:
   reason 1. *n* **a.** *(motive, cause)* ragione,…
   sole *n* *(fish)* sogliola

▸ Context:

handle 1. *n* …	(*of knife*)	manico, impugnatura;
(*of door, drawer*)		maniglia

corrente ... **3. *sf* (** *Elettr* **)** current;

| {current, electric current} | -- ELECTRICITY |

{current, stream}   -- GEOGRAPHY

{stream, flow, current}   -- GENERIC

albero 1. *sm* **a.** *(pianta)* tree

{ tree } -- a tall perennial woody plant having a main trunk ...
{ tree, tree diagram } -- a figure that branches from...

sogliola *sf (pesce)* sole

{ sole } -- right-eyed flatfish; many are valued as food; ...
    => { flatfish } -- any of several families of fishes having ...
{ sole } -- the underside of the foot
  => { area, region } – a part of an animal that has a special...

piega *sf* ... ; *(della pelle   )* fold;

{ fold, crease, ... } -- an angular shape made by folding

{ congregation, fold, faithful } -- a group of people who...

{ fold, plica } -- a folded part (as a fold of skin   or muscle)

{ fold, sheepfold, sheep pen, sheepcote }-- a pen for sheep

{ fold, folding } -- the act of folding; ...

pilastro *sm ...;(fig: sostegno)* pillar, mainstay

{pillar}

{column, tower, pillar}   {anchor, mainstay,...}

{pillar, mainstay}   {pillar, mainstay}

{column, pillar}   {mainstay}

{column, pillar}

- Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*, Mit Press, 1998
- Piek Vossen (ed.) *EuroWordNet: A multilingual database with lexical semantic networks*, Kluwer Academic, 1998
- L. Bentivogli and E. Pianta, "Looking for lexical gaps"*, Proc. of Euralex-2000*, Stuttgart, Germany, 2000.
- E. Pianta, L. Bentivogli and C. Girardi, "MultiWordNet: Developing an aligned multilingual database", *Proc. of 1st International WordNet Conference,* Mysore, India, 2002
- MultiWordNet homepage: http://multiwordnet.itc.it

Acknowledgments