

# ***Towards interfacing lexical and ontological resources***

Francesca Frontini & Monica Monachini  
ILC, CNR, Pisa

LOCI IV, Ontologies and Lexical Semantics  
Roma, 01/10/2011

# Computational lexica

During the last two decades, the Computational Linguistics community has dedicated considerable effort to the research and development **Lexical Resources** (LRs), especially **Computational Lexicons**.

Examples of large and informationally rich LRs include

**WordNet**

**EuroWordNet**

**FRAMENET**

**SIMPLE**

# Ontological backbone of Lexica

These LRs, even though belonging to different linguistic approaches and theories, share a common element; all of them contain, explicitly or implicitly, an ontology as the means of organizing their structure.

**They are ontologically based lexicons.**

Word senses = ontological categories

Lexical relations = ontological relations (hyponymy = subsumption)

# Formal ontologies

Formal ontologies with respect to ontologically based lexica offer:

- **OWL**: ontology language
- W3C **standard** for Semantic Web apps.
- **Reasoning** capabilities

# Potential of ontologies for lexica

LRs can benefit from incorporating formal ontologies

- **Standardization:** conversion of NLP resources into OWL
- Increased **reasoning and inferencing potential:** offering a great potential for applications related to the Semantic Web and to Natural Language Understanding/Processing
- **Language independent** concepts (for multilingual resources)

# Potential of Lexicons for Ontologies

Ontologies in their turn can benefit from an easy access to typical linguistic information.

- **Richness** of information
- Access to common-sense and **linguistic knowledge**

# Hot issues

- From language to knowledge: Giving Lexicons a **Formal** Ontological Backbone
- From knowledge to language: **Mapping** Lexicons onto Formal Ontologies (**SUMO** and **DOLCE** are among the most widely used)
- **Reasoning** for Information Extraction and Text mining

# From language to knowledge

Lexical knowledge contained in LRs is exploited to **derive formal knowledge**.

Semantic Web applications.

Ex: **WordNet RDF / OWL**



# SIMPLE OWL (1)

Conversion of the **SIMPLE ontology** into the formal language OWL -  
**Enrichment** with multiple inheritance relations.

**PAROLE-SIMPLE-CLIPS (PSC)**  
Computational Lexicon for Italian

Generative lexicon → **Qualia structure**  
**Formal**  
**Constitutive**  
**Agentive**  
**Telic**

*Toral A., Monachini M., Muñoz R. 2007. Automatically converting and enriching a computational lexicon Ontology for NLP semantic tasks. LTC' 07*  
*Toral A., Monachini M. 2007. Formalising and bottom-up enriching the ontology of a Generative Lexicon . RANLP07*  
*Mascali, T. Automatic acquisition of qualia roles using an Italian semantically annotated corpus , Master Thesis*

# SIMPLE OWL (1)

Conversion of the **SIMPLE ontology** into the formal language OWL -  
**Enrichment** with multiple inheritance relations.

## PAROLE-SIMPLE-CLIPS (PSC)

Computational Lexicon for Italian

Generative lexicon → **Qualia structure**

**Formal** > Taxonomy

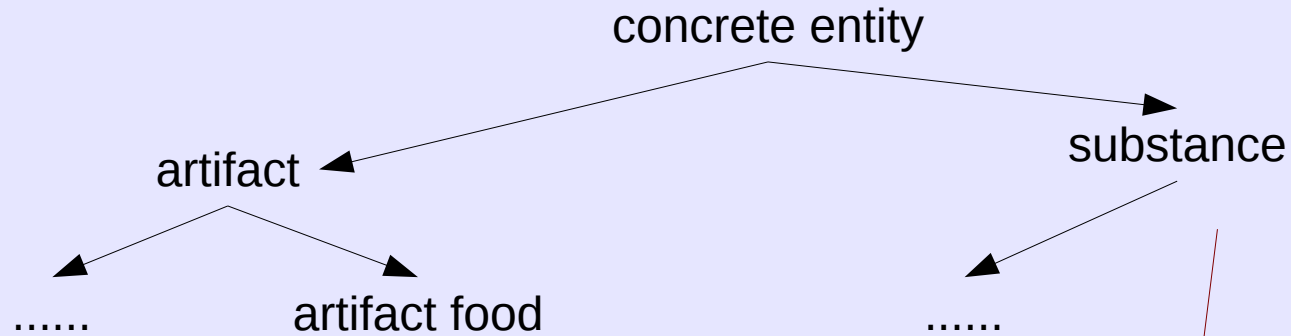
**Constitutive** > Constitution

**Agentive** > Origin

**Telic** > Function

# SIMPLE OWL 2

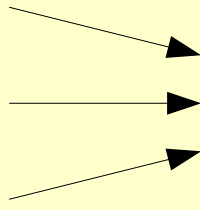
## SIMPLE ONTOLOGY



pizza

cake

...

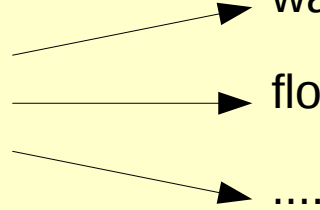


made\_of

water

flour

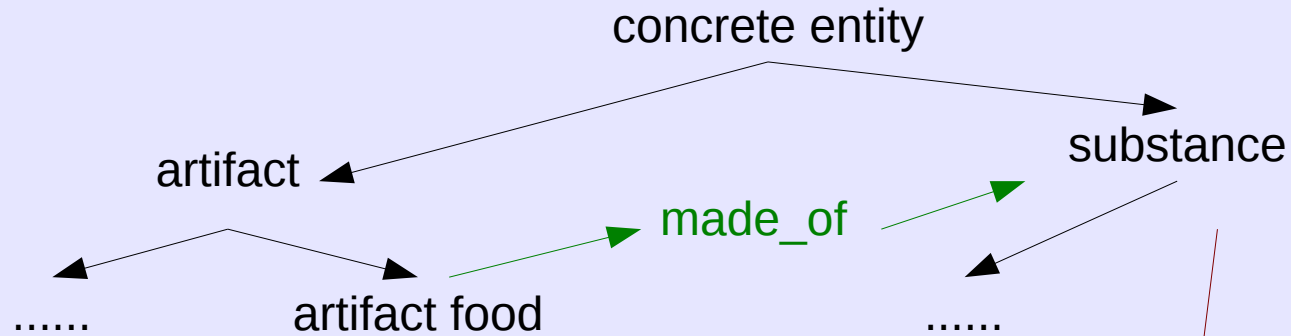
....



## SIMPLE LEXICON

# SIMPLE OWL 2

## SIMPLE ONTOLOGY



pizza

cake

...

made\_of

water

flour

....

## SIMPLE LEXICON

# Enriching with linguistic info

**Monnet** (Multilingual ONtologies for NETworked knowledge)

[www.monnet-project.eu](http://www.monnet-project.eu)

PROBLEM: mapping of labels of ontologies or taxonomies and natural language expressions occurring **in documents**

- **ontology lexicalisation**: takes as input an ontology and outputs an 'ontology-lexicon' for one language where **ontology labels** are enriched with linguistic annotation

- represented on the basis of the so-called '**lemon**' format (for the appropriate integration of lexical/linguistic and terminological information in ontologies)

GOAL: improvements both in the ontology-based *semantic annotation* of textual documents and in the **semi-automatic generation of ontologies** from text.

# Merging lexicons and ontologies

**Text mining issues: lack of standardization – lack of rich linguistic information**

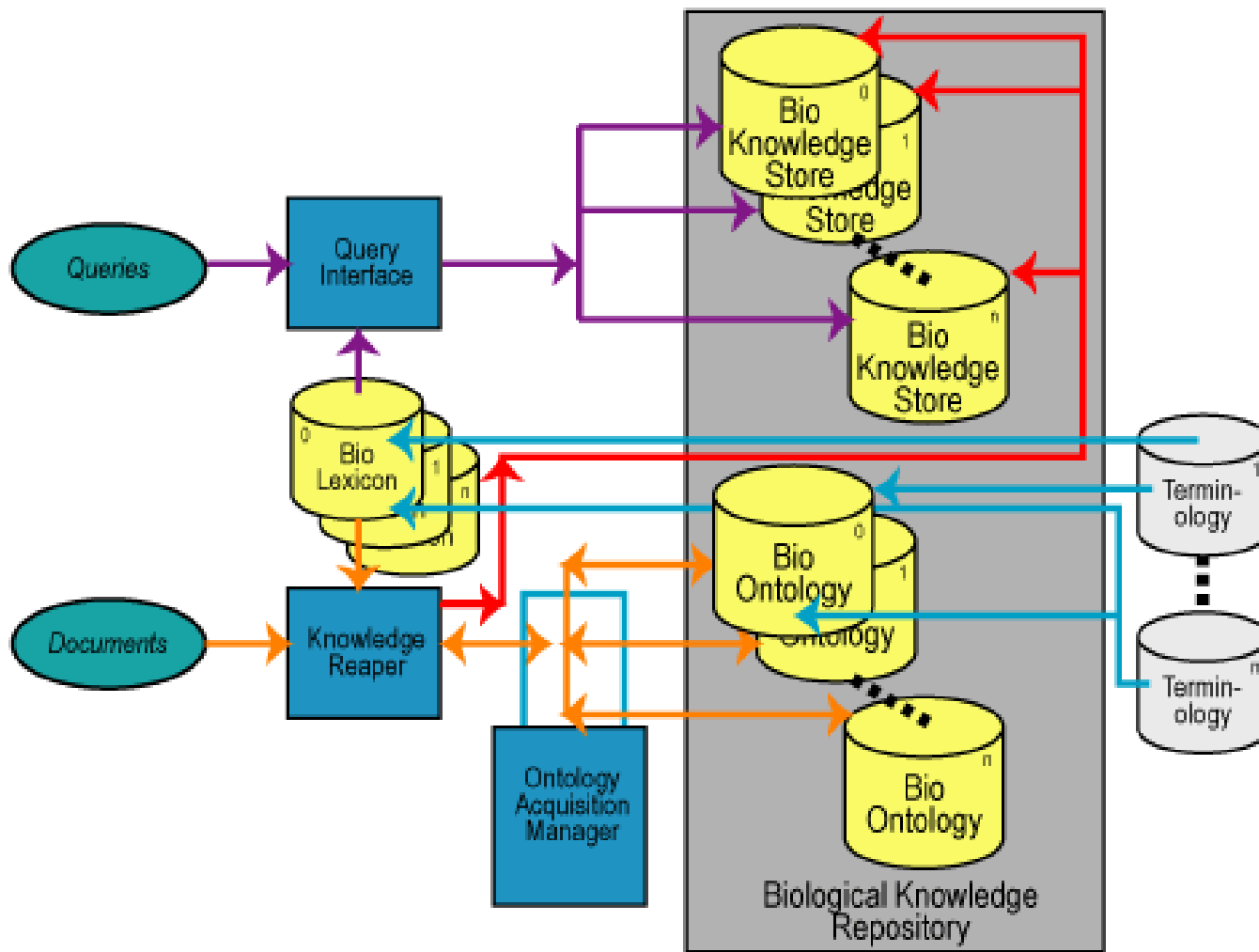
GOAL: integrate terminologies found in existing resources and exploit text mining techniques for automatic and incremental updates based on the literature

Generating a **unified resource** from several existing resources, such as databases, controlled vocabularies and ontologies

Enriching it with corpus driven entities (NER)  
Add **structured linguistic information**

# BOOTSREP: biology domain

[www.bootstrep.eu](http://www.bootstrep.eu)



- to generate **bio-lexicon** and a **bio-ontology** using text mining techniques

- to **harmonise** and augment existing bio-terminologies and bio-ontologies

- to allow **multilingual** querying of scientific knowledge and databases

**LMF**, an ISO standard

## Mining of environmental texts:

- enrichment of **lexicons**
- enrichment of **ontologies**
- **mapping** of lexicons and ontologies
- mapping of lexicons in **different languages** via ontology
- linguistic, semantic and conceptual **annotation** of texts in different languages
- multilingual text mining



# KYOTO - Knowledge modeling

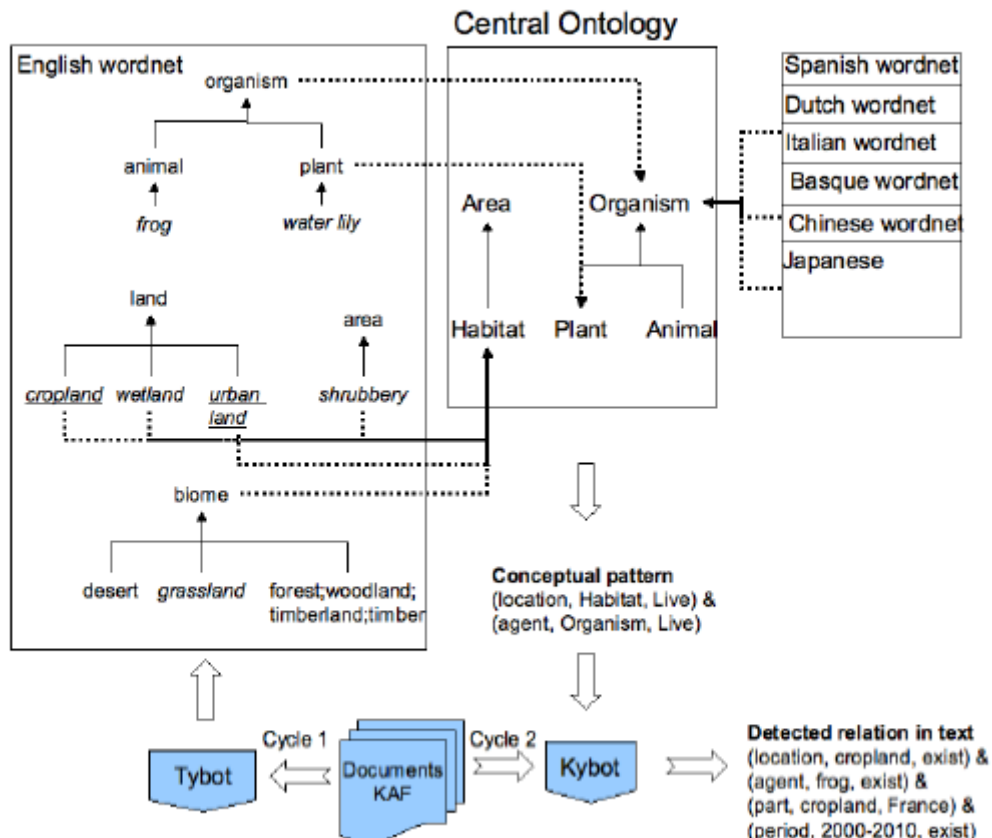


Figure 1: Two Cycles of processing in KYOTO

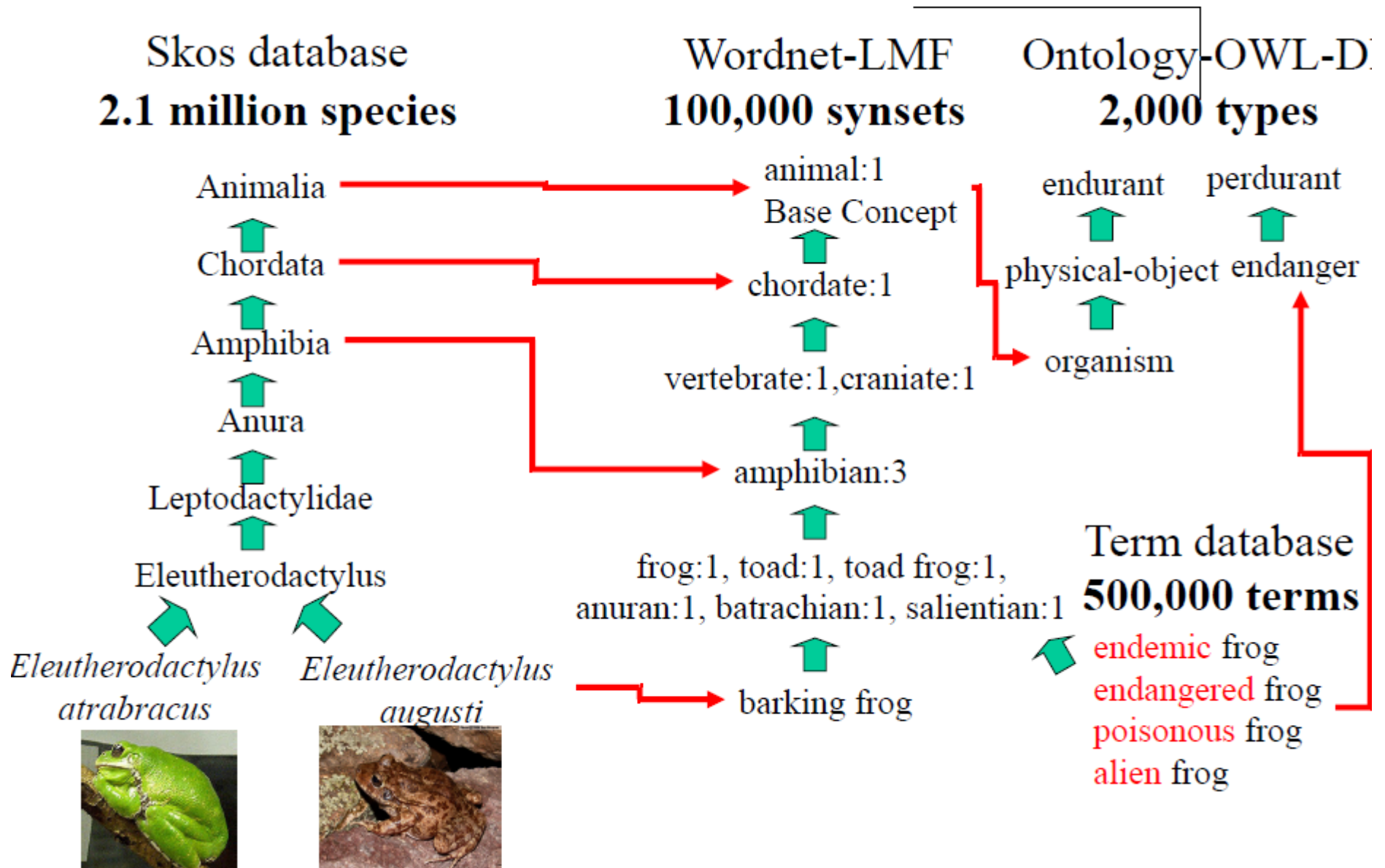
Performed  
by experts via Wikyoto

Term extraction from corpus  
> outputs term hierarchies

Mapping of terms onto  
WordNet and Ontologies  
> **Enriched Domain  
WordNets**

**Mapping** of WordNets onto  
an Ontology (enriched  
version of DOLCE) via BASE  
CONCEPTS

# KYOTO – division of labour



# KYOTO – annotating knowledge

Enriching documents in KAF notation - layers:

- **NER** – for georef (link to **Geonames**)
- **WSD** – disambiguates and adds semantic information > (domain) WordNet synsets
- **Ontotagging** : uses the linking between synsets and ontology noted to add concepts

# KYOTO: Fact mining

**Kyobot profile**: simple xml structure that specifies:

- a set of **variables**, with for each variable a set of constraints
- **relations** between variables
- some kind of **output** format

Profiles can combine logical schemes with linguistic patterns

Kybots can run sequentially

# KYOTO: Kybot input

```
<kprofile>
  <variables>
    <var name="x" type="term" pos="N"/>
    <var name="y" type="term"
      lemma="produce | generate | release | ! create"/>
    <var name="z" type="term"
      reference="DOLCE-Lite.owl#contamination_pollution"
      reftype="SubClassOf"/>
  </variables>
  <relations>
    <root span="y"/>
    <rel span="x" pivot="y" direction="preceding"/>
    <rel span="z" pivot="y" direction="following"/>
  </relations>
  <events>
    <event target="$y/@tid" lemma="$y/@lemma" pos="$y/@pos"/>
    <role target="$x/@tid" rtype="agent" lemma="$x/@lemma"/>
    <role target="$z/@tid" rtype="patient" lemma="$z/@lemma"/>$
  </events>
</kprofile>
```

# KYOTO: ontotagged KAF

```
<term lemma="pollution" pos="N" tid="t13444" type="open">
  <externalReferences>
    <externalRef reference="eng-30-00191142-n" reftype="baseConcept" resource="wn30g"/>
    <externalRef reference="Kyoto#change-eng-3.0-00191142-n" reftype="sc_subClassOf"
resource="ontology">
      <externalRef reftype="SubClassOf" reference="DOLCE-Lite.owl#contamination pollution"/>
      <externalRef reftype="SubClassOf" reference="DOLCE-Lite.owl#accomplishment" status="implied"/>
      <externalRef reftype="SubClassOf" reference="DOLCE-Lite.owl#event" status="implied"/>
      <externalRef reftype="SubClassOf" reference="DOLCE-Lite.owl#perdurant" status="implied"/>
      <externalRef reftype="DOLCE-Lite.owl#part" reference="DOLCE-Lite.owl#perdurant" status="implied"/>
      <externalRef reftype="DOLCE-Lite.owl#specific-constant-constituent" reference="DOLCE-
Lite.owl#perdurant"
        status="implied"/>
      <externalRef reftype="DOLCE-Lite.owl#has-quality" reference="DOLCE-Lite.owl#temporal-quality"
status="implied"/>
      <externalRef reftype="SubClassOf" reference="DOLCE-Lite.owl#spatio-temporal-particular"
status="implied"/>
      <externalRef reftype="DOLCE-Lite.owl#participant" reference="DOLCE-Lite.owl#endurant"
status="implied"/>
      <externalRef reftype="DOLCE-Lite.owl#has-quality" reference="DOLCE-Lite.owl#temporal-location_q"
status="implied"/>
      <externalRef reftype="SubClassOf" reference="DOLCE-Lite.owl#particular" status="implied"/>
    </externalRef>
  </externalReferences>
</term>
```

# KYOTO: Kybot output

```
<kybotOut>
  <doc name="11767.mw.wsd.ne.onto.kaf">
    <event eid="e1" lemma="generate" pos="V" target="t3504"/>
    <role rid="r1" lemma="industry" rtype="agent" target="t3493" pos="N"
event="e1"/>
    <role rid="r2" lemma="pollution" rtype="patient" target="t3495" pos="N"
event="e1"/>
  </doc>

  <doc name="16266.mw.wsd.ne.onto.kaf">
    <event eid="e2" lemma="release" pos="V" target="t97"/>
    <role rid="r3" lemma="fuel" rtype="agent" target="t96" pos="N" event="e2"/>
    <role rid="r4" lemma="exhaust_gas" rtype="patient" target="t101" pos="V"
event="e2"/>
  </doc>
</kybotOut>
```

# KYOTO: Semantic Search

Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti Aiuto

http://kyoto.inron.nl/kyoto-kybot/web/name.do

superare - Dizionario I... Kyoto - TWiki SIS Miller Italia Webmail L...

ff - Diz... PREZZI & S... Offerte Hot... Almanacco ... Web search... http://...kybot/ htt...do Kyoto Searc... Kyoto Searc... http://...line=8

**KYOTO** Query: inquinamento in america BNC 100

Search

cause

- 3 done-by:land
- 3 done-by:water
- 3 done-by:watershed

Location

- 9 Chesape
- 4 Humber
- 4 Virginia
- 2 Killingho

Result

- 2 patient w
- 1 patient a
- 1 patient b
- 1 patient c

100 Events

Prob. v	Event	Cause	Result	Location	Date	Other	F
1.52	<a href="#">pollution</a>	done-by:chemical fertilizer					2
1.52	<a href="#">pollution</a>					participant:load	2
1.32	<a href="#">pollution</a>	simple-cause-of subdivision and		<a href="#">Chesapeake (populated place)</a>			2
						participant:reduction effort	2
						participant:area	2
							2
1.27	<a href="#">pollution</a>	done-by:phosphorus, done-by:sediment, and done-by:water		<a href="#">Pennsylvania (first-order administrative division)</a>			2
1.18	<a href="#">pollution</a>	done-by:water					2

**Label:** *pollution*

**Phrase:** The Chesapeake Bay and its tributaries are unhealthy primarily because of pollution from excess nitrogen, phosphorus and sediment entering the water..

**Cause:** done-by:phosphorus, done-by:sediment, and done-by:nitrogen.

**Location:** [Pennsylvania \(first-order administrative division\)](#).






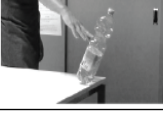



**Page:** 2:6.

start



# Going Multimodal: **IMAGACT**

lablita.dit.unifi.it

ACTION TYPE	DEFINITION (and example)	EQUIVALENT VERBS
A1 	Continuing generic pressure, with the sole result that the object (or body part) is pressed <i>(the doctor palpates the abdomen)</i>	premere, pigiare, spingere
A2 	Pressure on only one side that brings about the reduction of the volume of the object <i>(Jane compresses the garbage)</i>	premere, pigiare, schiacciare, pressare, comprimere, spingere
A3 	Pressure on two or more sides that brings about a reduction of the volume of the object <i>(John squeezes the toothpaste tube)</i>	schiacciare, spremere, comprimere
A4 	Non continuous (brief) pressure <i>(Jane presses the button)</i>	premere, pigiare, schiacciare, spingere
A5 	Continuous pressure that accompanies the object in the transition <i>(Jane pushes the trolley)</i>	spingere
A6 	Impulse that distances the object from the agent <i>(Jane pushes away the bottle)</i>	spingere
A7 	Pressure that causes damage to the object <i>(John crushes the tomato)</i>	schiacciare
A8 	Pressure that inserts the object into something <i>(John pushes the plasticine into the mould)</i>	premere, pigiare, spingere
A9 	Internal pressure <i>(the Yogi pushes the stomach out)</i>	spingere

Building an inventory of **action types**:

Corpus driven manual analysis of action verbs






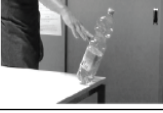



Corss-lingual comparison

Action types mapped onto WordNets and Ontology

**IMAGES + VIDEO** can enrich the resources

# Going Multimodal: **IMAGACT**

lablita.dit.unifi.it

ACTION TYPE	DEFINITION (and example)	EQUIVALENT VERBS
A1 	Continuing generic pressure, with the sole result that the object (or body part) is pressed <i>(the doctor palpates the abdomen)</i>	premere, pigiare, spingere
A2 	Pressure on only one side that brings about the reduction of the volume of the object <i>(Jane compresses the garbage)</i>	premere, pigiare, schiacciare, pressare, comprimere, spingere
A3 	Pressure on two or more sides that brings about a reduction of the volume of the object <i>(John squeezes the toothpaste tube)</i>	schiacciare, spremere, comprimere
A4 	Non continuous (brief) pressure <i>(Jane presses the button)</i>	premere, pigiare, schiacciare, spingere
A5 	Continuous pressure that accompanies the object in the transition <i>(Jane pushes the trolley)</i>	spingere
A6 	Impulse that distances the object from the agent <i>(Jane pushes away the bottle)</i>	spingere
A7 	Pressure that causes damage to the object <i>(John crushes the tomato)</i>	schiacciare
A8 	Pressure that inserts the object into something <i>(John pushes the plasticine into the mould)</i>	premere, pigiare, spingere
A9 	Internal pressure <i>(the Yogi pushes the stomach out)</i>	spingere

IWN  
pigiare [1]  
premere [1]  
spingere [5]

ILI - WN1.5  
Press[1]  
apply\_force\_to[1]  
apply\_pressure\_to[1]

Building an inventory of **action types**:

Corpus driven manual analysis of action verbs

Corss-lingual comparison

Action types mapped onto WordNets and Ontology

**IMAGES + VIDEO** can enrich the resources