Multilingual Terminology

Pierre Zweigenbaum

STIM : Mission de recherche en sciences et technologies
de l'information médicale,
DSI, Assistance Publique – Hôpitaux de Paris ;
CRIM, INaLCO ;
U729, INSERM
Paris, France

pz@biomath.jussieu.fr
http://estime.spim.jussieu.fr/~pz/

LREC 2006 : COCOSDA/WRITE workshop - 28/5/2006

More than 60 papers submitted to the Terminology track
(growing number)

General, misc	4
Terminology & knowledge acquisition	5
Relations extraction	5
Terminology and ontologies	9
Terminology extraction, creating terminologies	16
Total	39

Bente Maegaard, 27/5/2006

Terminology-oriented workshops

- Acquiring and representing multilingual, specialized lexicons: the case of biomedicine
    - 25 attendees whole day
    - distributed, collaborative development of multilingual {lexicon|terminology} (wiki-style)
- Terminology design: quality criteria and evaluation methods (TermEval)
    - today
- [OntoLex 2006: Interfacing Ontologies and Lexical Resources for Semantic Web Technologies]
    - ~40 attendees
    - "annotate" ontology with language information

Terminology: position and delimitation

Enumerate the terms that express the concepts of a given domain

- Domain-specific *vs* general
- Generally onomasiological *vs* semasiological
  - although corpus-based terminology construction an active domain

Terminology: position and delimitation

Enumerate the terms that express the concepts of a given domain

- Domain-specific *vs* general
- Generally onomasiological *vs* semasiological
  - although corpus-based terminology construction an active domain

Issues:

- Fuzzy border with specialized lexicons
- Fuzzy border with ontologies:
  - People often call "ontology"
    - a taxonomy
    - a structured terminology

$\rightarrow$ Structured terminology (often corpus-based) as a first step towards an ontology (whatever that may be)

Shareability and evaluation

Terminologies are purpose-oriented

E.g., 100+ different terminologies in the field of medicine in the UMLS Metathesaurus

- $\rightarrow$ Limits their shareability
- $\rightarrow$ Not only provide ready-made terminologies,
  - but also/instead provide methods and tools to design/build/adapt terminologies
- $\rightarrow$ Extremely difficult to evaluate,
  e.g., corpus-based term extraction
  - as opposed, e.g., to controlled indexing, where terms are given

- Natural given the onomasiological status of terms:
  - Concepts are "decorated" with terms from different languages
- Adopt a per-domain approach:
  involve relevant user groups
  - *E.g.*, current work on translation of LOINC terminology:
  - Involve biologists for translation of terminology of laboratory tests

How to foster cross-disciplinary collaboration

- ◁
  - Attract researchers from neighbouring,
    "application-oriented" fields
  - Promote and disseminate standards in these areas

- ▷
  - Raise interest and awareness of computational linguists for
    domain-oriented work (medicine/bioinformatics, law...)
  - [Computational linguists:] Publish in "application-oriented"
    conferences and journals

Summary

- Task-dependence limits shareability:
    - to which extent is this an impediment to actual reuse?
- Explicitly take into account purpose (task) when evaluating terminologies
    - and produce terminology tools instead of only terminologies
- Better articulate lexicon, terminology, ontology
- Multilingualism: terminologies are in a privileged position (but build on user groups)
- Specialized domains: foster cross-disciplinary collaboration

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS **Inserm**