Gerhard Budin
University of Vienna
Austrian Academy of Sciences

Lexicograffiti
European Parliament
Luxembourg
23rd of January, 2012
Overview

- My background, my perspectives

**First Part:**
- Trends analysis: eLexicography, Terminology, and Global Content Management
- Summary of Terminology Management principles

**Second Part:**
- Examples and Case Studies

**Conclusions**
My research background
My perspectives

- Socio-cognitive orientation: process-, user- and task-oriented, functional approach

- Computational orientation: data modelling, corpus-based perspectives

- Media orientation: multi- and cross-media, single-sourcing, re-usability of content

- Management perspectives: knowledge management; content management; terminology management; workflow management; quality management

- Empirical approach: inspired by best practices
First Part: Trends analysis: eLexicography

- Concurrent paradigm shifts:
  - From traditional lexicography to eLexicography:
    - corpus-driven analysis of linguistic usage
    - dictionaries as information and reference tools
    - on-the-fly generation of tailor-made and task-oriented lexical content
First Part: Trends analysis:
Terminology, and Global Content Management

- From top-down terminology work to
  terminology-based content management
  - domain-specific or multi-domain corpus-based term extraction and linguistic- and cognitive resource enrichment

- cross-media, multi-functional, multi-format, and single sourcing-oriented global content resource generation based on terminology management principles and methods
Some Terminology Management principles and methods

- What is terminology?
- What is terminology management?
- Who is doing multilingual terminology management and why?
- How to identify best practices? Which criteria can we use?
Terminology and its functions

- What is a terminology?
  - a structured set of concepts and terms of a specific subject field in a specific language
  - Terminology as an abstract noun denotes the subject field of terminology studies, terminology work, etc.
## Terminology and its functions

<table>
<thead>
<tr>
<th>Communication and discourse (mutual understanding)</th>
<th>Information (data storage, logistics)</th>
<th>Cognition (concept formation, creative thinking and naming the world)</th>
<th>Knowledge (dissemination, learning, storage, coding, etc.)</th>
<th>Professional work in science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>Industry</td>
<td>Business</td>
<td>Trade</td>
<td>Public and social affairs</td>
</tr>
<tr>
<td>Culture</td>
<td>Sports</td>
<td>Language policies</td>
<td>Language development</td>
<td></td>
</tr>
</tbody>
</table>

What is Terminology Management?

- A broad concept, covering a wide range of practical activities for manipulating terminological information for specific purposes

- Operationalizes theoretical principles of terminology as methodologies

- A type of Information Management

- A Key to Language Management, Knowledge Management, Content Management
Types of Terminology Management

- **Descriptive** Terminology Management
  - Aiming at documenting terminological diversity, for research purposes or for creating a sound basis for decisions to be taken

- Translation-Oriented Terminology Management: comparative approach, documenting cross-cultural differences in terminological structures in source and target languages

- Corpus-driven Terminology Management: usage-oriented, term extraction from real life discourse, computational terminology paradigm
Types of Terminology Management

- **Prescriptive** Terminology Management
  - a normative approach as part of language planning and technical and scientific standardization, decisions are taken on the basis of existing information sources, terminology collections, etc., aiming at reducing terminological complexity and diversity)
  - Standardizing terminological information  
    (engineering, natural sciences, medicine, etc.)
  - Standardizing the methods of terminology creation, of terminology management (term formation, quality management, process management, terminography)
Approaches (to all types of TM)

- Ad-hoc approach
  aiming at instant problem solving, as part of other processes [translation, technical documentation, technical standardization, etc.]

- Text-oriented approach

- Systematic approach
  consistent application of work methods, systematic problem solving, interaction of workflows

- Domain knowledge-oriented approach
Operational Principles

- Concept-orientation of any kind of terminological activity
  - Difference from lexicography (although mixed forms exist and are applied as well)
  - Knowledge-oriented (conceptual structures are knowledge structures)

- Systematic and adaptive approach
  - Applying work methodologies in a consistent way
  - Methods to be adapted and fine-tuned to specific purposes, application environments, projects, socio-cultural traditions, economic constraints, existing workflows, etc.
Operational Principles

- The sources of terminological information are always *documented*
  (in line with scientific traditions, as a pre-requisite for decision support in workflows, and for subsequent re-use of terminological resources)

- Management aspects
  (work flows and project management, economic and legal aspects, training and staff motivation, marketing, exploitation, dissemination)

- Cooperative approach

- Purpose-driven selection and application of *tools*

- Re-usability, open standards
Second Part: Examples and Case studies from our own research & development

more details in the afternoon session if desired

- I Multilingual glossary, dictionary, termbase, and ontology for global risk management communication

- II The LISE project – quality management for legal and administrative terminologies
The Making of... a Multilingual Glossary on Risk Management
Motivations and Methods: Terminologies for Risk Communication

The Role of LSP Lexicography in domain communication

- Increasing the “transparency” of terms
- Help negotiate a common understanding of terms in intra-, inter-disciplinary and transcultural discourse
- Help increase the consistency of risk discourse and increase understanding in target audiences
- Reduce unnecessary synonyms, disambiguate polysems, help separate homonyms
- Help create risk terminologies in many languages
- Support knowledge sharing and knowledge transfer in cooperative work environments
- Support translation work
The Domains of Risk Management

- Multidisciplinary, diverse, and fragmented – or
- Transdisciplinary, overlapping, converging, integrated, and complementary

The need for mediating between different approaches, cultures, and discourses:

- Technological, engineering, research, science
- Administration, legislation, monitoring
- Social, sociological, political, cultural
- Domain approaches (financial, ecological, chemical, safety, geographical, planning and forecast, health, etc.)
WIN Project
WP “Human Language Interoperability”

- Objectives
  - designed to support international risk management and risk communication processes

- Achieved results (with ongoing work)
  - Large parallel corpora collection with risk-related texts and lexical resources (fr, en, de, es, ro, fi, hu, ru)
  - Multilingual index with conceptual structure
  - Bibliography and codes of sources
  - Risk Ontology
  - Multilingual online terminology database
The global risk communication scenario

- Thesaurus building
- Creating multilingual terminology databases
- Creating multilingual text corpora
- Lexicographical glossary
- Semantic enrichment (conceptual links, frame semantics)
- Collection of relevant knowledge organization systems
- Annotation of resources
- Mark-up of resources (TBX)
- Ontology building
- Communication design
Integrated Workflows

Termbase:
- Export XML
- Domain Models – meta-models -> patterns

Text corpus:
- Term extraction – comparative testing ProTerm, MultiTerm Extract, MultiCorpora

Aligning with termbase
- Convert to RDF
- Ontology import -> editor

Mappings (TBX/TMF, XML, RDF, OWL, UML, comma delimited, RDB)
Term: acceptable risk

Definition: A technological activity that generates a risk for others can be termed acceptable if the activity respects ethical principles that are necessary for peaceful coexistence and/or if the activity renders social progress in a non-arbitrary way. A risk generating activity is acceptable in this sense if and only if the following holds: (1) all those who are subjected to the risk have given their informed consent to the activity and the conditions under which it is performed; (2) those who engage in such an activity without this informed consent can be held to full (unlimited, no caps) and unconditional (absolute) liability for any negative effects that the activity may cause to those who did...
The Glossary

- The glossary is used by risk managers, civil engineers, but also teachers, students, translators, journalists, etc.

- The purpose of such multilingual conceptual glossaries is to improve domain communication and to facilitate mutual understanding across linguistic boundaries.

- The multilingual glossary presented here includes 8 languages: English and French as main pivot languages, as well as German, Spanish, Romanian, Finnish, Hungarian, and Russian.

- It comprises about 230 central concepts of risk management with about 400 definitions and about 1400 terms representing these concepts in each language (including synonyms and hyperonyms), indicating the conceptual relations between the entries.
Thematic macro-structure of the glossary:
- Risk assessment and technology assessment
- Public perception of risk, planning, alarm,
- Risk events, equipment and operations, general terms
- Fire - events, equipment and operations
- Floods - events, equipment and operations
- Oil spills - events, equipment and operations.

Each glossary entry follows a micro-structure with elements:
- A concept number for a theme from the macro-structure
- The terms in 8 languages, with grammatical information
- The definitions in each language and their sources
- Related terms and expressions.
II
LISE
Legal Language Interoperability Services
Main objectives of the Project

- help terminology managers in public institutions improve the quality of their terminological resources

- The web-based, interactive terminology service is *work-flow* oriented and provides input and feedback from best practices in legal terminology management.

- To facilitate data expansion and terminological enrichment (adding more data such as definitions, concept relations; language expansion; re-purposing of data; working towards communicative goals such as transparency, clarity, precision, uniqueness, etc.

- Harmonisation work is supported upon request by and in cooperation with data owners

The project has received funding from the European Community (ICT-PSP 4th call) under Grant Agreement n° 270917.
Technological approach

- **Service Definitions**
  - The LISE Web Application design follows the SOA (Service-oriented architecture)
  - All LISE Services follow W3C definitions and the Web Services Description Language (WSDL).

- **Platforms**
  - LISE Services are designed to be platform-independent
WP 3- Legal Terminologies and Workflows

Tasks

- Analysis of existing workflows of terminology work
- Linking terminology workflows to test scenarios of the LISE web services
- Preparing a best practice guideline for optimised terminology management workflows

Deliverables

- D3.1 Report Analysis of Existing Terminology Workflows
- D3.2 Report Workflow Management for LISE
- D3.3 Guidelines for Collaborative Terminology Work
Conclusions

- not to separate or de-contextualize terminological resources from their “natural habitats” in order to use their full potential in further use (e.g. to explore the “real” meaning of a term in a particular occurrence)

- Interoperability is still an abstract notion in most cases: TBX needs better tools embedded in workflows

- Towards an ecosystem of terminological web services

- human dimension: pragmatic interoperability still missing → mutual understanding of people → socio-cognitive dimension to be focused on in R&D

- Integrative approaches in language industry
Thank you for your attention!

Questions & Answers
More information (publications, web links, resources, etc.) available in the afternoon session and upon request later on (by e-mail)

Gerhard Budin
Center for Translation Studies, University of Vienna
Institute of Corpus Linguistics and Text Technology
Austrian Academy of Sciences

gerhard.budin@univie.ac.at