Real-World Terminology Management

Sue Ellen Wright
Kent State University
Terminology Summer School

Acronyms

- TMM=Terminology Management
- TWF=Terminology Workflow
- KOS=Knowledge Organization System
- SKOS=Simple KOS
- OWL=Web Ontology Language
- i18n=Internationalization
- QA=Quality Assurance
- SL=Source Language
- TL=Target Language
- ISO=International Organization for Standardization
- ROI = Return on Investment
Terminology Management Spectrum

Continuum of practice in a range defined by multiple factors:
- Enterprise type and mission
- Criticality of terminology for core business
- Criticality of quality and branding concerns
- Recognition of tangible and intangible ROI
- Terminology user groups
- Buy-in by stakeholders (design, engineering, marketing, as well as client & in-country partners, but above all, top-level management)

Enterprise Type

Nature of enterprise
- Government
- Industry
- Research institutes
- Localization/translation bureaus
  - Nature of the client
  - Text type
  - Negotiated client/vendor agreements
- Web content management providers
- Freelance & in-house technical writers, translators
- Librarians & knowledge organization environments
TMM Mission & Strategic Position

- Support for national language policy
- Support for global enterprise activity
- Support for technical writing & translation
  - Ongoing, stable subject fields
  - Frequent one-off, unrelated, unanticipated subject fields
- TMM approaches
  - Prescriptive vs. descriptive
  - Ad hoc vs. systematic

TWF Task Issues

- Specific task types
  - Language planning
  - Standardization
  - Document production
  - Controlled language document production
  - Activity in localization & multilingual documentation environments
  - Support for machine vs. human-oriented translation
  - Content management in dynamically changing Web environments
  - Terminology management for enterprise solutions (whatever that is)
Knowledge-oriented TWF

Enterprise-related information and knowledge management
- Terminology as a function of taxonomy, ontology, and information retrieval (knowledge organization systems: KOS)
- KOS interaction with:
  - Monolingual technical writing & product development
  - Translation and localization
  - Multilingual technical writing
  - E-business solutions
  - Inventory control and logistics
  - General information retrieval and processing
  - Standards issues (SKOS and OWL)

Cost Effectiveness:
- Terminology management activity costs time and money. Is it economically feasible:
  - For individual translators?
  - For groups working together?
  - For bureaus and translation services in government and industry?
Criticality of Terminology

- Potential for market losses
- Potential for communicative losses
- Risk of product failure, human injury
- Adverse effects on branding efforts (marketing issues)
- Relative significance of terminology
  - To the process
  - To the product
  - Example: Terminology is more critical if you are selling software than if you are selling wheat.

ROI: Variable vs. Fixed Costs

- TMM tasks hidden in general overhead
- Cost of *not agreeing to uniform terminology* hidden in overhead costs
- Costs of correcting mistakes or recouping damage to branding difficult to document
- Compounded costs due to the persistence of defective communications
Fixed Costs

- The costs of doing systematic terminology work, on the other hand, are up front costs that can be calculated using traditional means.

Calculating Return on Investment

- Greater general applicability of specific terminological units (greater frequency) = greater the return on terminology management costs.
- Greater the quality, safety, or competition-related criticality = greater the return.
- The greater the degree of integration between straight CAT, TM and MT applications, the greater the payback in leverageable data.
These factors result in varying:
- Integrated and non-integrated workflow models
- Complexity with respect to the data model
- Investment of human resources in the form of
  - Term extraction
  - Context documentation
  - Concept definition (ideally by SL & TL subject experts)
  - Terminology product delivery
    - On-screen lookup
    - Web-based resources
    - Hard-copy output

A continuum of practice in a range defined by:
- Creation of non-systematic tab-delimited, text and job-oriented glossaries and spreadsheets
  - SL term + TL term
  - + POS, Gender, and/or a note
- Theoretically rich complex entry input models
  - Subject field classification
  - Client designation
  - Concept-oriented definition
  - SL term + TL term + POS + Term Type + Register + Context, etc.
- Documentation through source citations, graphics, etc.
What is workflow?

- **Workflow** is the operational aspect of a work procedure:
  - how tasks are structured
  - what their relative order is
  - how they are synchronized
  - who performs them (& where)
  - how information flows to support the tasks
  - and how tasks are being tracked

http://en.wikipedia.org/wiki/Workflow

Macrostructures & Task Structures

- Terminology workflow (TWF) as a task-set in overall enterprise workflow
- Terminology management (TMM) mission within the enterprise
- Workflow parameters
- Task issues
- Input issues
Evolving Workflow Models

- TWF as non-detailed item in overall workflow
- TWF plotted as task set during project planning
- TWF plotted independently for individual projects
- TWF plotted for ongoing activities involving ongoing global TMM
- Project-oriented workflow vs. ongoing global enterprise-related terminology management

Critical Workflow Parameters

- Dimensions:
  - Tasks (activities)
    - Time
      - How much time does it take?
      - How much time do you have?
    - Throughput
  - Resources
    - Human
    - Tools
    - "Raw Materials"
Critical Workflow Parameters

Dimensions:
- Cost accounting as a function of workflow management
  - Difficult to calculate because of individual task differences
  - Balance between intensity & granularity of TMM vs. the investment costs incurred for “doing” TMM
  - Balance between investment and incidence of reuse
  - Balance between investment and criticality of terminological accuracy

Project Inputs

Raw source text
- Frequently flawed (particularly in localization environments)
- Subject to ambiguity: polysemy, synonymy
- Term extraction from the source text
  - Human or automatic term recognition
  - Synonym identification and disambiguation
  - Co-occurrence issues with advanced automatic indexing software
- Determining target language equivalents
  - Verification issues as a function of workflow
Ordering and Synchronization

- Traditional TMM workflow position
- Rationalized workflow position
- Global TMM solutions
- Systematic TMM
- Information feedback loop

Traditional Position of Terminology Management (TMM) in Global Project Workflow

- Ad hoc TMM
- Reactive project-specific TMM
- No influence on document production, i18n

Workflow diagram ©Keiran Dunne 2005
Rationalized Project-Oriented TMM

- TMM as a function of QA (Quality Assurance) management
- TMM and QA upstreamed to planning stage
- Proactive TMM

Global Enterprise Terminology Management Solutions

- Corpus Generation & Maintenance
- Corpus Management
- Terminology Extraction
- Terminology Management

- Term Proposal
  - Origination
  - Input
  - Modification
  - Check
  - Verification
  - Approval
  - Publication
  - Use
  - Suggestion
  - Etc.
Terminology Management Transactions (ISO 12620)

- Proposal of a term / extraction of a term
- Origination: creation of a term entry
- Input: data input
- Modification: update and change
- Check by terminologist
- Verification by subject field experts
- Approval
- Withdrawal
- “Publication”
- Use “in the field”
- Modification suggestions
  (Standardization and language planning)

Task Definitions

- Origination: entry creation
- Input: population of data fields
- Modification: update & correction
- Check: terminology check
- Verification: subject field check (in-country reviewer?)
- Approval: sign-off by chief terminologist
- Publication: importation into master database, upload to Web, Internet, intranet, LAN, hardcopy, etc.
Systematic Term Management for Collections of Controlled Size

Corpus Selection → Terminology Management
   - Concept Selection
   - Concept System
   - Definition Creation
   - Term Assignment
   - Term Creation
   - Term Status Assignment

Language Planning
Prescriptive Terminology

Dynamic Systematic Management for Large Collections

Corpus Collection and/or Selection → Terminology Management
   - Concept Field Analysis
   - Concept Positioning
   - Definition Creation
   - Term Assignment
   - Term Creation
   - Term Status Assignment
   - Entry Creation (see previous)
   - Dynamic KOS generation

Ontology Link

KOS=Knowledge Organization System
Information Feedback Loop

Who Does What

- Project developers (engineers, designers, other subject experts)
  - Terminological inconsistencies
  - Uncontrolled coinage of neologisms
- Terminology project group representative
  - Technical writers
    - Frequently unaware of multilingual issues
    - Language specific problems
  - Translators
  - Knowledge engineers
  - Terminologists
Who Does what

- Master terminologist
  - Authority to change records in master file
  - Authority for final approval
- Language-specific terminologist
  - Source and target language research
- Data input specialist
- Subject-field specialist
  - Source language
  - Target language
  - Conflation and expansion of tasks according to need

Where do they do it?

- Workflow solutions and application design dependent on work group configurations:
  - Individual terminologist in one location
  - Small work group
  - Single local work group
  - Larger distributed network
  - Multilingual globally distributed network
Publication, Dissemination and Access

- Local master terminology database
- Web-based terminology resource
  - Web-server
  - Enterprise-based Intranet
  - LAN server
- Exportation to other TDB environments
- Hardcopy
  - Special project sub-setting

Special Outputs

- Project-specific term entry subsetting and exportation (electronic)
- Selective subset delivery, especially to freelance developers, tech writers, and translators
- Output of SL→TL glosses in hard copy
  - For subject area specialists in multilingual environments
  - For interpreters
Delivery to Technical Writers and Translators

- Translators: search for equivalent terms based on source language terms and concepts
- Technical writers: search for source language terms based on developers’ concepts
  - Frequently flawed
  - Frequently newly coined terms and concepts
- Delivery to and buy-in from technical writers is very difficult to achieve.

Solutions for Technical Writers and Program Developers

- Ontology-enabled terminology delivery
- Style sheets
- Automatic term use checkers
- Issues:
  - Resistance to synonymy control
  - Coinage of translation-unfriendly terms
  - Coinage of “cute” unmotivated terms
Translation-oriented Solutions

- Terminology database integrated with translation memory and text production software
- Concordance features for access to terms in context
- Multiple layers of information
  - Data category subsetting for rapid information retrieval
  - Optional view of full terminological entries

Tracking Issues

- Workflow-related database design issues
- Transaction-related data categories
  - Terminology transaction
  - Responsibility
  - Date
  - Authorizations
  - Task-related check-offs
  - Output and sort-related categories
  - Import / export tracking
  - Version tracking
Computer-mediated Workflow Management Systems

- SDL WorkFlow Service™
- Trados Global Content Management™
- Star Translation Workflow Server™
- Lionstream Workflow™
- LTC Organiser Multilingual Management and Workflow Control Software System™

Non-exhaustive list!

For More Information

- Sue Ellen Wright
  Institute for Applied Linguistics
  Kent State University
  109 Satterfield Hall
  Kent, Ohio 44242, USA

sellenwright@gmail.com