

Language Resource and Terminology Standards

TC 37 and Beyond

Gerhard Budin and Sue Ellen Wright, TSS 2011

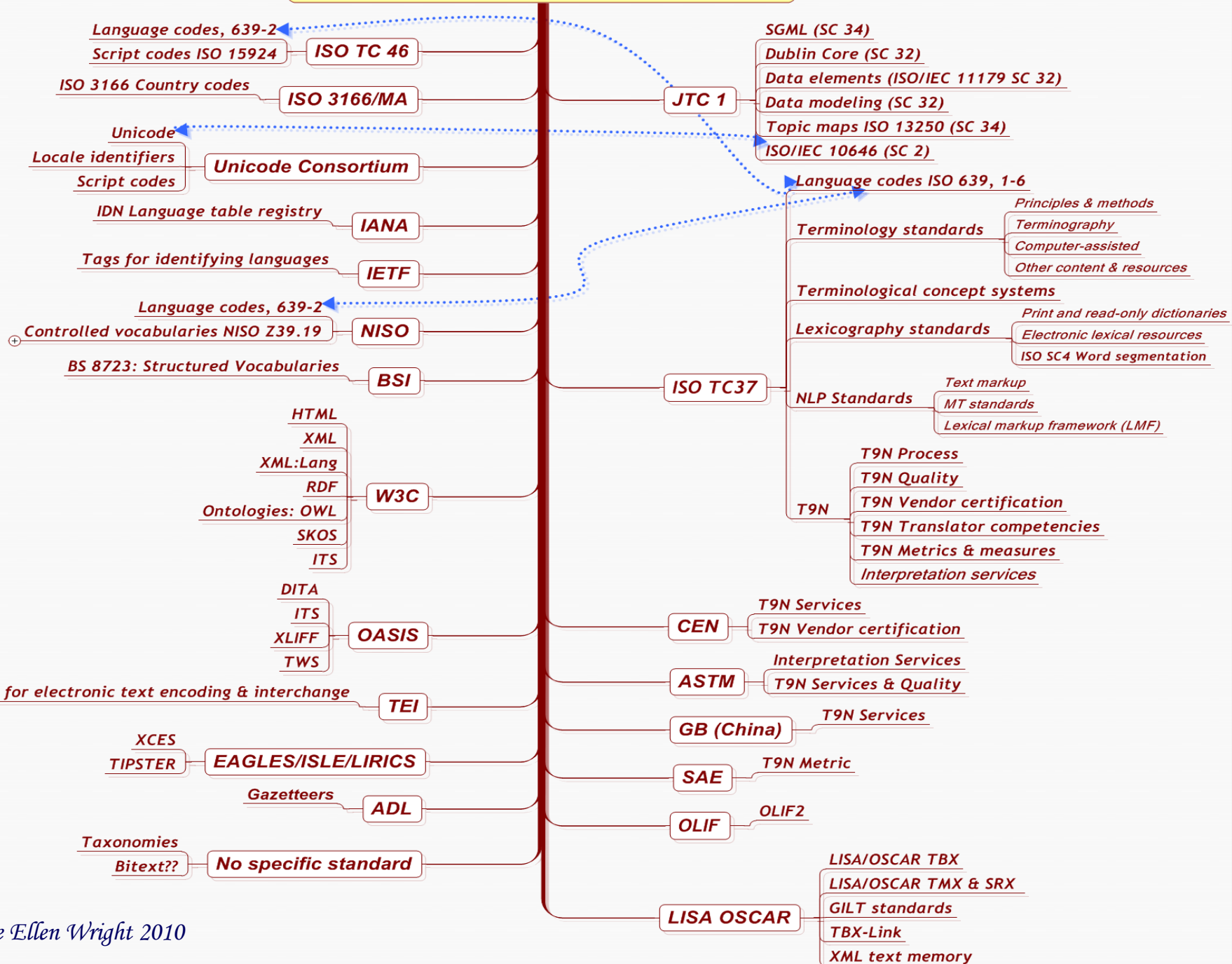


The Role of Standards

- Enabling Quality Management (products & services), as well as fulfilling safety requirements and consumer protection regulations
- Standards are the “nuts and bolts” of Interoperability: exchangeability, reusability
- Ideally, standards are implemented as (part of) technologies (functional standards), embedded in concrete applications and information ecosystems → standards as databases, standards as web services, etc.



Language and Knowledge Standards Groups



Standards Philosophies

- ISO: authorship by committee consensus with strong national component; copyrighted docs
- W3C, IETF, TEI, OASIS, others: public docs & widespread community comment & public consensus
- ISO model: pay to play, pay to use
 - 2nd wave hardcopy publishing mentality
- IETF: the standardizers are us
 - 3rd wave free for downloading approach
- -> Ongoing trend of convergence and cooperation, innovative business models integrated approaches



Language Standards in ISO/TC 37

- Different kinds of standards
 - By type of document
 - guide
 - standard
 - technical report
 - database
 - data model
 - software
 - By degree of normativity
 - bottom-up singular model
 - group agreement,
 - top down specification as a strong recommendation, or even as a law



Language Standards in ISO/TC 37

- By type of content/type of purpose & application
 - technical specification standards
 - process standards
 - conformance standards
 - safety standards
 - quality standards
 - service standards
 - (meta-) (data) (model) standards
 - terminology standards, incl. graphical symbols, cultural diversity management standards
 - coding standards



ISO Technical Committee 37, Terminology: Other language and content resources

- SC 1: Principles and methods
- SC 2: Terminographical and lexicographical working methods
 - Layout, lexicography, pragmatic applications; language codes, and translation management
- SC 3: Systems to manage terminology, knowledge and content
 - Computer assisted terminology management
- SC 4: Language resource management
 - Natural language processing and other language resources



ISO/TC 37/SC 1

ISO No.	Name
ISO 704:2009	Terminology work – Principles and methods
ISO 860:2007	Terminology work – Harmonization of concepts and terms
ISO 1087-1:2000	Terminology work – Vocabulary – Part 1: Theory and application
ISO 22134:2007	Practical guidelines for socioterminology
ISO/TR 24156:2008	Guidelines for using UML notation in terminology work
ISO 29383	Terminology policies – Development and implementation



ISO/TC 37/SC 2 – Language Codes

- ISO 639-1:2002 Codes for the representation of names of languages
– Part 1: Alpha-2 code
- ISO 639-2:1998 Codes for the representation of names of languages
– Part 2: Alpha-3 code
- ISO 639-3:2007 Codes for the representation of names of languages
– Part 3: Alpha-3 code for comprehensive coverage of languages
- ISO 639-4: 2009 Codes for the representation of names of languages
– Part 4: General principles of coding of the representation of names of languages and related entities, and application guidelines
- ISO 639-5:2008 Codes for the representation of names of languages
– Part 5: Alpha-3 code for language families and groups
- ISO 639-6:2009 Codes for the representation of names of languages
– Part 6: Alpha-4 code for comprehensive coverage of language variants



ISO/TC 37/SC 2

- ISO 1951:2007 Presentation/representation of entries in dictionaries – Requirements, recommendations and information
- ISO 10241-1.2 Terminological entries in standards – Part 1 General requirements and examples of presentation
- ISO 12199:2000 Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet
- ISO 12615 Bibliographic references and source identifiers for terminology
- ISO 15188:2001 Project management guidelines for terminology standardization
- ISO 23185:2009 Assessment and benchmarking of terminological resources – General concepts, principles and requirements
- ISO 22128:2008 Terminology products and services – Overview and guidance



ISO/TC 37/SC 2

Translation-Related Projects & Standards

Draft Number	Name
ISO 12616:2002	Translation-oriented terminography
ISO/CD TS 11669	Translation projects – General guidance
ISO/WD 13611	Interpreting/Interpretation processes – Community Interpreting/Interpretation
ISO/NP 14080	Assessment of translations



ISO/TC 37/SC 3

- ISO 12200:1999 Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange
- ISO 12620:2009 Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources -> ISOCat – standards-as-database
- ISO 16642:2003 Computer applications in terminology – Terminological markup framework
- ISO DIS 22274 Systems to manage terminology, knowledge and content – Design, implementation and maintenance of classification systems
- ISO 26162:2010 Systems to manage terminology, knowledge and content – Design, implementation and maintenance of terminology management systems
- ISO 30042:2008 Systems to manage terminology, knowledge and content – TermBase eXchange (TBX)



ISO/TC 37/SC 4

- ISO 24610-1:2006 Language resource management – Feature structures – Part 1: Feature structure representation
- ISO 24610-2 Language resource management – Feature structures – Part 2: Feature system declaration
- ISO/NWI 21829 Terminology for language resources
- ISO/FDIS 24614-1 Language resource management – Word segmentation written texts – Part 1: Basic concepts and general principles
- ISO/FDIS 24614-2 Language resource management – Word segmentation written texts – Part 2: Word segmentation for Chinese, Japanese, and Korean
- ISO/DIS 24611 Language resource management – Morpho-syntactic annotation framework
- ISO/DIS 24612 Language Resource Management – Linguistic Annotation Framework (LAF)
- ISO 24613:2008 Language resource management – Lexical markup framework (LMF)



ISO/TC 37/SC 4

- ISO/DIS 24619 Language resource management – Persistent identification and access in language technology applications
- ISO/DIS 24617-1 Language resource management – Semantic annotation framework (SemAF) – Part 1: Time and events
- ISO/CD 24617-2 Language resource management – Semantic annotation framework (SemAF) – Part 2: Dialogue Acts
- ISO/FDIS 24616 Linguistic resources management – Multilingual information framework



Potential Areas for T9N Standardization

T9N/L10N standards

Process & Terminology

Metrics & measures

SAE J2450 

ATA Framework for Grading 

LISA QA Model 

ISO Metric standard? 

GILT standards

GMX-V

SRX

GMX-C

GMX-Q

xml:tm

T9N Services

ASTM 

CEN 

GB (China)

Quality

Vendor certification

Translator competencies

Bitext??

XLIFF

TWS 

ISO 12620:1999/2010

- Replaces ISO 12620:1999, which provided a hardcopy list of data categories for use in terminology work
- 12620:2010 provides a framework for the TC 37 Data Category Registry – <http://www.isocat.org>
- Data categories for a range of Thematic Domains, including morphosyntax, language resource metadata, and terminology, among others



ISO 16642:2003 — TMF

- Provides a meta-model for designing terminology management systems
- Defines levels for distributing data elements in TMSs
- References data categories for use in TMSs as per ISO 12620
- Provides for interoperability between Terminology Markup Languages



ISO 30042:2008 – TBX

- Provides for a family of XML-based terminology markup languages for use in exchanging data between TMSs compliant with ISO 16642 and ISO 12620
- Provides a wide range of flexibility with respect to:
 - Level of data category detail
 - Vocabulary style (native TBX xcs representation vs. W3C .xsd)



Role of Import & Export Formats

- Legacy data
- Sharing data
- Different types of outputs (glossaries, dictionaries, XML files)
- Interaction with other tools (lexicographical works, glossaries, MT lexicons, etc.)



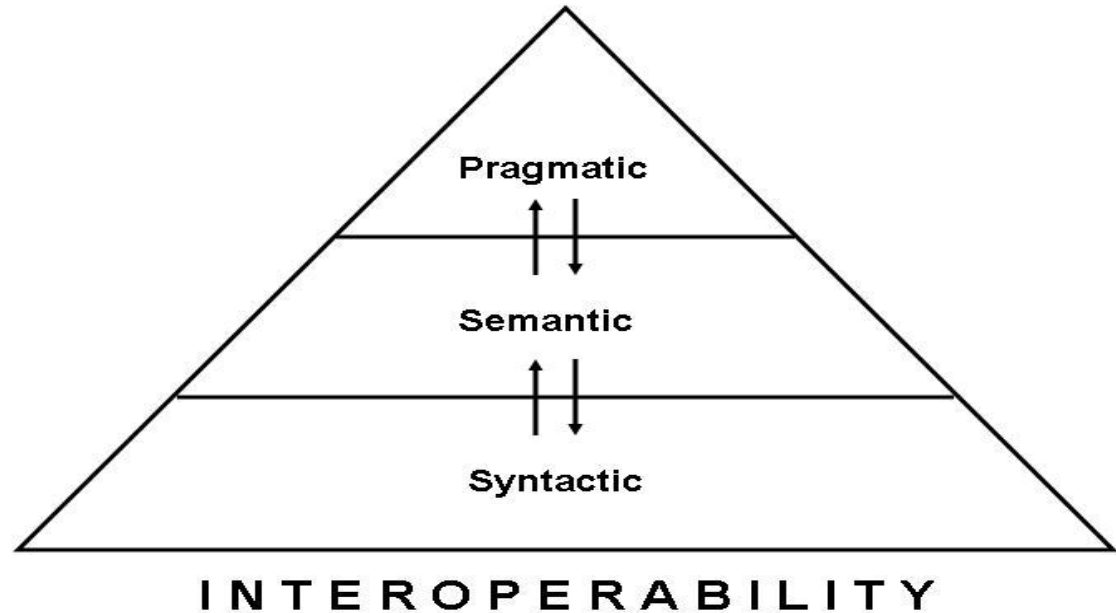
ISO 26162:2010(E)

Implementing TMSs

- *Systems to manage terminology, knowledge and content — Design, implementation and maintenance of Terminology Management Systems*
- Criteria for designing, implementing and maintaining terminology management systems (TMS)
- Rationale for using a TMS



Generic
interoperability
framework



Terminological
interoperability

