



Project acronym : EUROTERMBANK

Project title : Collection of Pan-European Terminology Resources through Cooperation of Terminology Institutions

Contract number : EDC - 22 267 - EUROTERMBANK

Starting date: 01/01/2005 **Ending date:** 31/12/2006

***D1.1 CURRENT STANDARDS AND
BEST PRACTICES ASSESSMENT REPORT***

Version: V 1.0
Date: 30/06/2005
Revised version: V 1.1
Date: 31/12/2005

Dissemination: *Public*

Contributors:

Sabiedrība ar ierobežotu atbildību Tilde
VIENIBAS GATVE 75A, 1004, RIGA,
(Tilde, Latvia)

The Institute for Information Management,
Fachhochschule Köln (University of Applied Sciences Cologne)
CLAUDIUS STRASSE 1, 50678, KOELN,
(IIM, Germany)

Center for Sprogteknologi, University of Copenhagen
NORREGADE 10, 1017, COPENHAGEN K,
(CST, Denmark)

Lietuvių kalbos institutas (The Institute of the Lithuanian Language)
P. VILEIŠIO G. 5, 10308, VILNIUS,
(LKI, Lithuania)

Terminology Commission of Latvian Academy of Science
AKADEMIJAS LAUKUMS 1, 1050, RIGA,
(LAS, Latvia)

MorphoLogic Számítástechnikai Korlátolt Felelősségű Társaság
ORBÁNHEGYI ÚT 5., 1126, BUDAPEST,
(MorphoLogic, Hungary)

Department of Estonian and Finno-Ugric Linguistics, Faculty of Philosophy,
Tartu Ülikool (University of Tartu)
ÜLIKOOI 18, 50090, TARTÜ,
(TÜ, Estonia)

OŚRODEK PRZETWARZANIA INFORMACJI (Information Processing Centre)
AL. NIEPODLEGŁOŚCI 188B, 00-608, WARSAW,
(OPI, Poland)

Authors:

Albina Aukšoriūtė, Institute of Lithuanian Language (LKI), Albinaa@lki.lt
Audronė Bielevičienė, Institute of Lithuanian Language (LKI), audrone@gironda.lt
Juris Borzovs, Latvian Academy of Sciences, juris.borzovs@riti.lv
Anna Braasch, Centre for Language Technology, anna@cst.dk
Anita Dravniece, Latvian Academy of Sciences, int@lza.lv
Nijole Dudlauskienė, Lithuanian Standards Board, nijole.dudlauskienė@lsd.lt
Lina Henriksen, Centre for Language Technology, lina@cst.dk
Heiki-Jaan Kaalep, Tartu University, hkaalep@psych.ut.ee
Danuta Kierzkowska, TEPIŚ Publishing House, dk@polishlaw.com.pl
Bente Maegaard, Centre for Language Technology, bente@cst.dk
Sussi Olsen, Centre for Language Technology, sussi@cst.dk

Claus Povlsen, Centre for Language Technology, claus@est.dk
Gábor Prószéky, Morphologic, proszeky@morphologic.hu
Marin Raguz, Institute for Information Management, marin.raguz@fh-koeln.de
Inke Raupach, Institute for Information Management, raupach@server.spr.fh-koeln.de
Klaus-Dirk Schmitz, Institute for Information Management, klaus.schmitz@fh-koeln.de
Krystyna Siwek, OPI, ksiwek@opi.org.pl
Raivis Skadins, Tilde, raivis.skadins@tilde.lv
Valentīna Skujiņa, Latvian Academy of Sciences, vaska@lza.lv
Robertas Stunzinas, Institute of Lithuanian Language (LKI), robertasst@lki.lt
Andrejs Vasiljevs, Tilde, andrejs@tilde.lv
Ainars Blaudums, Tilde, ainars.blaudums@tilde.lv
Uldis Priede, Tilde, uldis.priede@tilde.lv

Editors:
Centre for Language Technology, University of Copenhagen

D. 1.1 Contents

1	Introduction.....	9
2	Infrastructure of the new EU member countries within the terminology field.....	10
2.1	Estonia.....	10
2.2	Hungary.....	10
2.2.1	Terminology Council of the Hungarian Language.....	10
2.3	Latvia.....	11
2.4	Lithuania.....	12
2.5	Poland.....	13
3	Description of organisations working within the terminology field in the new EU member countries.....	14
3.1	Estonia.....	14
3.2	Hungary.....	15
3.3	Latvia.....	15
3.4	Lithuania.....	22
3.4.1	The Centre of Terminology.....	22
3.4.2	The State Commission of the Lithuanian Language.....	23
3.4.3	The Lithuanian Standards Board.....	24
3.5	Poland.....	26
3.5.1	The Polish Standardization Committee ‘PKN’.....	26
3.5.2	The Committee for European Integration ‘UKIE’.....	26
3.5.3	Council for the Polish Language.....	26
3.5.4	International Specialized Terminology Organization ‘ISTO’.....	27
3.5.5	Society of Economic, Legal and Court Translators TEPIS and PolTerm Bank.....	27
3.5.6	Universities.....	28
3.5.7	Published dictionaries.....	28
4	Terminology standardization.....	28
4.1	Major international standardization bodies.....	28
4.1.1	ISO standards in the terminology field.....	28
4.2	Major national standardization bodies.....	30
4.2.1	Estonia.....	30
4.2.2	Hungary.....	30
4.2.3	Latvia.....	31
4.2.4	Lithuania.....	32
4.2.5	Poland.....	34
4.2.6	Germany.....	36
5	General research principles in terminology work.....	37
5.1	Subject field classification.....	38
5.1.1	LENOCH universal classification system.....	38
5.1.2	Eurovoc.....	40
5.2	Identification and evaluation of specialized documentation.....	41
5.3	Identification of relevant concepts and relations.....	41
5.3.1	Conceptual analysis and concept system.....	42
5.3.2	Harmonization between languages.....	43
5.4	The terminological unit.....	43
5.4.1	Writing rules.....	43
6	Major existing terminology resources, bodies and networks.....	44
6.1	Termbanks, bodies and networks.....	44

6.1.1	IATE –Termbanks, bodies and networks.....	44
6.1.2	InfoTerm	46
6.1.3	TERMIUM.....	48
6.1.4	TSK.....	49
6.1.5	Polish terminology resources.....	50
6.1.5.1	The ‘PKN’ Polish Standardization Committee.....	50
6.1.5.2	UKIE terminology data base.....	51
6.1.5.3	PolTerm as a computerized collection of LSP-corpus-based terminology.....	51
6.1.6	Estonian terminology resources.....	52
6.1.7	Lithuanian language term base	53
6.1.8	Latvian terminology databases.....	53
6.1.9	SAP	54
6.1.10	Microsoft.....	54
6.2	Projects.....	55
6.2.1	INTERA.....	55
6.2.2	SALT.....	56
6.2.3	ENABLER.....	58
6.2.4	PHARE	58
6.3	Tools	59
6.3.1	Terminology management tools.....	59
6.3.2	Term extraction tools	60
6.3.2.1	Basic approaches to automatic term extraction.....	61
6.3.2.2	Evaluation of automatic term extraction.....	61
6.3.2.3	Some examples of term extraction tools.....	62
6.3.2.4	Conclusions on term extraction tools.....	62
6.3.3	Other tools.....	62
6.3.3.1	Search and replace tool: xPlace	62
6.3.3.2	Domain management in MemoQ / CATS, Kilgray's translation environment.....	63
7	Survey of best practices in terminology methodology and workflow	64
7.1	Overall workflow of terminology tasks	64
7.1.1	Estonian Legal Language Centre – Workflow.....	65
7.1.2	University of Tartu - Workflow	65
7.1.3	Latvia - Workflow.....	66
7.2	Classification systems.....	69
7.2.1	IATE – Classification systems.....	69
7.2.2	TERMIUM – Classification systems	69
7.2.3	Estonian Legal Language Centre – Classification systems	71
7.2.4	Lithuanian language term base - classification systems	71
7.2.5	Latvian language term base www.TermNet.lv – classification system	71
7.2.6	Hungary – an example of subject areas	72
7.3	Source identification	75
7.3.1	IATE – Source identification	75
7.3.2	TERMIUM – Source identification	76
7.3.3	TSK – Source identification.....	76
7.3.4	Estonian Legal Language Centre – Source identification.....	76
7.3.5	University of Tartu – Source identification	77
7.3.6	Lithuanian language term base - Source identification.....	77
7.3.7	Latvia – Source identification.....	77

7.4	Use of the internet as a resource	77
7.4.1	Types of search engines	78
7.4.1.1	TERMIUM – Types of search engines	78
7.4.1.2	TSK – Types of search engines.....	78
7.4.1.3	PolTerm – Types of search engines	78
7.4.2	Search methodologies	78
7.4.2.1	TERMIUM – Search methodologies	78
7.4.2.2	TSK – Search methodologies.....	79
7.4.2.3	PolTerm – Search methodologies	79
7.4.3	Evaluation of web site information.....	79
7.4.3.1	TERMIUM – Evaluation of web site information.....	80
7.4.3.2	TSK – Evaluation of web site information	80
7.4.3.3	PolTerm – Evaluation of web site information.....	80
7.4.3.4	Estonian Legal Language Centre – Evaluation of web site information	80
7.4.4	Creation of subject specific corpus	81
7.4.4.1	PolTerm – creation of subject specific corpus.....	81
7.5	Compilation of terms and conceptual analysis	81
7.5.1	Term extraction.....	82
7.5.1.1	TERMIUM – Term extraction	83
7.5.1.2	TSK – Term extraction	83
7.5.1.3	INTERA – Term extraction	83
7.5.1.4	PolTerm – Term extraction.....	85
7.5.1.5	Estonian Legal Language Centre – Term extraction	85
7.5.1.6	University of Tartu – Term extraction	85
7.5.1.7	Latvian partner – Term extraction	86
7.5.1.8	Morphologic – Term extraction in Hungary	86
7.5.2	Concept analysis	87
7.5.2.1	TSK – Concept analysis.....	87
7.5.2.2	PolTerm – Concept analysis	88
7.5.2.3	UKIE – Concept analysis.....	89
7.5.2.4	Estonian Legal Language Center - Concept analysis.....	89
7.5.3	Tools	90
7.5.3.1	TERMIUM - Term extraction tools	90
7.5.3.2	TSK - Term extraction tools	90
7.5.3.3	INTERA - Term extraction tools	90
7.5.3.4	PolTerm – Term extraction tools	90
7.6	Creation of terminological entries.....	91
7.6.1	Data categories in terminological entries.....	91
7.6.1.1	IATE – Data categories in terminological entries.....	91
7.6.1.2	INTERA - Data categories in terminological entries.....	92
7.6.1.3	TSK - Data categories in terminological entries.....	93
7.6.1.4	PolTerm – Data categories in terminological entries.....	94
7.6.1.5	UKIE – Data categories in terminological entries	95
7.6.1.6	Estonian Legal Language Centre - Data categories in terminological entries.....	97
7.6.1.7	University of Tartu - Data categories in terminological entries.....	98
7.6.2	Valid entries vs. concise, complementary and exhaustive entries	98
7.6.2.1	IATE - Valid entries vs. concise, complementary and exhaustive entries.....	99
7.6.2.2	PolTerm - Valid entries vs. concise, complementary and exhaustive entries.....	99

7.6.2.3	Estonian Legal Language Centre – Valid entries.....	100
7.6.2.4	University of Tartu – Valid entries	100
7.6.2.5	TC of LAS and LLI of LU – Valid entries vs. concise, complementary and exhaustive entries.....	100
7.6.3	Writing rules	101
7.6.3.1	IATE – Writing rules	101
7.6.3.2	TERMIUM – Writing Rules	103
7.6.3.3	TSK – Writing Rules	104
7.6.3.4	PolTerm – Writing Rules	104
7.6.3.5	University of Tartu – Writing Rules	104
7.7	Structures of terminological entries	104
7.7.1	IATE - Structures of terminological entries.....	105
7.7.2	INTERA– Structure of terminological entries	106
7.7.3	PolTerm – Structure of terminological entries.....	106
7.7.4	Estonian Legal Language Centre - Structure of terminological entries.....	107
7.7.5	University of Tartu - Structure of terminological entries	107
7.7.6	Lithuanian language term base – Structures of terminological entries.....	107
7.7.7	TermNet.lv Structure of terminological entries	108
7.8	Exchange formats.....	108
7.8.1	TBX (TermBase eXchange).....	108
7.8.2	TERMIUM – Exchange formats.....	110
7.8.3	TSK – Exchange formats	111
7.8.4	PolTerm – Exchange formats.....	111
7.8.5	Estonian Legal Language Centre – Exchange formats	111
7.8.6	University of Tartu – Exchange formats.....	111
7.8.7	TermNet.lv Exchange format.....	111
7.9	Management of terminological entries	112
7.9.1	Process requirements.....	113
7.9.1.1	TSK - Process requirements.....	113
7.9.1.2	PolTerm – Process requirements.....	113
7.9.1.3	Estonian Legal Language Centre – Process requirements	114
7.9.1.4	University of Tartu – Process requirements.....	114
7.9.2	Responsibilities	114
7.9.2.1	IATE - Responsibilities.....	114
7.9.2.2	TERMIUM - Responsibilities	115
7.9.2.3	TSK - Responsibilities	115
7.9.2.4	PolTerm - Responsibilities.....	115
7.9.2.5	Estonian Legal Language Centre – Responsibilities.....	115
7.9.2.6	University of Tartu – Responsibilities	115
7.9.3	Search and retrieval options.....	115
7.9.3.1	TERMIUM - Search and retrieval options.....	115
7.9.3.2	TSK - Search and retrieval options.....	116
7.9.3.3	PolTerm – Search and retrieval options.....	116
7.9.3.4	Lithuanian language term base - Search and retrieval options	116
7.10	Validation workflow	117
7.10.1	Process requirements.....	118
7.10.1.1	IATE - Process requirements	118
7.10.1.2	TSK - Process requirements.....	118

7.10.1.3	INTERA - Process requirements.....	118
7.10.1.4	PolTerm – Process requirements.....	119
7.10.1.5	Estonian Legal Language Centre – Process requirements.....	119
7.10.1.6	University of Tartu – Process requirements.....	119
7.10.2	Actors.....	119
7.10.2.1	IATE - Actors.....	119
7.10.2.2	TERMIUM - Actors.....	120
7.10.2.3	TSK - Actors.....	120
7.10.2.4	PolTerm: Actors.....	120
7.10.2.5	Estonian Legal Language Centre – Actors.....	120
7.10.2.6	University of Tartu – Actors.....	120
7.10.3	Feedback mechanisms.....	120
7.10.3.1	IATE - Feedback mechanisms.....	120
7.10.3.2	TSK - Feedback mechanisms.....	121
7.10.3.3	Estonian Legal Language Centre – Feedback mechanisms.....	121
7.10.3.4	University of Tartu – Feedback mechanisms.....	121
8	Conclusions.....	121
9	Appendices.....	122
9.1	ISO standards.....	122
9.1.1	ISO 704: Terminology work – Principles and methods.....	122
9.1.2	ISO 860:1996 Terminology work – Harmonization of concepts and terms.....	122
9.1.3	ISO 10241: International terminology standards – Preparation and layout.....	123
9.1.4	ISO 12200:1999 – Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) — Negotiated interchange.....	124
9.1.5	ISO 12620:1999 – Computer applications in terminology – Data categories.....	126
9.1.6	ISO 16642:2003 – Computer applications in terminology - Terminological markup framework (TMF).....	128
9.2	Principles in Latvian terminology work.....	131
9.2.1	Theoretical principles for terminology work in Latvia.....	131
9.2.1.1	Intralinguistic principles.....	131
9.2.1.2	Interlinguistic principles.....	134
9.2.1.3	Extralinguistic principles.....	134
9.2.1.4	Supplement: Ernest Drezen’s theoretical conclusions.....	135
9.2.2	Specific requirements for terms – Latvian experience.....	135
9.2.2.1	Main principles for term definitions.....	136
10	References.....	139
10.1	References from Estonia.....	139
10.2	References from Latvia.....	140
10.3	References from Poland.....	141
10.4	References from Hungary.....	146
10.5	References from Germany.....	147
10.6	References from Lithuania.....	148
10.7	References from Denmark.....	150

1 Introduction

In the world of today characterized by more and more globalization, the amount of cross national communication increases drastically.

In this context of multilingual interaction, possession and access to high quality terminological resources have become an extremely important issue. While it can be claimed that such terminological resources to a certain extent do exist for the so-called "old" EU Member States, this is by far not the case for the new EU member countries¹. In general, terminology resources in these countries need i.a. to be stored in a central pool, and to be made consistent and structurally compatible with respect to proven methodology. This report is to be understood as the first step towards reaching the overall goal: to improve the terminological infrastructure of the new Member States by making an inventory of international standards and best practices in terminological work and term management.

This report thus provides an overview of terminology standards, current terminology processes and best practices in selected terminology resources, bodies and projects. The purpose of the report is to form a basis for identification of a methodology for terminology processes in the new EU member countries and to ensure compatibility of terminological resources for data exchange and resource sharing.

Chapter 2 and 3 provide a description of the infrastructure of the new EU member countries within the terminology field.

Standardization is an important issue that takes place at international, national, regional and company level. Chapter 4 gives short introductions to major standardization bodies and to standards in the terminology field.

Chapter 5 is an introduction to terminology work and principles in general.

Chapter 6 provides introductory descriptions of the major existing terminology resources, bodies and networks that will form the basis for identification of best practices in terminology methodology and workflow.

In chapter 7 best practices regarding tasks, principles, tools etc. identified in the selected bodies, resources and networks are elaborated. In connection with each task or principle the particular resources applying this principle/task are stated and best practices recommended by these resources are described. The structure of this chapter is based on the tasks and principles identified during the project - mainly from the resources introduced in chapter 6, but also from standards and elsewhere to ensure that all relevant aspects of terminology work are covered, even if these aspects are not applied in the selected resources.

Finally, in chapter 8 key aspects are summarized.

¹ Which in this context are: Estonia, Hungary, Latvia, Lithuania and Poland

2 Infrastructure of the new EU member countries within the terminology field

2.1 Estonia

There is no central body responsible for the terminology development in Estonia. Terminology is developed by specialists (researchers, handbook and university textbook writers, technical translators etc) of the respective field. Typically, a field has a voluntary terminology committee that harmonizes the terminology of that field. The committee is not a legal entity; it may operate under a ministry, a non-profit organisation, or a university etc.

The following legal entities are responsible for some aspects of terminological work:

- Estonian Terminology Association
- Estonian Legal Language Centre
- Institute of the Estonian Language

2.2 Hungary

2.2.1 Terminology Council of the Hungarian Language

UNESCO has been active in terminology work during the last decades, establishing and unifying international norms, terms, classifications. Besides establishing international organisations, issuing decrees and guidelines, it also spurs the Member States to establish national/language-specific terminology coordination bodies.

In Hungary, terminology work was being done in several organizations, yet there was hardly any cooperation between actors in the field of terminology. During the last years, several organisations initiated a broad terminology dialogue on a national level.

Therefore, the Hungarian National Commission for UNESCO established the Terminology Council of the Hungarian Language (Magyar Nyelv Terminológiai Tanácsa, MATT) on the 12th of May, 2005 in Szombathely.

The tasks of the council are the following:

Basic research, applied research, education and training, PR activity, terminology and language policy, language planning, applying terminological data, methods, means in all areas of the information society, information collection and dissemination concerning terminology, supporting lexicography and discourse-specific lexicography;

Establishing and maintaining contact with the international circles of science;

Keeping in touch and consulting with government, business and public administration organisations.

To fulfil its tasks, the Council

drafts strategic recommendations for government, business and public administration organisations;

coordinates terminology work as needed;

enters into the network of **international terminology organisations**.

The Council members are terminology experts and representatives of institutions and organizations active in or making use of terminology.

The rules of procedure, the internal rules, the actual activities are all decided upon by an executive body elected from among the members.

As of today, the chairman of the commission is Prof. Dr. Vilmos Voigt, the vice-chairmen are Prof. Dr. János Pusztay and Ádám Kis.

Member institutions of the commission:

Állami Számvevőszék (State Audit Office)

Artisjus szerzői jogvédelmi iroda (Artisjus Copyright Office)

BME-UNESCO Információs és Trendkutató Központ (Budapest University of Technology-UNESCO Information and Trend Research Centre)

ELTE TTK UNESCO Multimédiapedagógiai és Információtechnológiai Központ (ELTE University of Budapest Faculty of Social Sciences-UNESCO Multimedia Education and IT Centre)

HUTerm

Igazságügyi Minisztérium (Hungarian Ministry of Justice)

Kilgray Kft.

KOPINT-DATORG Rt.

Központi Statisztikai Hivatal (Central Statistical Office)

Magyar Nemzeti Bank (Hungarian National Bank)

Magyar Szabványügyi Testület (Hungarian Body of Standards)

Magyar Tudományos Akadémia Nyelvtudományi Intézete (Institute of Linguistics, Hungarian Academy of Sciences)

MorphoLogic Kft.

Országos Széchényi Könyvtár (National Széchényi Library)

Országgyűlési Könyvtár (Library of the Parliament)

Pénzügyminisztérium (Hungarian Ministry of Finances)

Scriptum Informatikai Rt.

SZAK Publishers

2.3 Latvia

In Latvia, the main institution for the development of unified, coordinated and harmonized multibranch terminology for about 60 years is the Terminology Commission of the Latvian Academy of Sciences (TC of LAS) founded under the authority of LAS in 1946, but being founded earlier under the authority of Ministry of Education in 1919, and solving not only terminology but orthology issues as well. Until the middle of 1990ies subcommissions of the TC of LAS were the main compilers of subject-field terms and developers of subject-field term systems. The legal basis of the present activities of the TC of LAS is the State Language Law (in force since 1999) and the Regulations adopted by the Cabinet of Ministers. Decisions taken by the TC of LAS have the status of normative documents, and terms approved by TC of LAS are official.

TC of LAS has close links of different characters (institutional, juridical, subordinative, coordinative, cooperational, consulting, etc.) with a number of other commissions, state institutions, ministries, agencies, universities and research institutions (see the scheme “Institutional connections of the Terminology Commission of LAS” below).

TC of LAS has institutional contacts with the following institutions which are involved in implementation of the Latvian state language politics and in relevant terminology issues as well:

- State Language Commission (SLC)
- State Language Agency (SLA)
- State Language Centre
- Commission of Latvian Language Experts (CLLE)
- Latvian Language Institute (LLI)
- Terminology Department of LLI
- Translation and Terminology Centre (TTC)
- Latvian Standardization Organization “Latvijas Standarts”
- Tilde
- Other

See a more detailed description in Chapter 3, point 3.3.

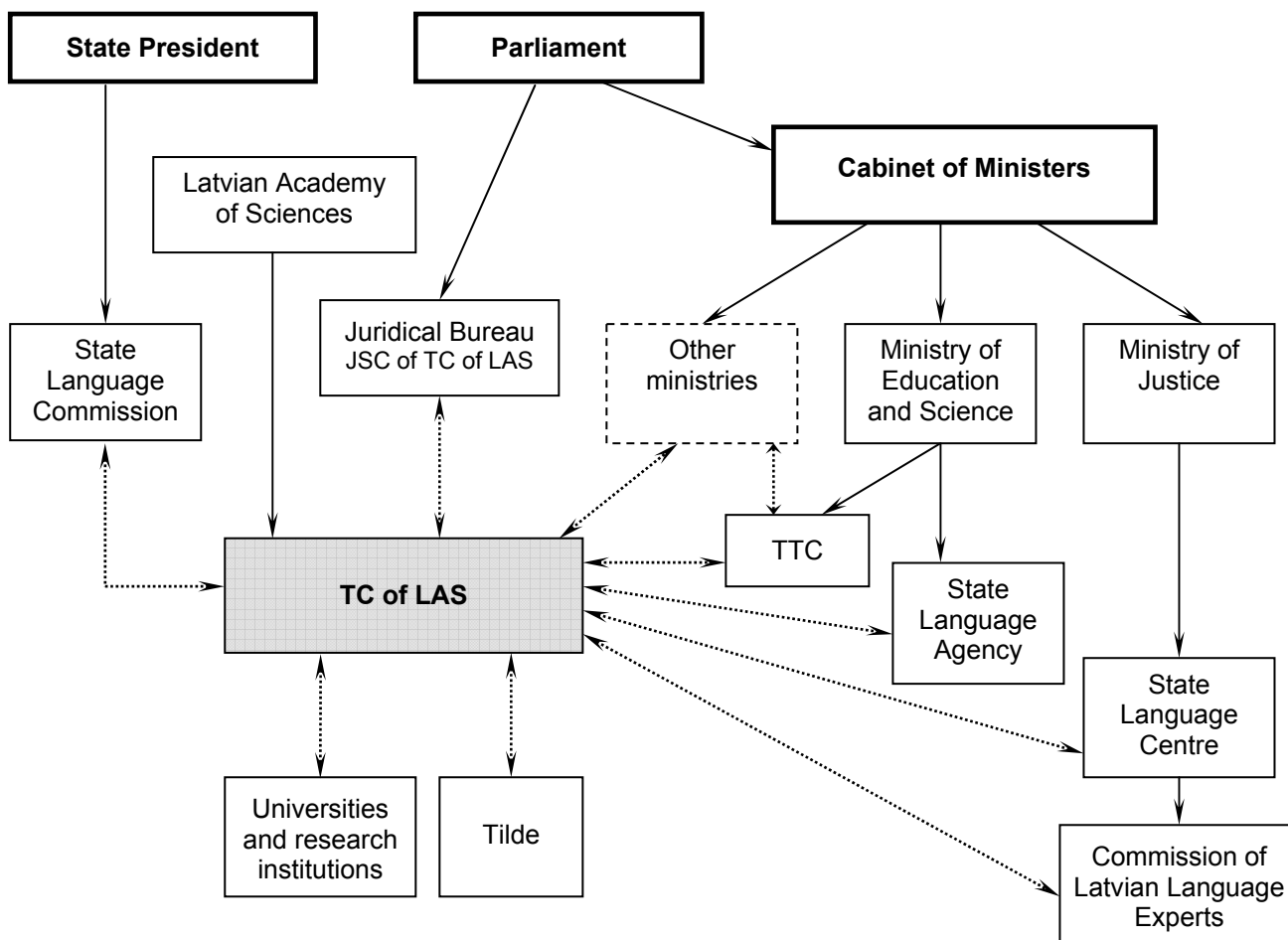


Figure 2.1. Institutional connections of the Terminology Commission of LAS

2.4 Lithuania

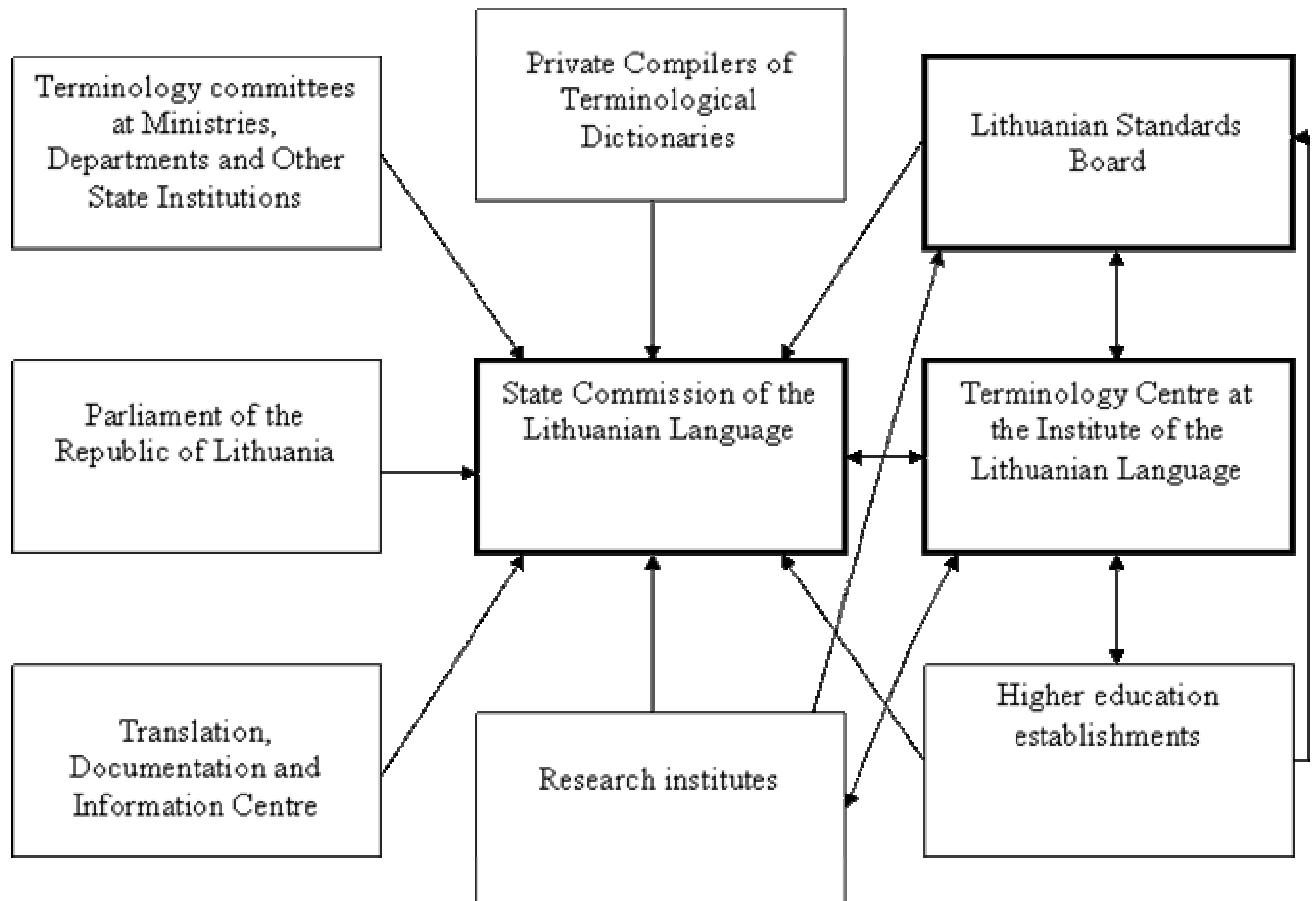


Figure 2.2. *Institutional connections of the Terminology Centre of LKI*

2.5 Poland

Terminology work in Poland is being performed in various forms by

1. PKN - a national standardization body
2. UKIE - a governmental office dealing with translation of UE legal instruments
3. RJP - a problem committee of the Polish Academy of Sciences
4. ISTO - an international organization registered in Poland
5. TEPIS - a specialized translators' professional association
6. UNIVERSITIES - units for scientific research
7. PUBLISHERS - editing dictionaries, glossaries and thesauri.

The above listed institutions are dealing either with methodology of terminology, creating terminology collections in particular fields or maintaining terminology resources.

1. The most comprehensive terminology work is being carried on by the **Polish Standardization Committee ‘PKN’** (Polski Komitet Normalizacyjny) in all fields of terminology work: methodology, creating and maintaining terminology resources (see a more detailed description in 4.2.5. below).
2. **The Office of the European Integration Committee ‘UKIE’** (Urząd Komitetu Integracji Europejskiej) is in possession of a bank of legal terms and phrases used in Polish translation of EU legal instruments passed in English/French/German (see a more detailed description in 3.5.2 below.)
3. Another institution dealing mainly with the methodology of creating specialized terminology is the **Council for the Polish Language** established in 1999 as a problem committee of the Polish Academy of Sciences (see a more detailed description in 3.5.3 below.)
4. **The International Specialized Terminology Organization ‘ISTO’**, founded in 2004, is a holder of a certain number of specialized terminology dictionaries and some issues of the journal **‘Neoterm’** dealing mainly with the methodology of specialized terminology (see a more detailed description in 3.5.4).
5. Under the auspices of the **Polish Society of Economic, Legal and Court Translators ‘TEPIS’** a specialized translation journal **‘Lingua Legis’** is edited as dealing with legal terminology and a term bank named **‘PolTerm’** is maintained by the TEPIS Publishing House (see a more detailed description in 3.5.5).
6. **Universities** are carrying on terminology work in their units for scientific research, mainly methodology-oriented and descriptive, whereas their term collections, if exist, require to be supplemented and converted into electronic version (see a more detailed description in 3.5.6).
7. **Published dictionaries, glossaries and thesauri** are edited in paper version, but sometimes are also accessible on-line on the Internet. For their lists and descriptions see 3.5.7 below.

3 Description of organisations working within the terminology field in the new EU member countries

3.1 Estonia

The Estonian Terminology Association (www.eter.ee), ETER is a non-profit organisation. The mission of ETER is to coordinate the work of LSP and terminology in Estonia, to develop the collections of terminology, to organize projects of terminology at the Estonian and European level.

Estonian Legal Language Centre (www.legaltext.ee)

ELLC is a public organisation, operating under the governance of the Ministry of Justice. The mission of the ELLC is to meet the legislative translation and terminology development needs of the Estonian Government.

The ELLC fulfils its mission by the following main activities:

- translation of Estonian legislation into English;
- translation of EC legislation into Estonian;
- creation and administration of a full-text database of legal translations and of a terminology database, and making such databases available to the public.

Translation priorities are set in close co-operation with the ministries and approved by the Council of Senior Civil Servants and the Minister of Justice.

Institute of the Estonian Language (www.eki.ee)

EKI is a public research and development organisation, operating under the governance of the Ministry of Education and Science. The mission of EKI is to research Estonian (modern Estonian, dialects, history of language, LSP etc), including terminology in Estonian. EKI is also responsible for language planning.

3.2 Hungary

In Hungary, the first major government effort in the terminology field was started by the Ministry of Justice in 1997. The aim was the creation of the official terminology database of the European Union. The Ministry established a Translation Coordination Unit which worked under the supervision of the High Department for European Community Law. The project consisted of translating 100,000 Official Journal pages, and during this work, the TCU created a terminology database of 23,000 entries.

The TCU did not employ in-house translators, translation was performed by translation companies successful at public procurement tenders. The unit consisted of 4-9 terminologists and lawyer-linguists. The legally approved texts were handed over to lawyers working at the Council and the Commission for final review, and then to OPOCE to get published in the Official Journal.

Terminologists collected the terms in a dedicated database called Termin. Entries appearing as approved have to be used by translators and revisers. The database has an internet interface available to everyone. It was also published recently, by three publishers: MorphoLogic&SZAK Kiadó, Akadémiai Kiadó and HVG ORAC.

The terminological database is based on English-Hungarian term pairs, French and German language versions aid in orientation and additional security.

The translation of the acquis ended in 2004, at the time of the EU accession of Hungary. The Ministry decided to upkeep its TCU, but the amount of work they perform dropped significantly and the TCU gradually decreases in size.

With the establishment of the Terminology Council of the Hungarian Language (MATT) in May 2005, any kind of work on the field of terminology was brought to a national, standardized level, integrating all previous efforts.

3.3 Latvia

Terminology Commission of the Latvian Academy of Sciences (TC of LAS)

Since 1946, the only institution for terminology work in Latvia for over a 50 year period was the Terminology Commission (TC) of the Latvian Academy of Sciences (LAS). It was founded by Government. From the beginning TC of LAS has subcommissions in 10 subject fields: law, history, economics, agriculture, philology, geography, geology, biology, medicine and technical sciences. The main tasks of subcommissions were to compile appropriate subject field terms and to develop subject field term systems.

At present the basis of the TC of LAS activities is the State Language Law accepted by Latvian Parliament in 1999 December 9 and coming into force since 2000 September 1, and the Regulations adopted by the Cabinet of Ministers, namely: Regulation of the Cabinet of Ministers of the Republic of Latvia No 405 "The Statute of the Terminology Commission of the Latvian Academy of Sciences" adopted in 2000 November 28; and Regulation of the Cabinet of Ministers of the Republic of Latvia No 36 "Amendment to the Regulation of the Cabinet of Ministers of the Republic of Latvia No 405 "The Statute of the Terminology Commission of the Latvian Academy of Sciences"" adopted in 2004 January 20.

According to the legislative acts it is stated that decisions taken and terms approved by the TC of LAS have the status of normative documents.

The main functions of the TC of LAS are:

- development of a uniform and scientifically grounded national terminology system harmonized with the international concept systems and theoretical principles of terminology work;
- assessment of terms submitted for approval (see scheme No 1 below);
- coordination of terminology work in Latvia;
- rendering of methodical and methodological help regarding subject-field terminology issues;
- approval of terms and definitions for usage in standards;
- solution of disputes between subject field specialists, etc.

The terminology work of TC of LAS is term-oriented and aimed at the development of a multi-branched term system according to the recommendations for terminology work given by the leading terminology experts. Presently the responsibility of the TC includes both the regular work as it has been carried out in the previous decades and solving of urgent tasks including those arising from Latvia's membership in the EU.

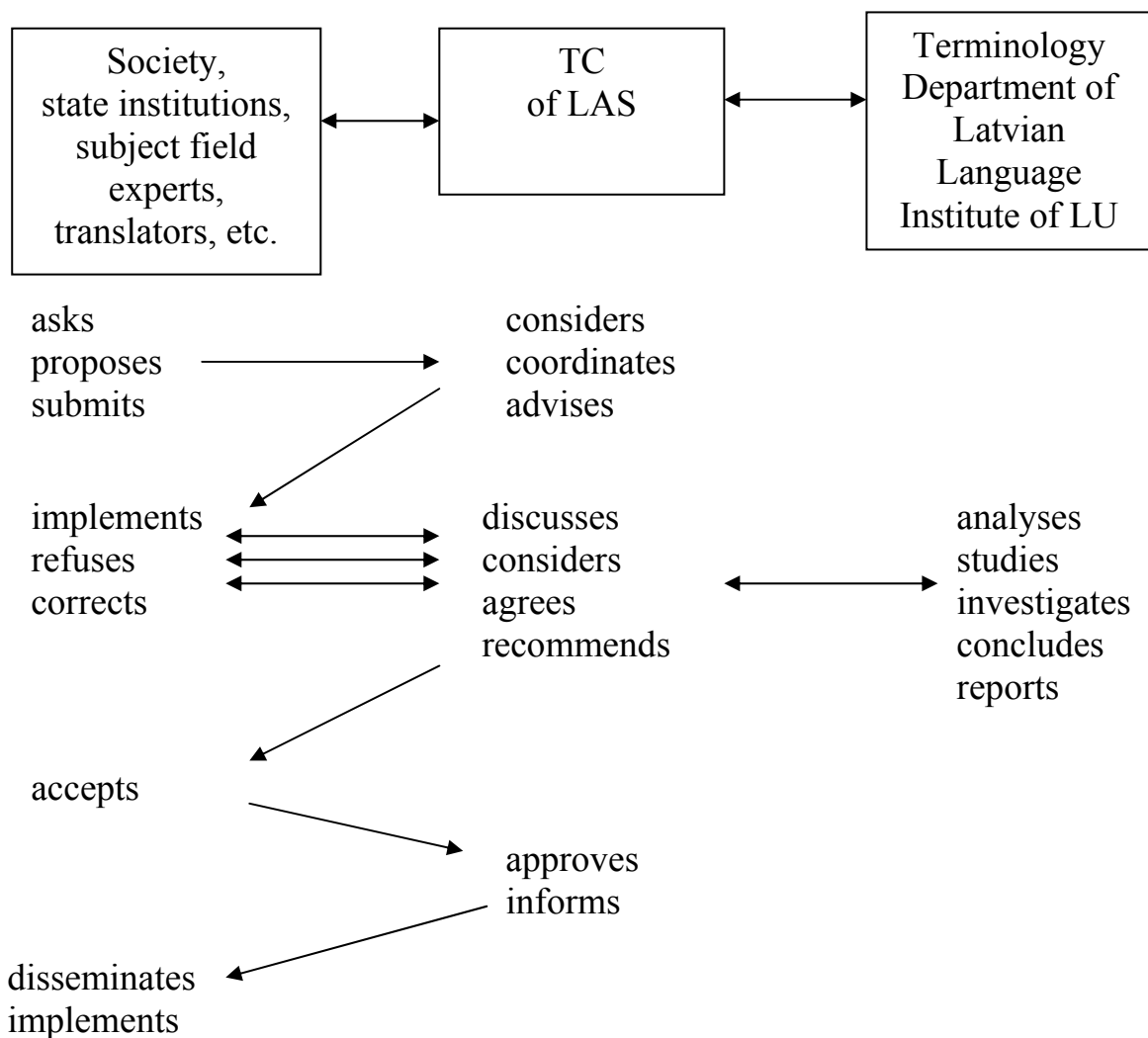


Figure 3.1. *Workflow of terminology tasks involving TC of LAS and LLI of LU*

N. B.! In all stages consultations with experts are practiced.

In the course of developing any subject-field terminology attention is paid to coordination of terminology at all possible levels: inside a branch, among branches, and up till the level of LSP and LGP, in general. Besides, as noted before, inter-lingual aspects of term harmonization on the conceptual (meaning) level are taken into account, too (see the scheme No 2).

M
E
A
N
I
N
G

and

F
O
R
M

L
E
V
E
L

M
E
A
N
I
N
G

L
E
V
E
L

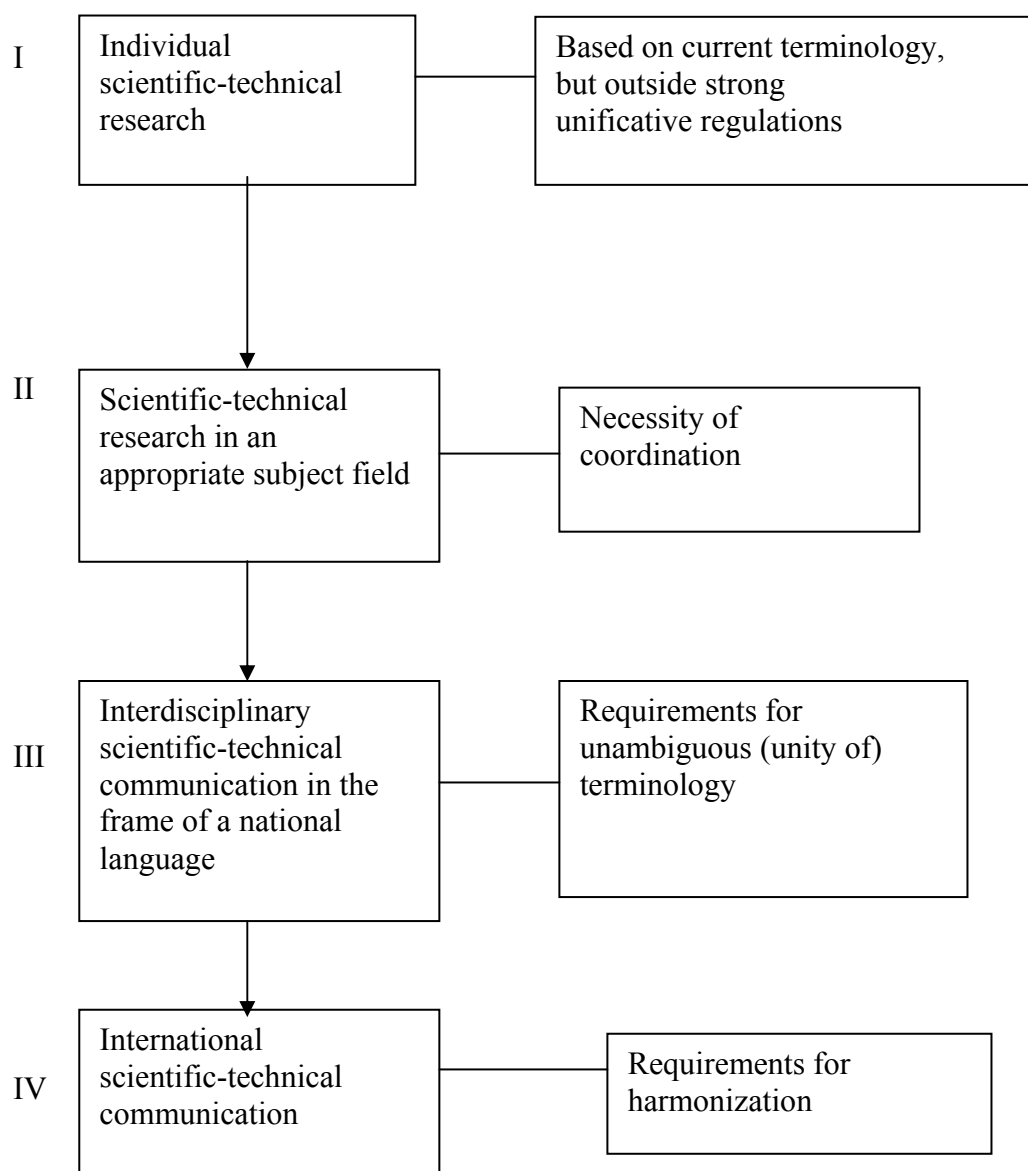


Figure 3.2. The course of the terminology development from the aspect of unification

At present the TC of LAS consists of 26 subject-field terminology subcommissions, which are working according to the general principles and methodology for terminology work accepted by TC of LAS and at the same time considering the peculiarity of each subject-field.

Subcommissions are founded on such subject-fields:

- Subcommission on Agricultural Equipment Terminology;
- Subcommission on Aviation Terminology;

- Subcommittee on Biology Terminology;
- Subcommittee on Chemistry Science and Technology Terminology;
- Subcommittee on Demography Terminology;
- Subcommittee on Economics Terminology;
- Subcommittee on Energetics Terminology;
- Subcommittee on Forestry Science Terminology;
- Subcommittee on Geology Terminology;
- Subcommittee on Hydro-engineering Terminology;
- Subcommittee on Heat, Gas and Water Technology Terminology;
- Subcommittee on History Terminology;
- Subcommittee on Information Technology and Telecommunications Terminology;
- Subcommittee on Legislation Terminology;
- Subcommittee on Library, Bibliography and Information Sciences Terminology;
- Subcommittee on Mathematics Terminology;
- Subcommittee on Medicine Terminology;
- Subcommittee on Mechanics and Engineering Industry Terminology;
- Subcommittee on Military Terminology;
- Subcommittee on Museology Terminology;
- Subcommittee on Philosophy Terminology;
- Subcommittee on Railway Transport Terminology;
- Subcommittee on Sport Terminology;
- Subcommittee on Textile Industry Terminology;
- Subcommittee on Tourism Terminology;
- Subcommittee on Vehicle Technology Terminology.

If necessary, sub-commissions are subdivided into sections, for instance, in the framework of biology separate sections for development of botany and zoology terminology are formed.

In accordance with the above the multi-branched term system is developed for term database of the TC of LAS.

The State Language Commission (SLC)

The State Language Commission (SLC) established under the auspices of the President of Latvia is the main institution determining the state language policy in Latvia. SLC is a collegiate institution without status of a legal entity (see the Regulation No. 186 of the Cabinet of Ministers from 14 May 2002 “The Statute of the State Language Commission”). The Commission was established in order to study the situation of the state language in Latvia and to develop suggestions how to for strengthen the position, improve and ensure sustainable development of the Latvian language.

SLC has developed the conception of a state language policy, as well as the state programme for state language development and research, which determine its implementation strategy.

The issues of terminology do not fall directly within the competence of SLC. Still these are included in the general state language development policy programme.

The State Language Agency (SLA)

The State Language Agency (SLA) is a government body, established under the supervision of Ministry of Education and Science, Republic of Latvia. This process was initiated by SLC. Pursuant

to the Regulation No 667 of the Cabinet of Ministers of 25 November 2003 “the Statute of the State language Agency” the tasks of SLA are as follows:

- to analyse the situation of state language and the dynamics of socio-linguistic processes;
- to establish a consulting service network and to provide consulting on issues of state language;
- to promote the learning and use of the state language;
- to organize the development and publishing of methodological and information material;
- to promote the development of the state language in the field of IT;
- to support development and publishing of updated state language normative sources;
- to ensure the development of terminology of the state language by creating term databases.

The State Language Centre (SLC)

The State Language Centre (SLC) is a direct management institution under the control of Minister of Justice. The purpose of the Centre work is to implement the state policy, performing supervision the observance of regulative acts and control in the field of the state language use, as well as protecting the rights and interests of the state language users (see Regulation No. 202 of the Cabinet of Ministers from 22 March 2005 “The Statute of the State Language Centre”). Among other tasks SLC aims at improvement of the state language normative acts, development and publishing of informational materials.

The Statute of the Terminology Commission (TC) of the Latvian Academy of Sciences provides for co-operation between TC and SLC in regulation of strategic issues of terminology policy, as well as to turn to SLC in order to ensure the observance of TC’s decisions.

The Latvian Language Institute (LLI)

The Latvian Language Institute (LLI) was founded in 1992 on the basis of the Institute of the Latvian Language and Literature, and from the beginning it was founded under the authority of the Latvian Academy of Sciences. Since the 1994 it is working under the authority of the Latvian University (LU). At present, the LLI of LU is a scientific institute with main research directions:

- lexicography theory and practice;
- language history and development of Latvian grammar system;
- dialectology;
- terminology theory and practice; etc.

The Terminology Department of LLI

Terminology Department is a structural unit in the Latvian Language Institute with main task to investigate terminological theoretical and practical issues.

The Translation and Terminology Centre (TTC)

The Translation and Terminology Centre (TTC) – the second relevant institution on terminology issues in Latvia – was founded in 1997 by the Ministry of Foreign Affairs of Latvian Republic with main function of translation of EU legislative acts, in other words – term-saturated LSP texts. For a period about 7 years there was a close collaboration between these both institutions – TC of LAS and TTC.

Since November 2004 the Translation and Terminology Centre (TTC) is reorganized into a state agency under supervision of the State Chancery, since June 2005 – under the Ministry of Education and Science. Pursuant its Statute (see Regulation No 826 of the Cabinet of Ministers from 5

October 2005) the aim of TTC is to provide the translations of the documents of state and international organizations for the purposes of state administration and the society, as well as to submit proposals for development and standardization of terminology. The functions of TTC also include translation of Latvian legislation into languages of EU Member States along with the development of translation methodology. During the translation of EU legislative acts TTC has developed (in many cases context-oriented) term database.

The Terminology Commission and TTC continue to co-operate in the field of terminology development. TTC staff members take part in the work of some TC sub-commissions.

Latvian Standardization Organization „Latvijas Standarts” (LVS)

Latvian Standardization Organization “Latvijas Standarts” (LVS), being founded in 1999, is the national standardization body (Ltd) of Latvia and its main tasks are to provide juridical and natural persons with information on standardization, developing of the national standards, to adapt international standards and to maintain the register of adapted Latvian standards (home pages of LVS).

See a list of LVS standards harmonized with terminology approved by TC of LAS in 4.2.3

Tilde

Tilde, being established in 1991, is a leading Baltic IT company specializing in language technologies, multilingual and Internet software, localization.

In language technologies, the aim of Tilde is to provide language technologies for the languages of the Baltic countries that would be equivalent to the support for the major languages of the world.

As a member of the Information Technology and Telecommunications Terminology Subcommittee of the TC of LAS, Tilde actively participates in terminology development process.

Tilde provides quality localization services for all three languages of the Baltic countries. Such companies as IBM, Microsoft, Hewlett Packard, Nokia and others use Tilde’s localization services.

The company has established good cooperation with research institutes, educational and state organizations, such as the Academy of Science and the Institute of Mathematics and Computer Science.

Other institutions involved in the terminology process are:

- The Central Statistics Department, dealing with the terminology of different branches of economy;
- The Service for Plant Protection, managing the names of plants and the means of their protection;
- The State Fishery Department, managing the names of fish and fishing equipment;
- The State Agency of Medical Substances, that manages the names of medicines and pharmacological substances and preparations;
- The Bank of Latvia that has organized a special terminology development group in the field of international accountancy standards.

3.4 Lithuania

There are three institutions which work in the field of terminology in Lithuania – The Centre of Terminology at the Institute of Lithuanian Language, The State Commission of the Lithuanian Language and The Department of Standardisation.

3.4.1 The Centre of Terminology

Terminological work at the Institute of Lithuanian Language was started in 1941. A separate Department of Terminology was founded in 1991 and it was reorganized into the Centre of Terminology in 2003. Dr Albina Auksoriūtė is Head of the Centre and the staff of twelve includes four people with doctorates.

Principal activities:

- Research into the history of Lithuanian terminology and theoretical issues of present-day terminology.
- Establishment of principles and norms of Lithuanian terminology and terminography.
- Practical work of normalisation of terminology in various fields, creation, evaluation and normalisation of new terms.
- Training of specialists in the field of terminology, consultation to compilers of terminological dictionaries.

Theoretical research into Lithuanian terminology

Articles by employees of the Centre are usually published in *Terminologija*, *Acta Linguistica Lithuanica*, *Kalbos kultūra*, *Lituanistica* and in foreign scientific publications.

The Centre publishes a terminological magazine called *Terminologija* (ISSN 1392-267X), which has been published annually by the Institute of Lithuanian Language since 1994. The publication deals with the theory and history of terminology and terminography, terminological analysis, scientific regulation, standardisation, its link with the theory of standardisation of the lexis of common language, terminological and lexicographical matters of scientific language, the correlation between scientific and folk terminology. It presents articles, comments, reviews, notices, information and an annotated bibliography of terminological dictionaries.

Terminologists from the Institute participated in the preparation of the 3rd and 4th edition of the *Dictionary of Modern Lithuanian* (1993, 2000) and its computer version (2002). In 1980 Dr. Stasys Keinys published *Terminologijos abėcėlė (ABC of Terminology)* and in 1992 a group of authors issued *Terminologijos taisymai (Terminology Corrections)*.

The latest books prepared by the terminologists from the Institute are a monograph by Dr. Kazimieras Gaivenis *Lietuvių terminologija: teorijos ir tvarkybos metmenys (Lithuanian terminology: an outline on theory and practice)* (2002) and a selection of scientific articles *Terminologijos istorijos ir dabarties problemos (Problems of the past and the present of terminology)* (2004).

The Centre organizes terminological conferences and seminars:

1. A scientific seminar about synonymy in terminology (2002).
2. An international conference *Terminology at the beginning of the third millennium* (2003).

3. An international seminar *Problems and tasks of Estonian, Latvian and Lithuanian terminology in the European Union* (2004).

Practical work of the Centre of Terminology

For 50 years nearly all the most important dictionaries of terms were prepared with the participation of terminologists from the Institute. Most employees of the Centre participate in various editorial boards of terminological dictionaries, where they are included as consultants. Until 1996 the Department of terminology was responsible for approving terminological dictionaries and standards, later the State Commission of the Lithuanian Language took over those duties. Now the Centre of Terminology reviews terminological dictionaries and standards and carries out expertise on behalf of the Language Commission. Employees of the Centre of Terminology assist the Language Commission as experts and participate in the work of the sub-committee of terminology in this commission. Centre of Terminology also advises various institutions by email, telephone or post.

3.4.2 The State Commission of the Lithuanian Language

The State Commission of the Lithuanian Language was established in 1990 by the Act of Parliament. It collectively takes decisions about the language policy, normalization and standardisation of the Lithuanian language and implementation of the State language law, which was approved by the Parliament in 1995.

The Commission itself consists of 17 members from scientific and educational institutions. There are 7 sub-committees of

- orthography, punctuation and grammar;
- pronunciation and accentuation;
- onomastics;
- dictionary;
- language technologies;
- evaluation of textbooks;
- terminology.

The Sub-committee of Terminology

The head of the sub committee of terminology is a professor Albertas Rosinas from Vilnius University and four employees of the Centre of Terminology are its members. Practically all research staff of the Centre work as experts for the Commission and review dictionaries and standards of terms on its behalf.

In 1997 the Commission approved the regulations for the approval of terminological dictionaries. Dictionaries, published in a large number of copies and not for the internal usage of a particular institution ought to seek the approval of the Commission, though in practice this requirement is not being enforced. Most significant dictionaries of terms published recently got this approval. Compilers of dictionaries seek it, because it is a “quality certificate”. Seven out of fifteen terminological dictionaries published in Lithuania 2001, had the approval of the Commission and eight out of nineteen dictionaries in 2002.

For the preparation of a terminological dictionary there is a requirement to include a terminologist in the group of compilers, although this requirement is not always observed. When compilers seek approval, the Commission appoints two experts perform the expertise – a specialist of a particular

field and a terminologist. After the dictionary is discussed in the sub-committee of terminology, it receives the approval.

3.4.3 The Lithuanian Standards Board

National standards body of Lithuania (Lithuanian Standards Board) was established in 1990. According the Law on standardization of Lithuania the main responsibility of national standards body – to establish and to improve national standardization system of Lithuania and to take part as members in the activity of the European standardization organizations (ESO – CEN, CENELEC and ETSI) and as correspondent member - in the international organizations for standardization (ISO and IEC). Responsible for preparation of Lithuanian standards are technical committees of Lithuanian Standards Board (LST/TC), established as mirror technical committees of ESO and ISO or IEC. In field there technical committees on European level do not exist, LST/TC are active correspondent members of ISO/TC.

The number of LST/TC in 2005 is 71, the list of LST/TC: <http://www.lsd.lt>.

In 1996 LST/TC 37 *Terminology* was established as a mirror technical committee of ISO/TC 37 *Terminology and other language resources*. In 1999 LST/TC 37 *Terminology* became correspondent member of ISO/TC 37 *Terminology and other languages resources*.

There are representatives of 35 members in the LST/TC 37:

- Vilnius University, Vilnius Gediminas' Technical University, Kaunas Technology University, others universities, science and education institutions;
- Members of other LST/TC's;
- Governmental institutions;
- State enterprises;
- Other Organizations.

In 2000 five sub-committees (SC) of LST/TC 37 were established.

LST/TC 37 SC 1 as correspondent member of ISO/TC 37 *Terminology and other languages resources* is responsible for adoption of international standards prepared by ISO/TC 37 as Lithuanian standards.

Other responsibilities of subcommittees of LST/TC 37 is expertise of Lithuanian terms created by LST/TC in different fields of activity according International Classification of Standards (ICS) in the process of preparation of Lithuanian versions of European or international standards and adoption as Lithuanian standards or creation of Lithuanian terminology standards.

- SC 1 Principles and methods of terminology - expertise of terms according ICS 01 used in Lithuanian standards, prepared by LST/TC 37, 47, 65, 69;
-
- SC 2 Electrical engineering and electronics - expertise of terms according ICS 07, 13, 17, 19, 21, 23, 27, 29, 31, 33, 35, 37, 39, 97, used in Lithuanian standards, prepared by LST/TC 1, 4, 5, 8, 18, 22, 24, 35, 36, 45, 55
- SC 3 Building constructions - expertise of terms according ICS 03, 13, 19, 21, 23, 45, 47, 73, 77, 79, 91, 93, 95, 97 used in Lithuanian standards, prepared by LST/TC 12, 14, 17, 19, 22, 23, 25, 26, 28, 30, 35, 38, 39, 46, 50, 59, 60, 61;
-

- SC 4 Mechanics, road vehicle and railway engineering - expertise of terms according ICS 03, 21, 25, 43, 45, 47, 49, 53, 55, used in Lithuanian standards, prepared by LST/TC 12, 14, 17, 19, 22, 23, 25, 26, 28, 30, 35, 38, 39, 41, 48, 49, 52, 53, 56, 63, 66;
-
- SC 5 Health, environment and agriculture - expertise of terms according ICS 01, 11, 13, 19, 65, 67, 71, 71, 75, 83, 85, 87, used in Lithuanian standards, prepared by LST/TC 2, 3, 6, 7, 10, 11, 15, 16, 27, 29, 32, 33, 36, 51, 54, 62, 64, 68, 70, 71
-

After expertise of LST/TC 37 *Terminology* all Lithuanian terminology standards are offered to the State Lithuanian Language Commission for approval. Terminology used in Lithuanian terminology standards has the topmost level of reliability of terminology used in technical fields related with standardization.

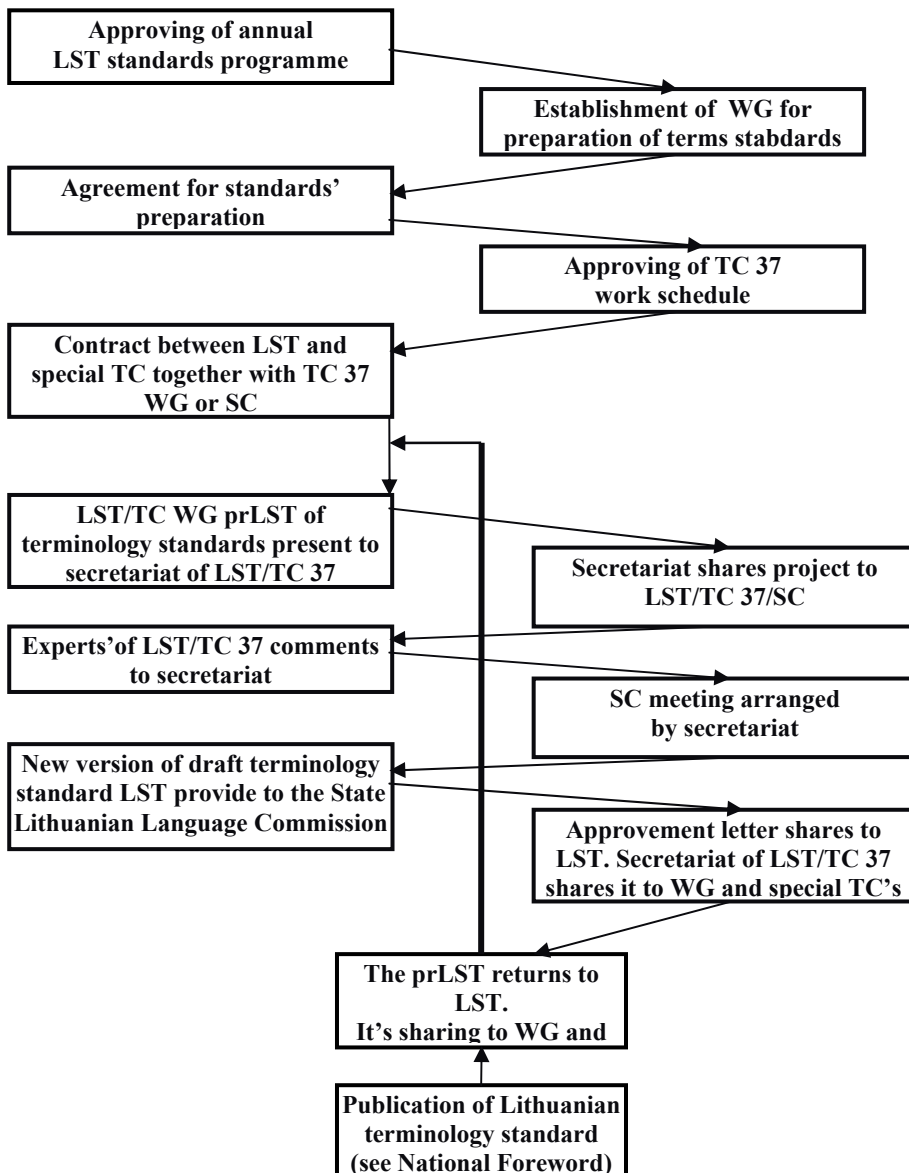


Figure 3.3. Flowchart of expertise of terms used in Lithuanian standards

Excerptions from the LST 0-6:

“The Lithuanian terms have been approved by the State Lithuanian Language Commission”.

“The Lithuanian terms have been approved by LST/TC 37 *Terminology*”.

The Lithuanian Standards Board has a database LST-TERM of standardized terms and definitions used in the Lithuanian terminology standards and in the clauses *Terms and definitions* of non-terminology standards, which contains about 19 000 entries from 531 Lithuanian standards including as Lithuanian standards adopted European deliverables and ISO publications. LST-TERM covers not only Lithuanian terms, but also German, English, and French equivalents. In such cases when other equivalents from official sources exist, in Lithuanian terminology standards and LST-TERM Latin, Russian or other languages equivalents are included.

LST-TERM is accessible to registered users on the internet.

3.5 Poland

3.5.1 The Polish Standardization Committee ‘PKN’

A description of activities carried on by the Polish Standardization Committee ‘PKN’ is presented under items 4.2.5.1 and in 6.5.1.

3.5.2 The Committee for European Integration ‘UKIE’

The Committee for European Integration was established by virtue of the Act of 8 August 1996, is a supreme governmental administration body competent for programming and coordination of policy relating to Poland’s integration with the European Union, programming and coordination of Poland’s actions adjusting Poland to European standards, as well as for coordination of state administration activities in the field of foreign assistance obtained.

Within the framework of its activities, the Department for Translation of the Committee Office was a unit responsible for coordinating translation of *acquis communautaire* into Polish. The translation process, involving consultations with other ministries and central administration bodies, was a highly complex task and required efforts both on the part of professional translators and the public administration. Poland, just as all other countries that became Member States of the European Union in 2004, was obliged to translate into Polish the entire body of *acquis communautaire*, estimated to contain about 80 thousand pages of the Official Journal of European Communities.

A by-product of translation activities was the creation of the UKIE bank of translation-oriented legal terminology, described under 6.5.2.

3.5.3 Council for the Polish Language

The Council for the Polish Language was established by virtue of the Act of 7 October 1999 on the Polish Language and acting as a problem committee of the Polish Academy of Sciences; it is the institution providing opinions and advice in the matters relating to the usage of the Polish language. The scope of the Council’s activities include the permanent care of the appropriate development of the modern Polish language used in public life, legal acts, press, radio, television and public

administration, including special usage in various fields of science and technology. The Council's opinions and suggestions are published in their periodicals and accessible on-line on the internet. Either the Act establishing the Council, or its plans, do not provide for the future foundation of any terminology collections, but certainly their terminology work will be very important and useful for specialists coining new Polish terminology in all fields of science and technology.

So far, seven commissions have been established within the Committee: Commission for the Language of Law, Commission for the Language of Medicine and Natural Sciences, Commission for Computer Science, Commission for the Religious Language, Commission for Orthography and Onomastics, Commission for the Contemporary Language, and the Commission for Language Teaching.

3.5.4 International Specialized Terminology Organization 'ISTO'

The International Specialized Terminology Organization 'ISTO' was founded in 2004. The ISTO Inauguration Congress was held in Warsaw in July 2004. The ISTO was established as a result of transformation of the former International Organization of Unification of Terminological Neologisms (IOUNT) affiliated to the United Nations as a non-governmental organization since 1987. All the publications edited by the IOUNT have been transferred to the ISTO, presently being in possession of all publications, including dictionaries of specialized terminology.

Members of IOUNT were expert philologists, linguists and specialists in various fields of science involved in the theory and methodology of terminology work carried on in many different countries such as Poland, Latvia, Slovenia, Russia, Belarus, Namibia, Republic of South Africa, and others. The IOUNT was founded by Zygmunt Stoberski, author of numerous publications dealing with internationalisms as terminological neologisms.

'Neoterm' Journal. In a specialist IOUNT journal "Neoterm", issued in the years from 1984 to 1997, there were published many articles concerning general principles of terminology work in such domains as technology, chemistry, physics, nuclear energy, astronautics, statics, dynamics, physical culture, sports, architecture, environment, microbiology, medicine, economy, information science, communication, history, and music.

3.5.5 Society of Economic, Legal and Court Translators TEPIS and PolTerm Bank

Founded in 1990, TEPIS stands for the Polish Society of Economic, Legal and Court Translators and is a member of the International Federation of Translators. TEPIS aims are to enrich and disseminate the knowledge of the art of translation, including methodology of legal terminology described in numerous articles in the journal 'Lingua Legis' and the terminology developed by the TEPIS Publishing House, holder of the PolTerm bank of legal terminology in Polish-English and Polish-German. TEPIS website basic glossaries have been put on line with the access restricted to TEPIS members on the website: <http://www.tepis.org.pl>

'Lingua Legis' is a journal of the Polish Society of Economic, Legal, and Court Translators designed for specialized translators, in particular for legal and economic translators. The TEPIS Publishing House has edited thirteen issues of the journal, altogether containing over 130 articles dealing with the methodology of legal translation, often including lists of legal or specialized terms, then made accessible on-line at TEPIS website.

The PolTerm bank of legal terminology, developed and held by the TEPIS Publishing House, has always aimed to prepare, edit, and publish high-quality translations of Polish legislative instruments based on standardized terminology with one equivalent for one concept and to assist legal translators in finding right terms to express denotation of the Polish legal terms in English. Now, PolTerm is a medium-sized terminological database stored and updated in one of the recognized computer-aided translation (CAT) programs, the current number of entries amounting to 10,500 and covering a wide range of branches of law; it is based on 42 Acts of the Polish Parliament translated into English and regularly updated, published in “The Polish Law Collection” (PLC) issued by the TEPIS Publishing House: a two-volume loose-leaf and CD collection subscribed by over 300 institutions (a.o. the Sejm’s Chancellery, the National Bank of Poland) and firms (a.o. IKEA, ABB, KGHM, Ernst & Young, Baker & McKenzie, Deloitte & Touche, etc.).

3.5.6 Universities

The ‘Synaba’ database is a collection of current and completed research that have been carried out by university units. The database contains over 140 thousand records concerning mainly methodology-oriented and descriptive research in terminology. The detailed list of the ‘Synaba’ research works is in possession of the Information Processing Centre in Warsaw.

The same kind of research activities are carried on by several scientific NGO societies, such as Bialystok Scientific Society, Polish Section of the International Specialized Terminology Organization, Polish Linguistic Society, Polish Association of Applied Linguistics, Polish Society for Russian Studies, Polish Society for German Studies, and Polish Society of Economic Legal and Court Translators ‘TEPIS’.

3.5.7 Published dictionaries

A number of dictionaries edited by private publishers are available on the Polish market in paper version. They cover terminology in the various branches of science and knowledge in general, such as technology (metallurgy, mining, hydrogeology, food industry, transportation, patents, computer science), natural sciences (physics, biology, biochemistry, molecular biology, astronomy), medicine, economy (business, finance, taxation, insurance), history, theology and photography. The detailed list of dictionaries is in possession of the Information Processing Centre in Warsaw.

4 Terminology standardization

4.1 Major international standardization bodies

4.1.1 ISO standards in the terminology field

ISO is a global network of the national standards institutes of 151 countries, on the basis of one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system. The aim of the network is to identify what international standards are required by business, government and society, develop them in partnership with the sectors that will put them to use,

adopt them in transparent procedures based on national input and deliver them to be implemented worldwide.

ISO's work programme ranges from standards for traditional activities, such as agriculture and construction, through mechanical engineering, manufacturing and distribution, to transport, medical devices, the latest in information and communication technology developments, and to standards for services. New growth areas in the coming years are the environment, the service sectors, security and good managerial and organizational practice.

The ISO members propose new standards, participate in their development and, in collaboration with the secretariat, provide support for the 3000 technical committees and subcommittees that actually develop the standards.

The technical committee responsible for the development of standards for terminology is *TC 37 - Terminology and language and content resources*. The scope of this technical committee is the *standardization of principles, methods and applications relating to terminology and other language resources and content resources in the contexts of multilingual communication and cultural diversity*. It consists of four subcommittees which are listed below together with the standards developed by each subcommittee. Brief descriptions of some of the most relevant standards are given in Section 9 Appendices – ISO Standards. These are marked on the list.

TC 37/SC 1 - Principles and Methods

ISO 704:2000: Terminology work – Principles and methods, see appendix 9.1

ISO 860:1996: Terminology work -- Harmonization of concepts and terms, see appendix 9.2

ISO 1087-1:2000: Terminology work -- Vocabulary -- Part 1: Theory and application

TC 37/SC 2 - Terminography and lexicography

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 1951:1997: Lexicographical symbols and typographical conventions for use in terminography

ISO 10241:1992: International terminology standards -- Preparation and layout, see appendix 9.3

ISO 12199:2000: Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet

ISO 12615:2004: Bibliographic references and source identifiers for terminology work

ISO 12616:2002: Translation-oriented terminography

ISO 15188:2001: Project management guidelines for terminology standardization

TC 37/SC 3 - Computer applications for terminology

ISO 1087-2:2000: Terminology work -- Vocabulary -- Part 2: Computer applications

ISO 12200:1999: Computer applications in terminology -- Machine-readable terminology interchange format (MARTIF) -- Negotiated interchange, see appendix 9.4

ISO 12620:1999: Computer applications in terminology -- Data categories, see appendix 9.5

ISO 16642:2003: Computer applications in terminology -- Terminological markup framework, see appendix 9.6.

TC 37/SC 4 - Language resource management

This subcommittee has not yet developed any standards but several standards are under development.

4.2 Major national standardization bodies

4.2.1 Estonia

Various terminology committees, connected with various public and non-profit organisations:

- Aviation (Estonian Civil Aviation Administration)
- Botany (Tartu University)
- Bookkeeping (at Estonian Bookkeeping Committee - www.easb.ee)
- Ecology (Estonian Water Association, Estonian Geology Association and others)
- Electro-energy (Estonian Society for Electrical Power Engineering)
- Jurisdiction (Estonian Legal Language Centre)
- Geography (Estonian Geography Association)
- Librarianship (the National Library of Estonia has a terminology committee and it co-operate with the Estonian Librarians' Association).
- Meat production (The Estonian Chamber of Agriculture and Commerce)
- Marine (Estonian Maritime Administration)
- Medicine (Ministry of Social Affairs)
- Military (section of translation and editing *of the Estonian National Defence College*)
- Ornithology (Estonian Ornithology Association, 5 members)
- Pedagogy (Ministry of Education and Research of Estonia)
- Radiology (Estonian Society of Radiology)
- Rescue service (Väike-Maarja Rescue School and Estonian Rescue Board)

4.2.2 Hungary

The national standardization body of Hungary is the Hungarian Standards Institution (MSZT). The first standards had appeared in Hungarian language in 1868 (the rails and sleepers of the railway of Fiume). This was followed by the standardization of the building materials, the firetools, the electrotechnical tools and methods, among others with the management of experts such as Miklós Ybl, architect. The institutionalized standardization began in 1921 by the foundation of the Hungarian Industrial Standardizing Committee of which the president was Kálmán Kandó. This committee worked as a separate professional body with small transformations till 1948, till the socialization. Since then it lived on as a governmental organization till 1995 (see Law XXVIII of 1995 on national standardization), when it became a separate, non-profit public body again – on the name Magyar Szabványügyi Testület (Hungarian Standards Institution) – because of partially external (joining of Hungary to the EU), on the other hand internal (market-economical relations) causes. From then the standardization is a supporting (and not prohibitive) tool of the participants of the market. They attend in the development of the standards on a national, European and international level not by necessity but in order to vindicate their interests.

The international acknowledgement of the standardization work of the MSZT is indicative of that it is full member of every international and European standardization organization. On the field of certification its system certification is honoured world wide by its membership of the International Certification Network (IQNet), its employee certification is honoured in Europe by the acknowledgement of the European Organization for Quality (EOQ), its standard product certification is honoured all over the world by its membership of the international standards organizations. The high level domestic acceptance is an essential term of these acknowledgements. Based on the professional plane of the standardization activity existing from 1868, the preparedness and human attitude of the participant experts, the transformations based on the results and

experiences of the former decades generating appreciation at a European and world wide level as well, the real meaning of **MSZT** letter word is **Marketability, Skill and Zealous Trustworthiness**.

Hungary is represented by MSZT in the following international and European standards organizations:

- International Organization for Standardization (ISO)
- International Electrotechnical Commission (IEC)
- European Committee for Standardization (CEN)
- European Committee for Electrotechnical Standardization (CENELEC)
- European Telecommunication Standards Institute (ETSI)

MSZT is a member institution of the Terminology Council of the Hungarian Language (Magyar Nyelv Terminológiai Tanácsa, MATT).

4.2.3 Latvia

In accordance with the State Language Law, the terminology Commission of LAS is the main standardization body in the terminology field in Latvia. The main functions of TC of LAS include the approval of new terms and definitions and passing decisions on issues in using Latvian terms (see about TC of LAS and their subcommissions in 3.3.1).

At the same time, there is a formal standardization body – the Latvian Standardization Organization “Latvijas Standarts” (LVS) (“Standard of Latvia”), which encompasses two departments: Information Department and Standardization Department. The Standardization Department manages contacts with different Technical Committees (see additional information in 3.3.8).

The list of LVS standards harmonized with terminology approved by TC of LAS:

1. **Izdevējprodukcija un izlaides ziņas. Terminu un definīcijas.** — LRS, 6 - 91. Rīga, 1991. — 31 lpp.
2. **Ugunsdrošība. Terminu un definīcijas.** — LVS 38 - 93. — Rīga, 1993. — 17 lpp.
3. **Kvalitātes pārvaldība un nodrošināšana.** — LVS ISO 8402: 1994. — 33 lpp.
4. **Kvalitātes sistēmas.** LVS EN ISO 9001: 2001. — 33 lpp.
5. **Kvalitātes sistēmas. Ražošanas, montāžas un apkalpošanas kvalitātes nodrošināšanas modelis.** — LVS EN ISO 9002: 1994. — 24 lpp.
6. **Kvalitātes sistēmas. Noslēguma pārbaudes un testēšanas kvalitātes nodrošināšanas modelis.** — LVS EN ISO 9003: 1994. — 21 lpp.
7. **Kalibrēšanas un testēšanas laboratoriju akreditācijas sistēmas. Darbības un atzīšanas vispārīgās prasības.** — LVS EN 45003. — 20 lpp.
8. **Drošuma un pakalpojumu kvalitāte.** — LVS IEC 50 - 191: 1990. — 63 lpp.
9. **Elektroenerģijas ražošana, pārvade un sadale. Vispārīgie termini.** — LVS IEC 50 - 601: 1985. — 19 lpp.

10. **Elektroenerģijas ražošana, pārvade un sadale. Elektroenerģijas ražošana.** — LVS IEC 50 - 602: 1983. — 23 lpp.
11. **Elektroenerģijas ražošana, pārvade un sadale. Energosistēmu attīstības un vadības plānošana.** — LVS IEC 50 - 603: 1986. — 30 lpp.
12. **Elektroenerģijas ražošana, pārvade un sadale. Elektroietaišu ekspluatācija.** — LVS IEC 50 - 604: 1987. — 37 lpp.
13. **Elektroenerģijas ražošana, pārvade un sadale. Elektriskās apakšstacijas.** — LVS IEC 50 - 605: 1983. — 21 lpp.
14. **Elektroenerģijas ražošana, pārvade un sadale. Elektroenerģijas tarifi.** — LVS IEC 50 - 691: 1973. — 27 lpp.
15. **Ugunsaisardzība. Vārdnīca. 1. daļa: Ugunsgrēks. Galvenie termini un parādības** (Fire protection – Vocabulary – Part 1: General terms and phenomena of fire). – LVS ISO 8421-1: 2003. – 9 lpp.
16. **Ugunsaisardzība. Vārdnīca. 2. daļa: Ēku ugunsaisardzība** (Fire protection – Vocabulary – Part 2: Structural fire protection). – LVS ISO 8421-2: 2003. – 6 lpp.
17. **Ugunsaisardzība. Vārdnīca. 3. daļa: Ugunsgrēka atklāšana un trauksmes izziņošana** (Fire protection – Vocabulary – Part 3: Fire detection and alarm). – LVS ISO 8421-3: 2003. – 9 lpp.
18. **Ugunsaisardzība. Vārdnīca. 4. daļa: Ugunsgrēka dzēšanas iekārtas** (Fire protection – Vocabulary – Part 4: Fire extinction equipment). – LVS ISO 8421-4:2003. – 17 lpp.
19. **Ugunsaisardzība. Vārdnīca. 5. daļa: Dūmu kontrole** (Fire protection – Vocabulary – Part 5: Smoke control). – LVS ISO 8421-5: 2003. – 5 lpp.
20. **Ugunsaisardzība. Vārdnīca. 6. daļa: Evakuācija un glabāšanas līdzekļi** (Fire protection – Vocabulary – Part 6: Evacuation and means of escape Bilingual edition). – LVS ISO 8421-6: 2003. – 8 lpp.
21. **Ugunsaisardzība. Vārdnīca. 7. daļa: Sprādzienbīstamības noteikšanas un novēršanas līdzekļi** (Fire protection – Vocabulary – Part 7: Explosion detection and suppression means Bilingual edition). – LVS ISO 8421-7: 2003. – 5 lpp.
22. **Ugunsaisardzība. Vārdnīca. 8. daļa: Ugunsdzēsības un glābšanas dienests un darbs ar bīstamiem materiāliem** (Fire protection – Vocabulary – Part 8: Terms specific to fire-fighting, rescue services and handling hazardous materials). – LVS ISO 8421-8: 2003. – 28 lpp.
23. **Ugunsdrošība. Vārdnīca** (Fire safety – Vocabulary). – LVS EN ISO 13943: 2004. – 28 lpp.

4.2.4 Lithuania

The main documents establishing requirements for terminology standardization in Lithuania are:

- 2000 - Law on Standardization (Official Gazete 2000, No. 35-972)
- The Law of the Republic of Lithuania on State Language (Official Gazete 1995, No.15-344)
- The Law of the Republic of Lithuania on bank of terms (Official Gazete 2004, No.7-129)
- 1998-01-29 Resolution of the State Lithuanian Language Commission No.18 Rules for Approval of Terminology Standards (Official Gazete 1998, No.18-446)

- 2000-06-05 Order of LST Director No.39 „Rules for Lay-out of Lithuanian terminology standards, Clauses of Terms and definitions and Symbols and abbreviations of not terminology standards“
- 2000-06-15 Order of LST Director Nr.42 “Rules for Preparation, Expertise and Approval of Terminology Standards”
- 2003-02-03 Order of LST Director No. V-5 “Order on the Drafting Rates of Lithuanian Standards”;
- Statute of the LST/TC 37 Terminology.

Requirements for preparation of Lithuanian terminology standards are also included in basic Lithuanian standards:

- LST 0-2: *Standardization. Part 2. Preparation of Lithuanian standards*;
- LST 0-3-1: *Standardization. Part 3. Drafting and presentation of Lithuanian Standards. Chapter 1. General provisions*;
- LST 0-3-2: *Standardization. Part 3. Content and Lay out of Lithuanian standards. Chapter 2. Terminology Standards (mod ISO 12041:)*;
- LST 0-6: *Standardization. Part 6. Adoption of publications of the European organizations for standardization as Lithuanian standards.*

Lithuanian standards adopted by LST TC 37 *Terminology* needed in the process of creation of Lithuanian terms and definitions included in Lithuanian terminology standards and clauses *Terms and definitions* of non-terminology Lithuanian standards are in the collection of Lithuanian standards:

LST EN 45020:2005 *Standartizacija ir su ja susijusi veikla. Bendrasis aiškinamasis žodynas (ISO/IEC vadovas 2:1996)* (Standardization and related activities – General vocabulary);

LST ISO 704:2005 *Terminologijos darbas. Principai ir metodai* (Terminology work - Principles and methods);

LST ISO 860:2005 *Terminologijos darbas. Sąvokų ir terminų darninimas (tpt ISO 860:1996)* (Terminology work - Harmonization of concepts and terms);

LST ISO 1087-1:2005 *Terminologijos darbas. Aiškinamasis žodynas. 1 dalis. Teorija ir taikymas (tpt ISO 1087-1:2000)* (Terminology work – Vocabulary – Part 1: Theory and application (idt ISO 1087-1:2000));

LST ISO 1087-2:2005 *Terminologijos darbas. Aiškinamasis žodynas. 2 dalis. Kompiuterių taikymas (tpt ISO 1087-2: 2000)* (Terminology work – Vocabulary - Part 2: Computer applications (idt ISO 1087-2:2000));

Pr LST ISO 1951:2005 *Terminografijoje vartojami leksikografiniai simboliai ir spaudos sutartiniai* (Lexicographical symbols and typographical conventions for use in terminography);

Pr LST ISO 12616:2005 *Vertimui skirta terminografija (tpt ISO 12616:2002)* (Lexicographical symbols and typographical conventions for use in terminography);

LST ISO 12199:2002 *Abėcėlinis daugiakalbių terminologinių ir leksikografinių duomenų, išreikštų lotyniškais rašmenimis, išdėstymas (tpt ISO 12199:2000)* (Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet (idt ISO 12199:2000));

LST ISO 12200:2002 *Kompiuterių taikymas terminologijoje. Kompiuterinio skaitymo terminų mainų formatas. Sutartiniai mainai (tpt ISO 12200:1999)* (Computer applications in terminology - Machine-readable terminology interchange format (MARTIF) - Negotiated interchange (idt ISO 12200:1999)); LST ISO 12620:2003 *Kompiuterių taikymas terminologijoje. Duomenų kategorijos (tpt ISO 12620:1999)* (Computer applications in terminology - Data categories (idt ISO 12620:1999)); LST ISO 15188:2003 *Terminijos standartizavimo projektų vadybos gairės (tpt ISO*

15188:2001) (Project management guidelines for terminology standardization (idt ISO 15188:2001));

LST ISO 6156:2003 *Keičiamo formato magnetinės juostos terminologiniams ir leksikografiniams įrašams (MATER) (tpt ISO 6156:1987)* (Magnetic tape exchange format for terminological/lexicographical records (MATER) (idt ISO 6156:1987)); LST ISO 639-1:2003 *Kalbų pavadinimų kodai. 1 dalis. Dviraidžiai kodai (tpt ISO 639-1:2002)* (Codes for the representation of names of languages - Part 1: Alpha-2 code (idt ISO 639-1:2002));

LST ISO 639-2:2003 *Kalbų pavadinimų kodai. 2 dalis. Triraidžiai kodai (tpt ISO 639-2:1998)*; (Codes for the representation of names of languages - Part 2: Alpha-3 code (idt ISO 639-2:1998));

LST ISO 12199:2002 *Abėcėlinis daugiakalbių terminologinių ir leksikografinių duomenų, išreikštų lotyniškais rašmenimis, išdėstymas (tpt ISO 12199:2000)* (Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet (idt ISO 12199:2000));

LST ISO 12200:2002 *Kompiuterių taikymas terminologijoje. Kompiuterinio skaitymo terminų mainų formatas. Sutartiniai mainai (tpt ISO 12200:1999)* (Computer applications in terminology - Machine-readable terminology interchange format (MARTIF) - Negotiated interchange (idt ISO 12200:1999));

LST ISO 12620:2003 *Kompiuterių taikymas terminologijoje. Duomenų kategorijos (tpt ISO 12620:1999)* (Computer applications in terminology - Data categories (idt ISO 12620:1999));

LST ISO 15188:2003 *Terminijos standartizavimo projektų vadybos gairės (tpt ISO 15188:2001)* (Project management guidelines for terminology standardization (idt ISO 15188:2001));

LST ISO 6156:2003 *Keičiamo formato magnetinės juostos terminologiniams ir leksikografiniams įrašams (MATER) (tpt ISO 6156:1987)* (Magnetic tape exchange format for terminological/lexicographical records (MATER) (idt ISO 6156:1987));

LST ISO 639-1:2003 *Kalbų pavadinimų kodai. 1 dalis. Dviraidžiai kodai (tpt ISO 639-1:2002)* (Codes for the representation of names of languages - Part 1: Alpha-2 code (idt ISO 639-1:2002));

LST ISO 639-2:2003 *Kalbų pavadinimų kodai. 2 dalis. Triraidžiai kodai (tpt ISO 639-2:1998)* (Codes for the representation of names of languages - Part 2: Alpha-3 code (idt ISO 639-2:1998)

4.2.5 Poland

The Polish Committee for Standardization was founded in 1924, when also the first Polish Standard (PN) was introduced as an official document for voluntary use. In 1945 the ‘PKN’ Committee was reinstated, in 1947 it became a charter member of ISO, and in 1953 was charged with the responsibility of a central standardization body in Poland. In 1991 PKN became an affiliate of CEN and CENELEC and in 2004 became a full member therein. In 2003 the new Standardization Law entered into force and made the PKN standardization system fully compatible with the European one.

Activities in terminology work. The PKN Committee is carrying on regular terminological activities in two fields: through adaptation of ISO terminology standards and changing them into ‘PN-ISO’ standards, and through collecting data bases of unified terminology extracted from Polish standards dealing with terminology issues.

Technical Committee No. 256. The task of adapting ISO standards and transforming them into Polish “PN-ISO” standards was assigned to Technical Committee No. 256 for Methods and Principles of Terminology Work in 1994. Its activities included principles of creating concept systems, principles and methods of defining concepts, principles and methods of selecting and creating terms, principles and methods of implementing international and regional terminology standards in national terminology standards; preparation and layout of standards and other

terminology documents, including arrangement, content and layout of terminological entries, lexicographical symbols and typographical conventions to be applied in standards and other terminology documents, codes for the representation of names of languages, alphabetical organization of alphanumerical terminological data; computational aids in terminology work, including terminological data elements, the format of records serving the interchange of terminological data, the principles of setting up terminological databases and management thereof, the application of SGML (Standardized General Mark-Up Language) in terminology work.

Methodology for standardization work of TC 256. Technical Committee No. 256 is carrying on its activities in compliance with the *RPN-007 Rules for introduction of international standards into Polish standards* passed in 1997.

The Rules determine comprehensively the detailed subject and field of its application, general principles and methods of preparing international standards to be introduced, methods of their updating, editing and translating, elements of the standard and its layout, etc.

Composition of the Technical Committee No. 256. Chairperson of the TC 256 is Prof. Dr. Hab. Wanda Zmarzer of the University of Warsaw, Faculty of Applied Linguistics, Chair of Specialized Languages. In 2005 the Committee has 11 other members, namely: Prof. Dr. Hab. Jerzy Lukszyn, Prof. Dr. Hab. Jan Lewandowski, Prof. Dr. Hab. Kazimierz Luciński, Dr. Danuta Kierzkowska, Dr. Aniela Topulos, Ms. Hanna Kłodnicka, Mr. Mariusz Górnicz, Ms. Jadwiga Różycka, Ms. Irena Wojtasiak, Mr. Piotr Ubysz, Mr. Radosław Pyra.

List of Polish standards developed by TC 256:

PN-ISO 860:1998 *Działalność terminologiczna – Opracowanie i układ.*

[Terminology work - Harmonization of concepts and terms]

PN-ISO 6156:1998 *Format do wymiany rekordów terminologicznych/leksykograficznych na taśmie magnetycznej (MATER)*

[Magnetic tape exchange format for terminological/lexicographical records (MATER)]

PN-N-02001:1994 *Podstawy działalności normalizacyjnej – Metodyka prac normalizacyjnych*

[Bases for the standardization activities - Methodology for standardization works]

PN-ISO 1087-2:2001 *Działalność terminologiczna – Terminologia – Zastosowanie komputerów*

[Terminology work - Vocabulary - Part 2: Computer applications] PN-ISO 10241:1997 **Normy terminologiczne – Opracowanie i układ** [Terminology standards - Preparation and layout]

PN-ISO 1951:1998 *Użycie symboli leksykograficznych i konwencji typograficznych w terminografii*

[Lexicographical symbols and typographical conventions for use in terminography]

PN-ISO 639-2:2001 *Kody nazw języków – Kod trzyliterowy*

[Codes for the representation of names of languages - Part 2: Alpha-3 code]

PN-ISO 12620:2002 *Zastosowanie komputerów w terminologii – Kategorie danych*

[Computer applications in terminology - Data categories]

PN-ISO 12200:2002 *Zastosowanie komputerów w terminologii - Format wymiany danych terminologicznych odczytywanych komputerowo (MARTIF)*

[Computer applications in terminology - Machine-readable terminology interchange format (MARTIF) - Negotiated interchange]

PN-ISO 12199:2005 *Porządek alfabetyczny wielojęzycznych danych terminologicznych i leksykograficznych przedstawianych alfabetem łacińskim*

[Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet]

PN-ISO 1087-1:2004 *Działalność terminologiczna – Terminologia – Część 1. Teoria i zastosowanie* [Terminology work - Vocabulary - Part 1: Theory and application]

PN-ISO 704:2005 *Działalność terminologiczna – Zasady i metody*
[Terminology work - Principles and methods]

PN-ISO 15188:2005 *Wytyczne zarządzania przedsięwzięciem w zakresie normalizacji terminologii*
[Project management guidelines for terminology standardization project]

PN-ISO 639-1:2005 *Kody nazw języków – Kod dwuliterowy*
Codes for the representation of names of languages - Part 1: Alpha-2 code

PN-ISO 12616:2005 *Terminografia translacyjna*
[Translation-oriented terminography]

Unified PKN Terminology Bank ‘BTZ’. The bank contains about 77 thousand records concerning terms extracted from Polish standards (for more detailed description see item 6.5.1.).

4.2.6 Germany

DIN, the German Institute for Standardization (Deutsches Institut für Normung e. V.) was founded in 1917. Its head office is in Berlin. Since 1975 it has been recognized by the German government as the national standards body and represents German interests at international and European level.

DIN offers a forum in which representatives from the manufacturing industries, consumer organizations, commerce, the trades, service industries, science, technical inspectorates, government, in short anyone with an interest in standardization, may meet in order to discuss and define their specific standardization requirements and to record the results as German Standards.

Standardization as undertaken by DIN is a service that aims to benefit the entire community. The results of its work have a significant influence on economic performance at both company and national level. A research project completed in 2000 confirmed the annual benefit to the German economy as being 1 % of GNP, or approx. US\$ 15 billion.

DIN Standards promote rationalization, quality assurance, safety, and environmental protection as well as improving communication between industry, technology, science, government and the public domain.

The main activity of DIN is the development of technical rules. In drawing up a new organizational structure, a clear line has been drawn between standardization on the one side and business activities on the other. The objective of DIN is to create standards for the benefit of the economy and of society as a whole. The business activities of the companies within the DIN Group are profit-oriented. The income generated by the subsidiary companies of DIN, and from those companies in which it is a shareholder, represents the single largest contribution to the financing of the not-for-profit core activities of DIN.

The input of external experts into standardization is organized in standards committees and their subsidiary working bodies. One standards committee is responsible for each distinct area of activity and also coordinates the corresponding standardization work at European and international level. As a rule, the standards committee in DIN comprise a number of technical committees. There are currently 76 standards committees, in which some 26,000 external experts are working as voluntary delegates on the standards. Draft standards are published for public comment, and all comments are reviewed before final publication of the standard. Published standards are reviewed for continuing

relevance every five years, at least. In 2004, the number of DIN standards amounts to nearly 29,000, from which 15,200 are available in English. DIN achieved in 2004 a turnover of 56 million EUR.

The standards are published and sold by the publishing house Beuth Verlag that specializes in sales and distribution of standards, directives and other normative documents.

Within terminology standardization two main directions can be distinguished: the standardization of concepts and the respective terms and the definition of principles and guidelines for terminology work and terminology standardization. The standardization of concepts and terms is in general carried out by subcommittees of the respective technical Standards Committees. They include the standardized terminology in the corresponding technical standard or publish it in the form of specific terminology standards.

The concepts enclosed in DIN standards with their terminological representations, definitions and further information are documented in a terminology database called DIN-TERM. As the standards often contain foreign-language terms and definitions, DIN-TERM covers not only German terms, but also English and French equivalents. The service responsible for DIN-TERM plans to make the data available to users in the form of technical dictionaries and electronic databases. Today DIN-TERM contains more than 210,000 entries.

For the standardization in the field of the principles of terminology work the Terminology Standards Committee (Normenausschuss Terminologie, NAT) is responsible. It focuses on the fundamental significance of technical language for standardization in general as well as on tools for terminology work, translation and lexicography. The primary areas of responsibility are: principles of concept and term formation, elaboration and configuration of technical dictionaries, computer applications for terminology work and lexicography, terminology of the terminology work, terminological practice and technical translation. NAT represents German interests in ISO/TC 37 *Terminology and other language resources*.

5 General research principles in terminology work

Terminology work (or a terminological resource) is often confused with lexicology work (or a lexical resource), but these concepts really represent distinct disciplines.

Lexicology concerns the study of words in general and lexicography concerns the development of general language dictionaries. These dictionaries are word-based and each entry contains a headword and all the different senses of this headword. In other words polysemes (same form, but different meaning) are grouped together in the same entry and synonyms (different form, but same meaning) are scattered throughout the dictionary.

Terminology concerns the study of specialised language terms within an area of specialist knowledge and terminography concerns the development of a terminological resource, sometimes also called a glossary, a specialised vocabulary, a term collection etc. A terminological resource is concept-based and an entry contains all the terms that designate one particular concept. This means that synonyms are grouped together in the same entry and polysemes are presented separately in different entries.

This chapter is concerned with some general principles of terminology work and the vast majority of the information found in this chapter is a paraphrase of information contained in ISO standards. The purpose of this chapter is to give a coherent overview of processes and methods often applied in terminology work.

5.1 Subject field classification

Concepts in a term collection are organized in relation to a subject field classification system. It is essential to study concepts (and their terms) in relation to a subject field in order to delimit the scope and focus on certain characteristics of a particular concept. The term *emission* within the subject field Automotive Engineering might e.g. mean ‘the process of emitting exhaust gases’ and the same term within the subject field Environmental Pollution might mean ‘the exhaust gases themselves’.

A company can choose to adopt or adapt an existing classification system or a new one can be developed from scratch. Quite a number of existing classification systems is available, but most of them are documentary and there is a difference between classifying documents and concepts. It is of course essential that a classification system corresponds exactly to the activity field of a particular company.

5.1.1 LENOCH universal classification system

The Lenocho classification system is a complex and fine-grained classification system covering a wide range of subject fields and widely used for both terminology and documentation, among others in Eurodicautom, the multilingual term bank of the European Commission (now obsolete, integrated into IATE, but still to be accessed at the web at <http://europa.eu.int/eurodicautom/Controller>) and in the term bank of the Swiss Federation .

At the top level Lenocho works with 48 subject fields each labelled with a two-letter code

AD	Management in the public and private sector
AG	agriculture, fisheries, forestry - food processing industries
AR	Art
AS	Insurance
AT	Nuclear industry (with applied atomic and nuclear physics)
AU	Automation (includes telecommunications and computers)
BA	Building industry
BZ	Botany and zoology
CE	The European Communities
CH	Chemistry
CO	Commerce - movement of goods
DE	Defense
DI	Documentation and information
DO	Domestic economy
EC	Economics
ED	Education

EL	Electrical engineering and energy
EN	Environment
ER	Earth resources - energy
FI	Financial affairs - taxation - customs
GE	Generic civilization - heritage
GO	The cosmos
HI	History, ethnology, manners and customs
IC	The chemical industry
IN	Various industries and crafts
JU	Law
LA	Language and literature
MA	Mathematics
ME	Medicine
MG	Mechanical engineering
MI	Mining
NO	Standards, measures and testing
OO	News-systems and communications
OR	International organisations
PG	Printing and publishing
PH	Physics
PO	Politics
RP	Religion and philosophy
RS	Risk Management - security
SC	Co-operatives
SI	Iron and steel industries
SO	Man and society
SP	Sports, entertainments and leisure
ST	Statistics
TE	Technical and industry in general
TR	Transport
TS	Land and property
TV	Labour

Each of the subject fields has a number of subclasses from about 10 to about 100, which are labeled with a letter or digit following the subject field code. Some times even an additional level is provided, thus allowing for a very precise classification of concepts, e.g.

AT Nuclear industry (with applied atomic and nuclear physics)

AT1 – general

AT2 – institutions – public offices – organizations

AT3 – publications

AT3#1 documentation & information systems
AT4 – nuclear fission
AT5 – nuclear fusion
AT6 – applied nuclear science
AT7 – nuclear propulsion & nuclear powering
AT8 – nuclear safeguards – hazard control
AT9 – machines – apparatus – processes
ATA – radio-active materials
ATW – man & society
ATY – training & education
ATZ – professions & careers

As a direct consequence of its fine-graininess, some term bank administrators have found Lench too complex to manage in, making it hard for the translator to classify the terms correctly, as it appears from Ball and Rummel (2001):

“The IATE work group responsible for questions related to the content of the database (Data Content Group) decided to adopt the EuroVoc thesaurus for the domain classification of entries in the EU term base. The alternative proposition, the Lench classification that is used in the Eurodicautom database, was regarded as a complex, very rich, fine-grained system, that allows for a very precise classification of concepts. This positive characterisation is at the same time the reason why, after some discussions, it was decided to vote for EuroVoc: Lench demands expertise in classification. Translators would be able to enter first-level codes, but the allocation of lower-level codes would have to be done by experts.”

Still Lench is one of the most widely used classification systems and constitutes the basic classification, possibly with some adjustments, of many term banks.

5.1.2 Eurovoc

Eurovoc is a multilingual thesaurus, developed in accordance with the ISO **ISO 2788-1986** - Guidelines for the establishment and development of monolingual thesauri and **ISO 5964-1985** - Guidelines for the establishment and development of multilingual thesauri. It covers the fields in which the European Communities are active, provides a means of indexing the documents in the documentation systems of the European institutions and of their users and is currently used by the European Parliament, the Office for Official Publications of the European Communities, national and regional parliaments in Europe, national government departments and certain European organisations. Eurovoc exists in **16 official languages of the European Union** (Spanish, Czech, Danish, German, Greek, English, French, Italian, **Lithuanian**, **Hungarian**, Dutch, Portuguese, Slovak, Slovene, Finnish and Swedish). In addition to these versions, it has been translated by the parliaments of a number of countries (Albania, Croatia, **Latvia**, **Poland**, Romania and Russia).

Eurovoc can be downloaded for free from the web site

http://europa.eu.int/celex/eurovoc/cgi/sga_doc?eurovoc_dif!SERVEUR/menu!prod!MENU&langue=EN in the 11 “old EU languages”. From this site it is also possible to navigate in the thesaurus. It does not appear from the site where to get access to the other language versions.

At the top level Eurovoc deals with the following 21 subject fields which are all of importance for the activities of the European institutions:

04 POLITICS

08	INTERNATIONAL RELATIONS
10	EUROPEAN COMMUNITIES
12	LAW
16	ECONOMICS
20	TRADE
24	FINANCE
28	SOCIAL QUESTIONS
32	EDUCATION AND COMMUNICATIONS
36	SCIENCE
40	BUSINESS AND COMPETITION
44	EMPLOYMENT AND WORKING CONDITIONS
48	TRANSPORT
52	ENVIRONMENT
56	AGRICULTURE, FORESTRY AND FISHERIES
60	AGRI-FOODSTUFFS
64	PRODUCTION, TECHNOLOGY AND RESEARCH
66	ENERGY
68	INDUSTRY
72	GEOGRAPHY
76	INTERNATIONAL ORGANIZATIONS

Some subject fields are more developed than others, depending on how closely related they are to the Community's centres of interest.

5.2 Identification and evaluation of specialized documentation

A large portion of the knowledge represented in a term collection is usually gathered from various written sources. Therefore sources are part of the backbone of a terminology collection and it is important that sources are selected and evaluated according to certain criteria. Sources should ideally be read and examined, perhaps by subject field experts, in order to select the documents that are most representative of the particular subject field.

5.3 Identification of relevant concepts and relations

A concept is a categorization or a mental representation of a set of objects that share the same properties or characteristics. Every single object in the world does not have its own individual name; that would make communication much too complicated. For example a number of objects in the world are human-built dwellings with enclosing walls and a roof. The concept or the mental representation of these objects is "house" (not the word "house", but the language independent image of a house). Possible designations of this concept are e.g. house, building, Haus, maison and many others.

Terminology work is always based on concepts and not the terms that designate these concepts. One reason is that differences between concepts are not necessarily transparent at the term level and similarly nearly identical terms in different languages do not necessarily imply that the concepts behind these terms are identical.

Concepts are not isolated items; they should be seen in relation to each other because the relations are important to elucidate delimitations and characteristics of each concept. Concepts are therefore organized in concept systems within the framework of the particular subject field.

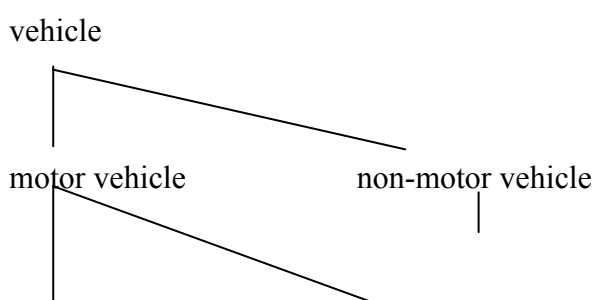
5.3.1 Conceptual analysis and concept system

The documentation identified for a terminological purpose is carefully examined and relevant terms and textual supports are extracted from the material (important textual supports are for example definitions, grammar information, context examples, collocations etc.).

The result of such an examination of the documentation material is a fairly unstructured amount of information. A list of terms extracted from some documentation could for example be the following:

Automobile
Car
Vehicle
Door
Wheel
Hand brake
Tyre
Taxi
Ambulance
Truck
Parking brake
Lorry
Cab
Wagon
Motor vehicle
Brake
Service brake
Hand brake
Emergency brake

A conceptual analysis involves an analysis of concept characteristics and will contribute to the structuring and harmonization of the gathered information. An element in the conceptual analysis is development of a concept system reflecting the concepts these terms represent and demonstrating their interrelations.



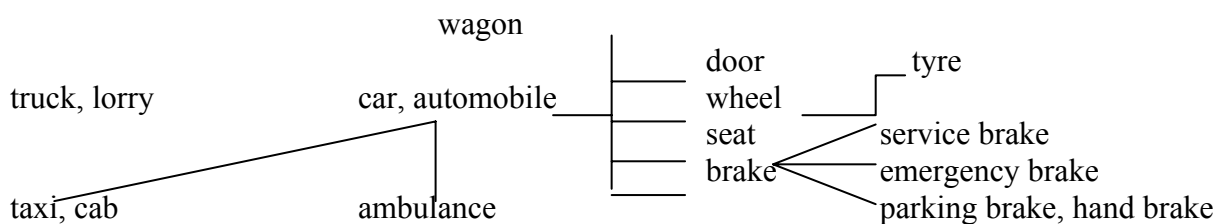


Figure 5.1. Fragmented concept system for vehicles

The above is a (very) fragmented concept system for vehicles. The concept system reveals synonyms, the structure of concepts and the different types of relations that hold between the concepts. For example, we have a generic relation between *vehicle* and *motor vehicle* showing that a motor vehicle is a kind of vehicle and a partitive relation between *car* and *brake* showing that a brake is part of a car. Partitive and generic relations are just some of the relations that may be used in a concept system. Other types of relations could be temporal, locative and associative relations.

5.3.2 Harmonization between languages

Communication between different nations is impeded by misleading similarities between terms in different languages and differences between concept systems covering the same subject field. Therefore a certain degree of harmonization is often necessary.

Concept harmonization concerns the elimination of minor differences between two or more concepts and term harmonization concerns the designation of one concept in different languages by terms that reflect the same characteristics or have the same form (or slightly different forms).

Harmonization should especially be considered if the subject field is well established or if it deals with concrete objects (e.g. tools, machinery) or if there is a tradition of standardization in the subject field.

5.4 The terminological unit

A term is a single word or a combination of words within a specialized domain always used to designate a particular concept. Terms are usually nouns, but they may also be verbs, adjectives or adverbs.

In connection with multiword terms the shortest possible form that designates the particular concept should be chosen. When it seems unclear whether a particular word of a multiword term should form part of the term or not, it should be tested whether the particular sequence of words occurs repeatedly in the documentation material.

5.4.1 Writing rules

Writing rules are important to regulate and standardize the contents of each textual support (textual supports are for example definitions, contexts and usage information). Writing rules should contain information about the *purpose* together with information about the *contents* of each textual support.

For example, writing rules concerning the textual support ‘definition’ might determine that the *purpose* of the definition is to be able to distinguish the particular concept from co-ordinate concepts of the concept system (this of course implies that concept systems exist). Concerning the *contents* of the ‘definition’ writing rules might stipulate that the definition should be only one sentence, start with a lower-case letter and describe the characteristics that differentiate the concept from co-ordinate concepts.

6 Major existing terminology resources, bodies and networks

6.1 Termbanks, bodies and networks

6.1.1 IATE –Termbanks, bodies and networks

IATE - A Single Terminology Database for the European Union

The Translation Centre launched the “IATE” (“**I**nter-**A**gency **T**erminology **E**xchange”) project in 1999; its initial objective was to create an infrastructure for the management of terminology for the Centre and the decentralised agencies of the Union. The other translation services joined this initiative in the same year and gave the project its truly inter-institutional status. Given this change in scope the acronym today stands for “**I**nter-**A**ctive **T**erminology for **E**urope”.

The project partners, who use and jointly finance the IATE database, are:

- European Commission
- Council
- Parliament
- Court of Auditors
- Economic and Social Committee
- Committee of the Regions
- Court of Justice
- Translation Centre for the Bodies of the EU
- European Investment bank
- European Central Bank.

The project was based on the outcome of an external study that had analysed the terminology arrangements of the EU services. The study recommended:

- incorporation of all existing terminology databases into a single new inter-institutional database;
- adoption of fully interactive creation of terminological data;
- reorganisation of terminology activity and reinforcement of staffing where necessary;
- integration of terminology tools, including tools designed to support terminology activities, into translation and office automation environment;
- deployment of new ergonomic and user-friendly interfaces;
- co-operative infrastructure for data management rational recovery of existing data.

After a development phase of three years the operational phase of the IATE project began in the summer of 2004. The system offers today the following

- One common database for all institutions and agencies containing all legacy data;
- On-line access in read and write mode, i.e. the possibility for users to carry out modifications, to add entries directly to the central database and thus to allow their colleagues to profit from this work immediately;
- Validation procedure to ensure quality. Possibility to define validation cycles, validation stages, user profiles, user roles etc. for each participating institution and agency;
- Management tools (e.g. for user management, data consolidation);
 - Features for large scale processing (export and import of data);
 - Reporting and auditing tools, e.g. the possibility to trace modifications in terminological entries;
 - A messaging system as communication mechanism between the actors in the terminology workflow;

IATE in Figures

Number of terms in the official EU languages

cs-Czech	16157
da-Danish	615516
de-German	1097792
el-Greek	524686
en-English	1503353
es-Spanish	632560
et-Estonian	14624
fi-Finnish	329634
fr-French	1427075
hu-Hungarian	22277
it-Italian	717718
lt-Lithuanian	29245
lv-Latvian	8556
mt-Maltese	556
nl-Dutch	710846
pl-Polish	21183
pt-Portuguese	543419
sk-Slovak	15917
sl-Slovenian	13313
sv-Swedish	317075
la-Latin	83352
ga-Irish	20572

Additional information

Language Independent Level	
Domain Notes	293053
Origin Notes	47887

Graphics	7
Language Level	
Definitions	98667
Notes	389161
Graphics	1
Term Level	
Notes	1023971
Term References	1933473
Contexts	34003
Context References	23162

User Activity

- Overall user population (i.e. user with a personal login): 1100
- Average number of queries per day: 10.000
- Average number of sessions per day: 5.700
- Average number of terms created per day: 550
- Average number of terms modified per day: 300
- Average number of terms validated per day: 300

6.1.2 InfoTerm

The International Information Centre for Terminology (Infoterm) was founded in 1971 by UNESCO, the United Nations Educational, Scientific and Cultural Organization, with the objective to support and co-ordinate international co-operation in the field of terminology. In 1996, Infoterm was re-organized and established as an independent non-profit organization.

Infoterm's mission is to promote and support the co-operation of existing and the establishment of new terminology centres and networks with the general aim to improve specialist communication, knowledge transfer and provision of content with a view to facilitate the participation of all in the global multilingual knowledge society.

In order to achieve this objective, Infoterm members co-operate in organizing a world-wide network of terminology centres and terminology networks with a view to:

- disseminating information on terminological activities as well as enhancing the awareness for the importance of terminology in all spheres of society,
- furthering the preparation of reusable terminologies by subject-field specialists in co-operation with terminologists,
- sharing the expertise regarding harmonized methods and guidelines for terminology management, the management of terminology centres, and for the use of terminological data, methods and tools in all applications where specialized information and knowledge are involved.

In discharging its functions Infoterm is co-operating with pertinent international, regional and national organizations and their respective members.

Infoterm's vision is to organize the methodological and organizational basis for a most efficient and effective preparation of terminologies in the form of net-based distributed co-operative terminology work under a comprehensive content management approach guaranteeing semantic interoperability across all application fields.

Infoterm members are either terminology organizations/institutions or specialized organizations/institutions with major terminological activities, which can be considered to be authorities in their field. They are exclusively public institutions, intergovernmental organizations and non-profit organizations. Members co-operate in organizing a worldwide network of terminology centres and terminology networks. Besides the dissemination of information on their activities they are obliged to enhance the awareness for the importance of terminology within their sphere of action.

Infoterms publishes, besides books on terminological issues, several quarterlies – the Infoterm Newsletter (INL), BiblioTerm (BIT) informing its readership about the latest publications in the field of terminology, StandardTerm (STT) providing up-to-date information on standardization in the field of terminology, including standardized guidelines for elaborating terminologies, and Terminology Standardization and Harmonization (TSH), a joint publication of the ISO/TC 37 Secretariat and Infoterm.

Special relations

Infoterm maintains special relations with a number of institutions and organizations, which support the aims and objectives of Infoterm:

- Austrian National Commission for UNESCO
- Austrian Federal Ministry for Economy and Labour (BMWA)
- Austrian Standards Institute (ON)
- CEN/ISSS Electronic Commerce Workshop (CEN/ISSS/WS/EC)
- East Asia Forum on Terminology (EAFTerm)
- International Organization for Standardization (ISO)
- World Health Organization (WHO).

Infoterm takes care of the secretariat of ISO/TC 37 “Terminology and other language resources” on behalf of the Austrian Standards Institute. The scope of ISO/TC 37 is defined as “Standardization of principles, methods and applications relating to terminology and other language resources.”

Infoterm Executive Board:

President: Klaus-Dirk Schmitz, Deutsches Informations- und Dokumentationszentrum für Terminologie (DEUTERM)

Vice-President: Håvard Hjulstad

Vice-President: Laurent Romary, Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)

Honorary Vice-President: Prof. Yunqi Jiao, China National Institute of Standardization (CNIS)

Please note that Infoterm is a network which co-ordinates terminology-related activities and publishes relevant literature, but doesn't collect or own terminology resources itself.

6.1.3 *TERMIUM*

Prefatory remark

The description of *TERMIUM* contained in the present report is extracted from the information provided on the home page http://www.termium.gc.ca/site/accueil_home_e.html, (last update 2005-03-18).

Background:

Translation between English and French being requested by governmental and federal bodies has a long tradition in Canada because of the fact that the country has two official languages. The Public Works and Government Services of Canada comprises therefore a Translation Bureau (funded in 1934) providing high-level translation and interpretation between these languages. Since 1987 also the development of terminology products and services, and terminology standardisation within the federal government are included into the responsibilities of the Translation Bureau. The standardisation work is carried out in the Terminology Standardisation Directorate (TSD) and through the Federal Terminology Directorate. The organisation is also internationally acknowledged within the field of terminology research and standardisation.

One of the objectives of the Translation Bureau and the TSD was to establish and improve the methodology used to carry out terminology research and standardization work and to develop the *TERMIUM*® data bank to store, manage and distribute the results of this work. Thus, *TERMIUM* builds upon the professional experience and know-how of translators and interpreters, and it is developed in co-operation with terminologists and technical professionals.

Today, *TERMIUM* is one of the largest termbanks of the world and is maintained continuously (approx. 100,000 modifications per year).

The content of the database is accessible to translators, technical writers and other professionals. Several spin-off products are also developed, such as an on-line linguistic tool the *TERMIUM Plus*® which is built on the top of the termbank being provided with writing assistance facilities in English and French and giving access to 13 electronic language resources.

The main content features of *TERMIUM* are:

- 3,500,000 terms and names in English and French
- standardized English and French terminology
- 100,000 terms and names in Spanish
- information types: synonyms, acronyms, abbreviations, definitions, contexts, phraseology units, examples of usage and observations
- subject fields: “almost every field of human endeavour is covered”

The experience related to working methodology is gained from the development of *TERMIUM* and standardisation work; it is formulated in a comprehensive tutorial, the Pavel Terminology Tutorial which is on-line and freely accessible through the internet at the address: http://www.termium.gc.ca/didacticiel_tutorial/english/lesson1/index_e.html.

The mission of this tutorial is to offer a self-study course sharing the knowledge acquired in translation and terminology standardisation processes. The guidelines and recommendations for terminology work given this tutorial are based on the comprehensive experience obtained from the variety of practical working tasks concurrent with expertise in terminology standardisation.

Therefore, the tutorial can be considered a documentation of best practices recognised in the development of the TERMIUM data bank.

6.1.4 TSK

The **Finnish Terminology Centre TSK** (Sanastokeskus TSK) offers information and expert services related to special language terminology, vocabularies and terminology work. TSK's main activities are terminology projects, termbank activities and term service.

Since 1974, TSK has been the only national terminology centre with the responsibility of co-ordinating all terminological activities, taking care of the special language planning in Finland in co-operation with Kotimaisten kielten tutkimuskeskus (The Research Institute for the Languages of Finland) and producing mono- and multilingual vocabularies and terminological databases in co-operation with subject field specialists.

TSK is a member of the **Nordterm** association (cf. section 6.1.2) and has a representative in its steering committee. TSK has participated in the planning and organization of Nordterm's courses and seminars and in the production of the Nordterm publications which deal with terminological principles and methods used by the Nordic terminology centres and institutes.

TSK is an independent, non-profit registered association free of all kinds of financial, political or other commitments. This means that TSK offers terminological services to all kinds of customers, whether in public or private sectors, and that the needs of the customers are equally tended to. According to the statutes of the association, TSK concentrates on offering information services for public purposes and not to making a profit. The member organizations of TSK have the right to decide upon the statutory matters of the association in two annual general meetings.

- *TSK is a producer of vocabularies and technical dictionaries*
One of the main activities of TSK is to produce vocabularies and technical dictionaries for various subject fields. These publications are produced in special terminology projects where a terminologist and a group of specialists work together. The methods used by these project groups are always based on systematic terminological principles and co-operation between subject field specialists and terminologists.
List of TSK's publications can be found on TSK's web site, <http://www.tsk.fi>.
- *TSK is a producer of quality on-demand term services*
Among the general public, TSK is probably best known for its on-demand term service, the only language service in Finland solving terminological problems of special languages.
- *TSK is developer and user of principles and methods of systematic terminology work* (cf. section 6.5.2.6)

As the official Finnish member body of ISO/TC 37, TSK has actively participated in the production of the international standards on terminology work.
(Terminology - Principles and Coordination, <http://www.infoterm.info/index.php> or <http://www.iso.org/iso/en/ISOOnline.frontpage>)

The research and development of terminological principles and methods are one of the most important activities of TSK. By teaching these principles and methods to subject field specialists and to language specialists, TSK contributes to the quality of all terminology work done in Finland.

TSK organizes courses in practical terminology work and in terminology project management for all those interested in terminological issues.

- *TSK is a developer and user of terminological databases*
TSK has several years of experience in developing methods and techniques for terminological databases and termbanks. TSK's own termbank TEPA, set up in 1985, is a multilingual database of technical terminology: in addition to Finnish terms and definitions it contains term equivalents in several languages of which English and Swedish are the most common ones. It contains now ca. 100,000 term entries, thus being one of the largest public termbanks in the world. For in-house use TSK has a termbank containing ca. 250 000 term entries.

6.1.5 Polish terminology resources

6.1.5.1 The 'PKN' Polish Standardization Committee

Unified Terminology Bank 'BTZ'. The Bank contains about 77 thousand records concerning terms extracted from Polish standards. The 'BTZ' Bank is held in ISIS program and includes the following information in their records: term, definition, remarks, synonyms, foreign language equivalents, etc. Recently, the PKN Committee has been involved in restructuring their data base through changing their software system into Integrated Computational System 'ZSI' which will take some time before making their data base available to the internet users.

Terminology collections published by the PKN Terminology Bank Committee. There has also been published by the PKN Committee a series of three publications containing collections of terms extracted from respective standards:

Standardized vocabulary – work safety and anti-fire technique is a publication prepared on the basis of terminological data contained in BTZ, maintained by the PKN Centre of Standardization Information. The publication includes the data obtained from Polish terminology standards covering work safety, the standards being, among others: PN-EN 132:1993, PN-EN 165:2000, PN-EN 614-1:1999, PN-EN 60743:2000, PN-ISO 4880:1999, PN-ISO 8421-1:1997, PN-ISO 8421-1/AK:1997, PN-ISO 8421-2:1997, PN-ISO 8421-3:1996, PN-ISO 8421-4:1998, PN-ISO 8421-5:1997, PN-ISO 8421-6:1997, PN-ISO 8421-8:1998;

Standardized vocabulary – information technology is a publication prepared on the basis of terminological data contained in BTZ, maintained by the PKN Centre of Standardization Information. The publication encompasses the data obtained from the Polish version of multi-sheet ISO/IEC 2382 Standard and the sheet standard PN-.../T-01016;

Standardized vocabulary – environmental protection is a publication prepared on the basis of terminological data contained in BTZ, maintained by the PKN Centre of Standardization Information. The publication encompasses the data obtained from Polish terminology standards covering environmental protection and related subject fields, last updated on September 31, 2000. It includes data obtained from the following standards: PN-ISO 4225:1999, PN-ISO 4225/Ak:1999, PN-ISO 6107-1:1994, PN-85/01705, PN-B-12079:1997, PN-74/J-01003.06, PN-73/M-52002.01, PN-R-04152:1997, PN-V-01007:2000, PN-Z-15010:1999.

International Thesaurus of Technology TIT. Besides, some thematic terminology collections are being prepared, such as ‘International Thesaurus of Technology TIT’ with descriptors: Pol, Eng, Fr; ascriptors: Pol, and others.

6.1.5.2 UKIE terminology data base

The UKIE terminology data base contains terminology collected while translating EU legal instruments into the Polish language. The data base contains, inter alia, legal terms extracted from Founding Treaties, Accession Treaties, and European Commission glossaries, such as: ‘Glossary of terms relating to the internal market’, ‘Treaty on European Union – vocabularium’, ‘Economy, finance, money’, ‘Regional policy’, ‘Community Customs Code – vocabularium’ and the internal publications of the Translation Department such as ‘Regular Report – vocabularium’, and ‘Europe Agreement – vocabularium’. All terms are in English and Polish, some of them have also French and German equivalents. The total number of records has reached 8000.

UKIE Terminology collections

Terminology work done in the Translation Department also covers preparation and issuance of terminology collections. Glossaries prepared at the European Commissions constitute a part of those collections. These glossaries have, as at the Commission, the status of working documents, and as such may only be disseminated gratuitously among state officials and education institutions. The list of the glossaries issued so far can also be found on websites of the Department.

The purpose of the glossaries is to promote the application of uniform terminology, they are of help to, inter alia, lawyers, employees of central and local government administration dealing with the European issues, and to translators working on translation of EU legislation into Polish.

Below we enumerate UKIE Translation Department’s published four-language glossaries:

- Glossary of the Community Customs Law Terms (CCC);
- Glossary of the Treaty on European Union Terms (TUE);
- Glossary of the European Agreement Terms (EA);
- Glossary of the Terms Relating to Internal Market (INT);
- Glossary: Economics – Finance – Money (EKOFIN);
- Glossary: Regional Policy (REG);
- Glossary: Regional Report of the European Commission (RAPORT)

6.1.5.3 PolTerm as a computerized collection of LSP-corpus-based terminology

PolTerm is stored in a computerized terminology processing system, which presupposes computer-assisted terminology management, i.e. compilation, storage and retrieval of terminology.

PolTerm is a bilingual LSP-corpus-based terminology collection. The LSP electronic corpus contains 42 consolidated texts of Polish legislative Acts and their English and German translations. Some more important Acts: the Civil Code Act, the Code of Commercial Partnerships and

Companies Act, the Freedom of Economic Activity Act, the Natural Persons' Income Tax Act, the Legal Persons' Income Tax Act, the Excise Duty Act, the Banking Law Act.

The PolTerm database consists mainly of one- to several-noun terms covering various domains of law and included into two main subject-fields: tax law and business law, as the reflection of the lexicon contained in the corpus, the current number of entries amounting to 10,500 and covering a wide range of branches of law. These broad subject-fields are further specified with regard to every term according to approved codes of 42 texts they refer to. Thus, every term in PolTerm has its own identification code. The same holds also for nomenclatures and phrases, mainly of verb-noun (-complement) type, which make up the rest of the database.

Status of the PolTerm terminology

The Polish term in each terminological record is presumed to be standardized in that the system of concepts (conceptual structure) underlying the legislative lexicon is presumed to be unified or consistent in other words, and thus their definitions within that system, even if covert, are presumed to ensure standardization of reference, i.e. the situation in which one concept is represented by one term throughout the entire system.

Thus the basic principle of secondary term formation in PolTerm is the preservation of the concept system consistency requirement so that one concept is expressed by one particular English term throughout the whole secondary terminology system (i.e. that the term is monosemic).

PolTerm as a terminological resource

PolTerm term collection is used in the in-house Polish-English legal translation by TEPIS terminologists headed by Danuta Kierzkowska and Tomasz Borkowski and by cooperating translators, individual legal translators and TEPIS Society members, including several hundred English-language translators, who have free internet access to selected subset of vocabulary from PolTerm labeled "basic" and containing almost 900 bilingual entries including terms and nomenclatures (occasionally phrases) excerpted from fourteen essential Acts (the "basic" project is an on-going activity).

6.1.6 Estonian terminology resources

The portals keeleveeb.edu.ee and www.keelevara.ee contain dictionaries and links to various Estonian mono- and bilingual dictionaries, both general-purpose and specialised. Keeleveeb is free of charge, while keelevara requires registration and payment for using some of its dictionaries.

Termbanks

- Esterm (www.legaltext.ee)
- Names of the birds of the World (www.eoy.ee)
- List of Birds of Europe (www.loodus.ee/eurolinnud)
- Estonian Plant Names (www.ut.ee/taimenimed)
- Lexicon of molecular-biology (www.tymri.ut.ee/sonastik.html)
- Lexicon of bank terminology (www.pangaliit.ee/pangandusinfo/s6nastik)
- Lexicon of rambblers' terminology (www.matk.ee/termin/sonastik.htm)

Networks

Estonian Terminology Association (over 60 individual and collective members) (www.eter.ee)

6.1.7 Lithuanian language term base

The Lithuanian language term base (Lietuvių kalbos terminynas) is created for terminology standardization and usage purposes (research work, terminology creation, planning and standardization, compiling of terminological dictionaries, vocabularies and glossaries). The term base is a source of information for linguists, translators and editors.

The Lithuanian language term base is created by the Laboratory for Computer Application in Research of the Institute of Mathematics and Informatics under the auspices of the State Commission of the Lithuanian language and consulting by the Institute of the Lithuanian language.

The term base is managed by using a Microsoft SQL Server 2000 database management system, and it is accessible on Internet for registered users. There are two types of users: ordinary users and special users. Ordinary users have limited access to data units and can get up to 100 data units within one request. Special users can get up to 500 data units within one request and have data a retention possibility.

In order to make use of the Lithuanian language term base by computer, it is necessary to open the website of the Institute of Mathematics and Informatics on the Internet at the address www.terminynas.lt , to have several files sent and implement them in computer.

6.1.8 Latvian terminology databases

In Latvia there are 3 important terminology databases. They all provide full access free of charge:

- www.termnet.lv
- completedb.ttc.lv
- www.termini.lv.

1. **TermNet** is hosted by Tilde in cooperation with Academy of Sciences of Latvia and is sistem-oriented. TermNet provides access to some 145.000 terms in different domains. The Latvian language terminology dictionaries have been put to an internet website in 2002. It was developed and is maintained by Tilde. It is also a portal where all new official terms get posted, and users can post comments on them. Many entries contain terms in up to 4 languages and definition. TermNet is hosted on a MS SQL Server 2000.

2. **CompleteDB.ttc.lv** is is hosted by the Translation and Terminology Centre. It was started by digitalizing terms created by Terminology Commission of Latvian Academy of Sciences (TC of LAS) in last 10 years as well as adding terms created in EU and NATO materials translation process. Many entries contain terms in up to 4 languages and/or usage examples (context). Database is continously updated with in-house created terms and several thousand terms approved by TC of LAS. Mostly used for translation work.

3. **Termini.lv** is hosted by TehnoMedia and provides IT&T and physics terms. It's mostly interesting as an advanced term search engine capable to recognize similar words, misspelled words and words derived from the same root. Created in 2002 using PHP and MySQL.

6.1.9 SAP

In this subsection, some experience acquired at Morphologic, Hungary concerning the use of SAP is presented.

SAP has a one-stop terminology interface integrated into its SAP R/3 system (transaction SE63 - Translation Environment), which, according to some translators at proz.com, is far from complete.

Similarly to other SAP products, the terminology tool also offers a wide range of user privilege management features, and terms can have different statuses of approval until they reach the status of approved.

Some SAP translators – especially the localisers – get access to this software and can suggest new terms, others are provided with bilingual MultiTerm glossaries to get help in their work. These glossaries are also available for sale for a sum of 110 euro, and contain entries in Bulgarian, Croatian, Czech, Danish, Dutch, English, Finnish, French, German, Hungarian, Italian, Norwegian, Polish, Portuguese, Russian, Slovakian, Slovenian, Spanish, Swedish and Turkish. (See SAP Note 565591)

So far, the database contains about 650,000 terminology entries in 20 European languages, and nearly 16,000 definitions of SAP concepts. SAP made public some of these definitions in English and German at

http://help.sap.com/saphelp_glossary/en/index.htm

http://help.sap.com/saphelp_glossary/de/index.htm

At SAP, the privileged languages are English and German, and all terms need to have equivalents in all languages.

Terminological entries are usually created by knowledge brokers and authors of texts as well in English or German. Entries include not only software-related entries (screen captions, etc.) but also entries appearing in training course materials and marketing materials. In Hungary, translators can also enter new terms, but superusers – consultants in Hungary – also need to approve them.

Entries include a wide range of information, including definition and part-of-speech information, but the emphasis falls on the source of the term.

At SAP, therefore, all users regardless of their nationality use the same terminology database, there are no competing databases.

6.1.10 Microsoft

The Microsoft company employs a particular term registration process, which is briefly described by Morphologic below.

Microsoft follows a systematic approach to software-encoded terminology, which starts during development time. Developers create terminology during program design and development in an intuitive/metaphoric way. Important terms – e.g. brand names, major technology names - are also checked by other personnel, sometimes even tested in a public opinion poll. The language of all source terms is English. The creation of the initial termbase is automatic: their own localisation software extracts all the string resources from the products.

Microsoft does not have real termbases, in the sense that definition is not an integral part of their terminology. They use terms to provide a consistent localisation to their products.

Microsoft employs a few (1-3) terminologists for every language they provide a product version for. These terminologists create a core termbase for each and every product, building on the terminology of former products and user responses too. The core termbase is then sent out to localisers, who have to create local versions of their products - and their terminology.

Most of the terms employed appear in screen captions, and termbases (in the form of source string – target string) are unique for each and every product and product version, and contain the screen captions and some help-specific terms. Non-software-related terms (which are only a few in number) are not collected in a single termbase, but Microsoft Press, the official publishing house of Microsoft, regularly updates its Microsoft Press Dictionary.

All string-termbases are available to developers for free at the following URL: <ftp://ftp.microsoft.com/developr/msdn/newup/Glossary>

The core termbases the work is based on are not published.

6.2 Projects

6.2.1 *INTERA*

INTERA (start Jan. 2003, duration 2 years) was a European project with two major objectives:

- to build an integrated European language resource area by connecting international, national and regional data centres,
- to produce new multilingual language resources.

The first goal involves the integration of a critical mass of different types of language resources with the help of metadata descriptions and the interlinking of the resulting distributed resource repository with an existing tool repository thus enabling users to directly start suitable tools on the included resources. INTERA anticipates that this integrated and interlinked metadata description domain will facilitate the access to language resources in Europe and help professionals in industry, the eContent business, research and education, and increase the usage of the resources already available.

The second goal addresses the lack of quality of multilingual resources, especially for the less widely spoken languages, including the Balkan ones, which are of crucial importance to the development of the eContent business. INTERA goes further ahead by developing exemplary methods for their business attractive production.

It aims to:

- deliver an integrated searchable and browsable European domain of metadata descriptions of language resources housed by various data centres
- deliver mechanisms that make it possible to immediately start a relevant tool after the user has found an interesting resource by interlinking the integrated resource repository with tool repositories
- make all elements and controlled vocabularies used compliant with current standards and integrate them in open terminology and RDF repositories
- localize the tools and descriptions to a large variety of languages
- produce multilingual resources for less widely spoken languages in order to improve the eContent situation for them
- develop a model which will help resource producers in creating such multilingual resources in an easier and more profitable way

- inform a large group of language resource users of the potential of the new mechanisms and resources, and to stimulate their use.

Of special interest to the EuroTermBank project are the achievements in the multilingual language resource production where four parallel corpora have been created from which extraction of multilingual terms have taken place, and the procedure developed on that subject.

The corpora created are

- Greek - English parallel corpora (4 MWs (Million words) in total, 2 MWs per language)
- Slovene - English parallel corpora (4 MWs in total, 2 MWs per language)
- Serbian - English parallel corpora (2 MWs in total, 1 MWs per language)
- Bulgarian - English parallel corpora (2 MWs in total, 1 MWs per language)

And the following domains were subject to term extraction:

- domain of law : Bulgarian, English, Greek, Serbian, Slovene
- domain of education : Bulgarian, English, Greek, Serbian
- domain of health : English, Greek, Serbian
- domain of tourism : English, Greek
- domain of environment : English, Greek.

The numbers of terminological entries are rather low, between 200 and 9000 entries per domain per language, but the procedures for corpora creation and term extraction are well described.

More information on the Intera project can be found at www.elda.org/rubrique22.html.

6.2.2 SALT

In the SALT (Standards-based Access to multilingual Lexicons and Terminologies) project a consortium of academic, government, association, and commercial groups in the United States and Europe worked together on the task of testing, refining and implementing a universal putting-together format for the interchange of terminology databases and machine translation lexicons.

The project responded to the fact that many organizations in the localization industry are using both human translation enhanced by productivity tools and machine translation (MT) with or without human post-editing. This duality of translation modes brings with it the need to integrate existing resources in the form of (a) the natural language processing (NLP) lexicons used in machine translation (categorized as *lexbases*) and (b) the concept-oriented terminology databases used in human-translation productivity tools (called *termbases*). The idea was to facilitate by this integration the consistency among various translation activities and to leverage data from expensive information sources for both the 'lex' side and the term side of language processing. The title of the project reflects the intention to provide a Standards-based Access service to multilingual NLP-Lexicon and human-oriented Terminology resources via an Internet website.

The SALT project combined two already finalized interchange formats: OLIF (Open Lexicon Interchange Format), which focuses on the interchange of data among lexbase resources from various machine translation systems, and MARTIF (ISO 12200:1999, MACHine-Readable Terminology Interchange Format), which facilitates the interchange of termbase resources with

conceptual data models ranging from simple to sophisticated. The goal of SALT was to integrate lexbase and termbase resources into a new kind of database, a lex/term-base called XLT (eXchange format for Lex/Term-data). XLT is based on XML (Xtensible Markup Language), which is a data format for structured document interchange on the Web and is under development by the World Wide Web Consortium.

Thus, XLT is an XML-compliant framework for defining a family of closely related term data exchange formats tailored to specific user groups. MARTIF is an SGML application that has been adapted to the XML world in anticipation of the adoption of an XML-Schema standard and has become the heart of XLT. The essence of OLIF, which is a tagged, but not SGML, application has been integrated into XLT by inserting the OLIF header into the XLT header, merging the OLIF Central Entry into the corresponding element of XLT taken from MARTIF, and adding to each XLT concept entry an optional NLP feature-value pair list that corresponds exactly the feature-value pair list of OLIF but is recast in MARTIF-style XML. In addition, the TMX method of documenting user-defined Unicode characters and the TMX meta-markup method of including presentational markup in running text (for contextual examples, etc) have been incorporated into XLT.

The SALT project was an open-source project creating open standards. Some of the results of the SALT project have been turned into ISO standards or have been integrated into revised ISO standards, and ISO IPR policies apply to these. Control of TBX has been handed over from the SALT project (by the European Commission as its legal representative) to the Localization Industry Standards Association (LISA) and its OSCAR (Open Standards for Container/Content Allowing Re-use) Special Interest Group. All work carried out by the SALT project was explicitly royalty free and all IPR donations to the SALT project were made under a royalty free license arrangement.

The overall goal of SALT was extremely practical. It was to reach “critical mass” with XLT so that tool developers, such as Star, Trados, EP, Logos, Systran, L&H, and Xerox, would incorporate some level of XLT support in their products so that various companies would provide on-going consulting services to anyone who wants to get their proprietary lex/term-data into XLT format or XLT data into their proprietary format.

Partners in the project were the Institut für Übersetzer- und Dolmetscherausbildung of the University of Vienna (Gerhard Budin) as project coordinator, the Institute for Information Management, University of Applied Sciences Cologne (IIM, Klaus-Dirk Schmitz), the Accademia Europea di Bolzano per la ricerca applicata e la formazione post-universitaria (Bruno Ciola), the University of Surrey (Khurshid Ahmad, Lee Gillam), the Laboratoire lorrain de recherche en informatique et ses applications (LORIA, Laurent Romary), the Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e. V. an der Universität des Saarlandes (Jörg Schütz) in Europe, as well as the Brigham Young University Translation Research Group (Alan K. Melby) and the Kent State University Institute for Applied Linguistics (Sue Ellen Wright) in the United States.

The project was co-funded within the EU Fifth Framework Programme during the period from January 2000 to December 2001.

6.2.3 ENABLER

The Enabler (European National Activities for Basic Language Resources) Thematic Network (EC project, started Dec. 2001) aims at improving cooperation among national activities established by national authorities for providing Language Resources (LRs) for their languages. The action aims at: establishing a regular exchange of information; identifying and fostering possible synergies and cooperation; promoting the compatibility and interoperability of their results, thus facilitating the successful transfer of technologies and tools among languages and the construction of multilingual LRs; increasing the visibility and the strategic impact of those national activities in the field of HLT; contributing to the creation of an overall framework in which the public and private sectors, national efforts and international coordination could cooperate in order to answer the IST need for LRs.

The main results within the Enabler Network were

- A survey of language resources (LRs), providing a global overview of National projects and activities on LRs of all kinds. It relates to 164 different resources from various countries and languages and concerns all the facets of LRs. Both the point of view of the LR producers and of the prospective users were taken into account.
- With the aim of optimizing the process of production and sharing of (multilingual) LRs, the Network promoted the compatibility and interoperability of LRs through cooperative work with projects, committees and communities in the different fields of LRs.
- Collection of validation methodologies of LRs representing current best practice in the area.
- A description of the industrial needs of LRs, with the aim of easing the exploitation of existing LRs and collecting ideas for future LRs. This was achieved by mapping the types of
- A number of initiatives with the objective of promoting LR production and management in the years to come, improving infrastructure and coordination activities for LRs.
 - The BLARK (Basic LAnguage Resource Kit) concept has been adopted and supported defining a further level The Extended Language Resource Kit (ELARK).
 - The promotion of the launch of a large initiative comprising the major LRs and HLT groups in Europe and world-wide for the creation of an open and distributed infrastructure for LRs.
 - Contribution to the design of an overall coordination and strategy in the field of LR. A new committee has been established in the field of Written LRs, the *International Coordination Committee for Written LRs and Evaluation (ICCWLRE)* continuing the Enabler mission but enlarging the scope beyond the European boundaries. This committee will cooperate with the COCOSDA (*International Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques*).

6.2.4 PHARE

The PHARE project on Terminological Dictionaries for Russian-Language Basic Schools is carried out at the University of Tartu, August 2004 – December 2005. The aim of the project is to create

basic terminological dictionaries for Russian-language basic schools of Estonia (learning years 7-9). There will be 12 dictionaries, and they will contain altogether 15,000 concepts. The dictionaries will be in Estonian, with the term translations in Russian.

6.3 Tools

6.3.1 Terminology management tools

Terminology management tools are computer programmes conceived especially for the management – the recording, processing, saving, and using – of linguistic data and their application in the area of technical writing, translation and terminology work. Most of them have a database-like structure less extensive than “real” databases and allow only the comparatively easy operations and commands necessary for terminology management.

While at the beginning of computer-aided terminology management most of the programmes were conceived with regard to a special language pair using the term-oriented or lexicographical approach, nowadays the most current programmes are designed following the concept-oriented approach required by terminology experts and determined in relevant ISO standards on terminology work. In these systems multilingual terminology data can be easily managed, each language serving as query or target language, according to the present needs. All the information on one concept – one or several terms either in one or in several languages, documented by further data categories, like definitions, contexts, grammatical information, etc. – is stored in one entry. Thus, these tools respond not only to bi- and multilingual translation necessities, but also to the requirements of monolingual terminology work in the area of language planning, terminology standardisation, technical writing and the use by experts of the respective subject field.

Furthermore, terminology management tools differ in their way of data structuring and modelling. In systems with a predefined entry structure the user has only few possibilities to adapt the database to his specific needs, whereas other systems allow – or force – him to define his own entry structure. Through the database definitions set up by the user himself before documenting the terminology, he is able to design a database tailored to his requirements and containing exactly the data categories he needs. In doing so, he can, if desired, manage other types of data apart from proper terminology, e. g. bibliographic information for the documentation of the references of the terminological data. Terminology management tools are equipped with various search, filter, import and export functions as well as, in some cases, the possibility to assign different kind of user and administrator rights.

In a modern terminology work and translation environment, an interaction between terminology management and other working tools, like word processing or DTP software, or translation memories, is often requested. Many of the terminology management tools are therefore integrated into a computer-aided translation environment consisting of translation memory, align tool, tag editor and, in some cases, term extracting tool, and provide interfaces to standard word processing and DTP software.

Some terminology management tools are not only available as stand-alone solutions, but also as multi-user systems, where different users can access the same terminology data stored in a database and work simultaneously with it; they are applied in translation and localisation companies

cooperating with freelance translators or integrated in the workflow of terminology creating and translating in IT and industrial companies.

During the last years, the current tendency goes towards interactive online terminology management. The new internet-based terminology management tools have a client-server-architecture allowing a functional division between the server performing the proper data base functions on a central computer system, and the client responsible for the interaction with the user on his own PC. Thus, terminologists, translators and experts scattered all over the world can work simultaneously with one terminology data base.

The most current terminology management tools are the following (please note that some of them are only available as an integrated part of a translation memory tool):

- across (Ahead Software AG, Germany)
- CATS - Computer-Aided Terminology System (CATS, Germany)
- GFT DataTerm (GFT GmbH, Germany)
- Lingo (Julia Emily Software, France)
- LingTools (Sietec Systemtechnik, Germany)
- LogiTerm (Terminotix Inc., Canada)
- MoBiDic (MorphoLogic, Hungary)
- MTX™ (LinguaTech, USA)
- MultiTerm (TRADOS GmbH, Germany)
- SDL TermBase (SDL, UK)
- TermStar (STAR AG, Switzerland)
- Termwatch (ATRIL Software SL, Spain)
- UniTerm (Acolada GmbH, Germany)
- Xerox Terminology Suite (XTS) (Xerox Multilingual Knowledge Management Solutions, France)

6.3.2 Term extraction tools

Term extraction tools may be used to help setting up terminology. Term extraction tools typically provide a list of potential terms, “term candidates”, from a corpus or from a text. The term candidates will have to be validated by a human user, so term extraction can be computer-supported, but will not be fully automatic.

A term extraction tool may be used when a term base for new domain has to be developed: First a corpus for the domain has to be collected, and then the tool will provide a list of potential terms. Similarly, term extraction may be used as one of the preparatory steps in a translation workflow: When a new translation job is started, it may be useful to go through the text and extract all new terms and make sure they are entered in the term base.

Term extraction can be either monolingual or bilingual. Monolingual term extraction provides a list of term candidates from a selected corpus or from a text as mentioned above. Bilingual term extraction will work on parallel texts, i.e. source texts with their target translation, and will identify the term candidates in the source text and their equivalents in the target text. This means that the first step, identification of the terms in the source text, is the same as for monolingual term extraction.

Below we will first present some basic approaches to obtaining the functionality of a term extraction tool, then we will provide some comments on how quality can be measured. Finally, we will mention a few examples of term extraction tools.

6.3.2.1 Basic approaches to automatic term extraction

There are two main approaches to automatic term extraction: *linguistic* and *statistical*.

Linguistic approaches make use of morphologic, syntactic or semantic information. They typically attempt to identify word combinations that match certain part-of-speech patterns, e.g. “adjective + noun”, “noun + ‘de’ + noun”, “noun + noun”, “noun + noun + noun”. So what is involved is simple, or shallow, analysis of the text. Obviously, in order to be able to recognise the nouns etc, a dictionary is needed, and the rules for word formation for terms will be language specific too. A list of stop words, i.e. words that cannot appear as part of a term candidate may be used to avoid some of the mistakes.

Statistical approaches basically attempt at identifying lexical items or combinations of lexical items that occur with a frequency higher than normal in the corpus. A statistical approach can obviously only be used if a reasonably large corpus is available, and will not work very well for a single text. The advantage of a statistical approach is that it is language independent; however, it also has disadvantages: a purely statistical approach is normally not very satisfactory, i.e. it does not find all the terms and/or it suggests too many candidates that are not terms.

Often the best solution is found by combining the two approaches in a hybrid solution, e.g. a statistical approach followed by a linguistic filtering.

6.3.2.2 Evaluation of automatic term extraction

Recall and *precision* are the standard evaluation parameters. The *recall* describes the ability to recognise all terms contained in a text or a collection of texts. It is defined as the number of correctly recognised terms divided by the total number of terms in the text. If the recall is 80%, this means that 20% of the terms have not been detected.

The *precision* describes the accuracy with which words and phrases are classified as terms. If too many candidates are suggested which turn out not to be terms, then the precision is low. The precision is defined as the number of correctly identified term candidates, divided by the total number of term candidates. If the precision is 80%, this means that 20% of the term candidates suggested are not terms.

High values for recall normally imply low precision and vice versa, so e.g. if it is very important to get a high recall, then the terminologist will have to delete many wrongly suggested terms.

6.3.2.3 Some examples of term extraction tools

Below we mention a few examples of term extraction tools. The list is by no means exhaustive and is meant only as examples.

1. TRADOS Term Extract. It can provide both monolingual and bilingual term extraction. The bilingual part requires access to a bilingual TM. Supports all European languages (Unicode). <http://www.translationzone.com/product.asp?ID=100>
2. Compendium Terminologist. This is a package that provides both monolingual and bilingual term extraction. The bilingual part requires access to a bilingual TM. Currently supporting English, German, French, Spanish. <http://www.compendium.com/jahia/Jahia/site/lingua/lang/en/pid/448>
3. SDLPhraseFinder. This package provides bilingual term extraction. Requires access to parallel texts. <http://www.sdl.com/products-translation/products/sdlphrasefinder-desktop.htm>
4. Xplanation (www.xplanation.com) offer a service of term extraction, but they seem not to sell the tool. They just opened offices in Vienna in order to be close to the new markets.

6.3.2.4 Conclusions on term extraction tools

A term extraction tool enhances productivity, but one cannot expect it to find all and only the terms.

6.3.3 *Other tools*

6.3.3.1 Search and replace tool: xPlace

A particular problem of terminology translation and management can be solved by using the xPlace terminology search and replace tool. As knowledge should be made available in Hungarian in a very short time after publication, translators have to meet tight translation deadlines. Also, many companies do not prepare for multilingual terminology management, and think it's the job of the translator to find out what terms they would like to see. In other cases, like the translation of the *acquis communautaire*, the terminology of any text needs to be harmonised with terminologies of other texts, and authoritative sources of terminology can sometimes appear after the term was used in other texts. Therefore, we need to provide a way to track and quickly change terminology in texts.

In languages with a relatively simple morphology, this is no problem as regular search/replace functionality makes the lion's share of the job, but in languages like Hungarian, where one word can take a myriad of forms, it is often very hard to find all word forms, because the common part of the stem and its inflected form can be very short and can appear in other words too. Once it is found, one also has to make the phonological agreements between the new stem and the suffixes. Therefore, Kilgray has developed a language-enabled search and replace tool called xPlace, which is a tool that can update documents to use the new terminology in a relatively short amount of time. xPlace is available at <http://www.kilgray.com>. With the use of xPlace, such search/replace operations usually take less than 5% of the time spent on search/replace before.

6.3.3.2 Domain management in MemoQ / CATS, Kilgray's translation environment

The domain is one of the core concepts of MemoQ, distinguishing it from its competitors. What's a domain?

The domain carries information concerning the categorisation of any text, including terms. All domains are interpreted as subtree to a tree hierarchy. The number of domains is not restricted, because the categorisation aspects of a text are not restricted either. Possible domains include: topic, organisation, client, style/register (letter, specification, short story, novel, etc.). We need many possibilities because on the one hand, uniformity is good for translation, while on the other hand we must provide the translator with the possibility of using idiomatic language – because customers like to get what they expect and require. E.g. the letter of Linus Thorvalds, written to the head of the European Commission's DG Informatique should not sound a technical description, should not contain Windows terminology, should contain Commission terminology instead of Parliament terminology, and host should not mean a person. This approach may seem overcomplicated, but as long as it does not appear so complex to the user, it's useful.

The system handles an arbitrary number of trees, and looks for similarity and correspondence between texts/terms and translations stored in the system. The domain of the text must be set during the preparation, but it's possible to mark different segments with different domain information. (Domain information must be stored together with the text, while memory domain information must be stored together with the memory.) The domain does not need to be re-set because the text or term entering the translation memory or the termbase is automatically set to the domain of the given segment.

When the engine finds similarity, it evaluates domain information as well. For practical reasons, the first domain is the „main domain” – we can suppose that the topic or subject area is more important than e.g. register. On the user interface, domain matching is presented with colour codes or in a different manner:

- a. there is an entry in the memory/termbase all the domains of which correspond to the text's domain or form its subtree.
- b. there is an entry in the memory/termbase the main domain of which corresponds to the text's domain or form its subtree.
- c. there is an entry in the memory/termbase at least one of the domains of which (but the main domain) corresponds to one of the text's domains (but the main domain) or form its subtree.
- d. there is an entry in the memory/termbase the domains of which do not correspond to the text's domain, but stringwise there is match.

The complex domain handling thus only means four colours for the translator, thus the user understands how much of the text is almost surely correct and how much needs to be carefully revised. Domain trespassability must be ensured if we want to provide efficient collaboration. Although everyone can define own domain structures (or adopt our suggestion), there's a need for a transfer format, thus everyone working in workgroups must map their domains to the ISO standard specification.

7 Survey of best practices in terminology methodology and workflow

7.1 Overall workflow of terminology tasks

Terminology, although having its own right, cannot be considered an independent discipline of work because it is closely related to translation, knowledge engineering, computational linguistics, lexicology, etc. On the other hand, terminology work carried out in various environments and for different purposes shares a number of essential methodical aspects and practices. This section deals with examples of work environments wherein practical terminology work is usually carried out and it provides an outline of the overall workflow and the most important tasks.

Terminology work can be requested in various environments, such as

- Monolingual environments: e.g. preparation of various kinds of documentation (technical, legal, economic, etc.), harmonisation of company language, etc.
Involved experts are e.g. domain experts, technical writers, knowledge base developers, etc.
- Bi- or multilingual environments: e.g. translation of (technical, legal, economic, etc.) documentation, domain language teaching, building of terminology bases for translators, term standardization, communication between company branches, etc.
Involved experts are e.g. translators, domain experts, linguists, domain language learners, viz. technicians, etc.

Depending on the organisation, body or institution where the terminology work is carried out and on the purpose of the work, the workflow can be organised differently. In each case, some basic questions arise.

Terminology work is very often requested by translators of domain specific documentation, thus this environment can be taken as a practical model scenario for terminology work. Details of the **workflow from translation to terminology** are for instance described by the following information:

- Where in the workflow arises the need for a term
- Who puts the request forward to whom
- Who deals with the request
- Which types of tools are at the translator's disposal

Details of the **terminology work** comprise tasks as

- What types of terminology work are carried out
- What kinds of experts are involved (e.g. a domain expert, company terminologist, etc.)
- What are the steps (e.g. search for terminology in reference documents, term bases, term entry creation, reviews, etc.)

Often, a graphic illustration of the workflow e.g. in form of a flowchart is very useful.

7.1.1 Estonian Legal Language Centre – Workflow

The current practice of the *Estonian Legal Language Centre*, as described below, is defined in its in-house document “Guidelines for Terminology research”.

The task of Estonian Legal language Centre is to translate European and Estonian legislation. As a rule, terminology work precedes actual translation of the documents. When a new document has to be translated, it is first passed to a terminologist who identifies the terms, checks the existence of their translations, and if a translation is missing, creates one. The actual translation of the document starts only after the terminologist has finished her work. It is believed that thus the terminology of the translations will be more consistent.

The first thing a terminologist does is to scan the new document. The aim of scanning is to identify the subject field of the document and all the terms that need to be in the database of the Centre.

If a term is in the database already, the terminologist checks whether it is indeed the same term used in the same sense in the new document. If yes, then she adds the new document’s id number to the term’s entry, thus linking the term with the new document. If not, then the terminologist must first clean up the whole nest of closely related terms and meanings.

If the terminologist finds, in the process of scanning, that the text is of a very specialised nature and has to be translated by a specialist of the subject field, she will make a suggestion to find such a translator from outside of the Legal Language Centre.

The results of the terminologist’s work are kept in two places: in a folder of paper documents and in the term database of the Legal Language Centre. The folder contains more information, and it is less structured than the entry of the database. The folder is like a draft of the final term entry of the database. It contains, among other information, the pros and cons of the suggested terms and translations, copies of the source materials, references to experts.

If a translator, during her work, meets a terminology-related translation problem, she may conduct the terminological research herself. Ultimately, however, she will contact the terminologist, who will then check and possibly modify the problematic terms in the database.

The terminologists and translators use Trados Translator's Workbench and MultiTerm in their work.

7.1.2 University of Tartu - Workflow

The current practice of the *University of Tartu*, as described below, is followed in a PHARE project „Terminological Dictionaries for Russian-Language Basic Schools”.

The workflow is characteristic of a project of creating 12 terminological dictionaries (PHARE project „Terminological Dictionaries for Russian-Language Basic Schools”). The dictionaries contain 1000 – 1500 concepts each, approximately 15,000 altogether. The source language terms, definitions, and examples are in Estonian; the term equivalents are in Russian. The dictionaries are aimed at pupils of the age 14 – 16.

The workflow contains the following steps:

1. Dictionary authors (teachers, university staff, i.e. not professional lexicographers or terminologists) identify terms to be included in the dictionaries

2. The term lists are evaluated by external experts of the corresponding subject fields
3. The dictionary authors create definitions and examples, using a concept-based approach (different from the term-based approach of step 1). The terms, initially chosen by the authors, may appear to be synonyms and thus belong to the same concept entry. The authors use a MS Access database, MS Excel spreadsheet or MS Word document to create the entries. Eventually, all the input is converted into a MS Access database
4. The definitions are evaluated by external experts
5. The dictionary databases are converted to XML
6. Grammatical information is automatically added, reference entries are automatically created, and lay-out information is automatically deduced from the XML tags
7. The result is evaluated by external experts.

After feedback from steps 2, 4 and 7, the authors modify their dictionaries, if necessary.

7.1.3 Latvia - Workflow

The main challenge of a language of a small country is to prove that it is able to coexist with other EU languages – to discuss the same problems and to perform the same tasks as they do. Correspondent words in different languages may have different semantics, but by translating them we have to achieve unambiguous meanings. For establishing mutual collaboration on different levels, correspondence between terms in different languages has to be established.

The need for a term mainly arises during:

- research process
- translating of LSP texts
- advisory services.

Everyone has a right to put the request forward to an expert or to an advisory body.

Domain experts are main persons to deal with the request. Main types of tools for translators are term databases, term standards, term dictionaries and the like.

Where do the terms which need coining into Latvian appear from? Where in the workflow arises the need of a term?

The sources of them may be different. In some cases a dictionary or a glossary from a foreign language has to be translated. Laws and regulations from European Union have to be translated into Latvian, and they abound with specific terms and concepts, the equivalents of which need to be created and thoroughly analysed.

Lots of terms have to be adjusted in economic structures. By expanding e-commerce (m-commerce) and i-business, the necessity of providing corresponding terminology is very important. The development of all the branches of ICT with the flow of new terms is very demanding, and the new terms have to be coined rather quickly, as not to allow to appear some English slang at its place.

Further, standards of all branches of science and technologies, which need to be translated into Latvian and have their Latvian equivalents, etc.

Who puts the request forward to whom? Who deals with the request?

The initiators of a term may be different organizations and individuals dealing with all kinds of social, legal, technological, medical, economic activities; translators and education managers,

banking officers, newspaper writers and publication editors, standard developers and even students and front-end users of different kind of devices.

The request for a new term is addressed to a corresponding branch of a subcommission of the TC of LAS, which at its meeting analyses the requested term and its explanations. Explanations of the very term may be found in online dictionaries, and they may have different degree of detail. The task of subcommission members is to select the essential features and to propose the adequate Latvian equivalent of the term. The TC of LAS hosts 26 subcommissions, and everyone of them is responsible for its branch of terminology.

As an example, let us consider the work of Information Technology and Telecommunications (IT&T) subcommission. This subcommission which includes different technology experts is meeting twice a month. The agenda of the meeting comprises a list of English and Latvian terms to be approved and their corresponding explanations in English and Latvian. This list is sent to the subcommission members some days before the meeting, so they may elaborate their concept of the semantics of the new term and propose an adequate Latvian term. The members of this mailing list may exchange their opinions, propose new terms, start discussions and also revert to the previously adopted terms. The main source of terms are terms of developing technologies and ISO standards. Since the middle of 1990's the IT&T subcommission has had 251 meetings.

There are two basic methods for creating and implementing terms: "from the top" (top down) – some regulation defines the very term which has to be used in a specific case, and from the "bottom" (bottom up) – the term is chosen from the vocabulary which is used in everyday life.

By developing different branches of Latvian terminology a combination of both methods is used. The very term is chosen or coined by specialists of the corresponding branch or by translators working in this branch, as almost every field of economic, social, technological, cultural etc. activities is matched to a corresponding terminology commission with the main goal of creating and approving terms of this very branch.

After the harmonization of terms within a particular branch, they have to be harmonized with terms of related branches and with the whole Latvian lexical system. The TC of LAS is responsible for this task. TC of LAS is also the main arbiter to decide about the borrowing or not borrowing English terms which from EU documentation are invading the Latvian language and for creating new Latvian terms if the corresponding term could not be found in the present vocabulary.

TC of LAS every month has to examine, adjust, analyse and approve 400 or 500 terms. Some of them may be created on the spot, some are demanding serious consideration. Nevertheless the avalanche of new emerging terms is a large one, and it is not always easy to keep peace with it.

The main stages of the development of Branch Terminology

In the workflow of the development of branch terminology several stages are distinguished.

Stage 1 – the summarization of branch terms made by specialists of the branch.

At this stage it is important to define the limits of branch terminology and the criterions for term selection and the sources of them.

At present the sequence of operations might be as follows:

- to find out terms used in laws and regulations, and to retrieve terms created in the previous periods of term development (first of all terms which were approved by TC of LAS)
- to find out the international standards of terms and their definitions
- to use foreign term dictionaries, also on-line dictionaries
- to use Latvian dictionaries as a tool for defining more exactly the semantics of the term and for delimitation of concepts.

Stage 2 – the development of a system of branch terms. This task is done by branch specialists in collaboration with a language specialist according to the before mentioned theoretical principles and proposed guideline.

Stage 3 – the approval of branch terms.

For approval of the list of terms a resolution of a meeting of TC of LAS or a resolution of an expert authorized by the TC of LAS is needed.

The terms are approved if they fit in the Latvian term and term creation model system (not contradict to it), and if the specific features of the branch and the needs of international harmonization of terms are taken into account.

The terms approved by TC of LAS are published in a special brochure or dictionary. The most important resolutions of TC of LAS and approved terms, which are intended for a wide use in the society are published in the newspaper “Latvijas Vēstnesis” and in a central newspaper.

The organizing platform for terminology work of TC of LAS are stated by the Statute of TC of LAS.

Taking into account that TC of LAS has a long-term experience in harmonizing the terminology of related subject fields, in cases when terminology work is started in a new subject field, where terminology had never been developed before, it is recommended to establish contacts with TC of LAS to consult and make out the best practice.

Which types of tools are at the translator’s disposal?

Every translator has to work not only with a dictionary and glossary, but also with a lot of databases, which are provided online by the Internet. The term creator is equipped with different types of Internet browsers, providing access to databases of the needed branch.

The task of terminology subcommissions is not only to create proper Latvian terminology of their corresponding branch, but also to host a database of the very domain of terms which their experts have approved and are responsible for. To unify the way of creating these databases, to address inherent challenges in terminology management cost-effectively and to facilitate exchange of information between users, all these terminology subcommissions use TRADOS MultiTerm terminology management system. They maintain the database by adding new terms and their explanations. If the term is presented in different branches, the consolidation of different meanings of it is explored and the corresponding Latvian term (terms) is proposed. A database joining together these specialized databases is created. This database is available to Internet users.

TRADOS MultiTerm is the world's leading terminology management solution and is specifically designed to enable corporate organisations to:

1. build and manage a corporate term base in a cost-effective manner;
2. eliminate non-productive time/cost spent on correcting erroneous term usage;
3. ensure accurate and consistent usage by all players in the product/service lifecycle;
4. support distributed teams of contributors: authors; terminologists; editors; reviewers and users;
5. publish corporate terminology within and outside the organisation cost-efficiently;
6. maintain central control of critical terminology assets.

7.2 Classification systems

This section describes the use of classification systems in terminology work on the basis of the following definitions of the term *classification system*.

- The first definition reflects the standardisation point of view (originating from ISO 12620:1999E, p. 23):

“A classification system is the arrangement of concepts into classes and their subdivision to express the relations among them. [...] The *classification system* used in a terminology collection can appear as a header code to indicate the classification system used throughout a terminology file or document, e.g. UDC, BRT, etc. or it can vary among entries and be reported separately in each entry. The *classes* themselves are covered by the subject field data category.”

- The second one has the practical aspect in its focus (Source: <http://www.termium.gc.ca>, The Pavel Terminology Tutorial)

“A structured scheme for categorizing knowledge, entities or objects to improve access or study, created according to alphabetical, associative, hierarchical, numerical, ideological, spatial, chronological, or other criteria.”

In practical work, a classification system is used to organise the terms of a term collection which implies that

- a classification system should cover all subject fields (domains), in which an organisation does terminology work
- concepts (and terms) are examined in relation to a subject field, thus one of the most important tasks is to understand and define exactly what a subject field covers.

There are a number of established and agreed classification systems which may constitute a basis for terminology work in a given organisation, for a particular purpose. Such classification systems are set up for various purposes covering selected subject fields such as the medicine and health care domain, biology, economics, law, etc.

This section presents the classification systems which are used as a basis in the various, selected term collections to arrange concepts (and terms) supplied with the list of relevant subject fields which are covered by the classification system(s) in question.

7.2.1 IATE – Classification systems

The classification system used by IATE is the Eurovoc system, see chapter 5.1.2.

7.2.2 TERMIUM – Classification systems

Prefatory remarks

All information given on the various subtasks within this section related to the TERMIUM ® termbank is based on the home page mentioned above in Section 6.1.4, where also the *Pavel Terminology Tutorial*, henceforth PTT can be found at the following web address: http://www.termium.gc.ca/didacticiel_tutorial/english/lesson1/index_e.html.

This interactive tutorial is a systematic summing up of the considerable experience on terminology work gathered by the staff of the Government of Canada's Translation Bureau Translation Bureau where the TERMIUM ® data bank is developed. The tutorial is elaborated in co-operation with experts within various fields of language technology and comprises also internationally recognised terminology standardisation principles and methods.

Classification systems may differ in complexity, level of detail, perspective and focus according to the particular purpose for which they are designed. According to the general principle of studying and describing concepts and their terms in relation to their subject field, and to organise terminological data in the TERMIUM ® data bank, an appropriate classification system is worked out.

This classification system is characterised by the following features:

- It covers all fields relevant to the Translation Bureau's work (cf. Section 6.1.4), viz. "almost every field of human endeavour is covered" (quoted from the homepage http://www.termium.gc.ca/site/pourquoi_why_e.html)
- Some subject areas are more elaborated than others because of the fact that various levels of specialisation in working areas of translation, etc. results in various needs of detail in the classification.

The PPT provides a brief description of the subject-field classification system of TERMIUM ® : "In this classification system, sets of concepts and their terminology are organized into 24 broad subject fields. On average, each broad subject field, or class, is divided into 10 to 12 subject fields (divisions), each of which is, in turn, divided into sub-fields. This gives a total of about 1,600 classification nodes. This system continues to be adopted and adapted elsewhere by language professionals responsible for establishing smaller terminology databases.

The following table lists the broad K class (Electronics and Informatics) of the TERMIUM® *Subject-field Classification Guide*, as well as divisions of this class."

K. Electronics and Informatics

DIVISIONS

KA	Cybernetic Systems
KB	Informatics
KC	Electronic Systems
KD	Computer Hardware
KE	Software
KF	Automatic Control Engineering
KG	Electronics
K-	Provisional Classification

(Source: PTT, Section 2.2.3. Subject-field Classification Systems)

The TERMIUM ® classification system is widely adopted and adapted in smaller terminology projects because of its broad coverage and elaborateness.

In PTT the following practical recommendations are given as regards best practices:

The most important introductory step is to delimit the scope of the terminological activities and to identify the focus and knowledge structure within the subject field selected.

If an organisation, etc. wants to adopt the TERMIUM® or another existing classification system, it is necessary to check the coverage of the classification system adopted against the own products, workflow and activities. It is likely that some areas must be elaborated more in detail because of the project's individual focus on different subject fields and areas.

7.2.3 Estonian Legal Language Centre – Classification systems

Lenoch Universal Classification System, also used by *Eurodicautom*

7.2.4 Lithuanian language term base - classification systems

Terminological data search of Lithuanian language term base is multiple-characteristic. Primarily retrieval request can be made according to subject field. At this moment there are 25 subject fields: automatics, automobile road, aviation, botany, chemistry, electrotechnics, physics, geodesy, geology and physical geography, hydraulic engineering, information, informatics, coastal science, linguistics, computer programs, mathematics, medicine, meteorology, pedagogy, polytechnics, psychology, radioelectronics, sociology, sports, textile. Requests can be made to all subject fields.

7.2.5 Latvian language term base www.TermNet.lv – classification system

Historically, the *Lenoch Universal Classification System* was used for categorizing terminology in LAS. Thus it is still used in LAS and in TermNet portal.

Currently, there are 35 terminology categories/domains in www.TermNet.lv:

Domain	Languages	Number of terms
Building; Construction; Civil Engineering	LV-RU	41502
Customs	LV-EN	12788
Railway	LV-RU	10126
Economics	LV-EN-GE-RU	9959
Agronomy	LV-RU	9230
Geology	LV-RU	7861
Agricultural Machinery	LV-RU	7323
Information Technology, Telecommunications	LV-EN-RU	5004
Archaeology	LV-EN-GE-RU	3915
Glass Technology	LV-RU	3790
Card files of Terminology Committee of Academy of Sciences of Latvia	LV-RU	3728
Graphic Arts, Publishing	LV-EN-GE-RU	3622
Agriculture	LV-EN-RU	3592
Ceramic Art	LV-RU	3335
Medicine	LV-RU	3128
Primary Legislation of EU	LV-EN	1651
Museology (Museum Science)	LV-EN-GE-RU	1638
Chemistry	LV-EN	1619
Botany	LV-LA	1352
Library Science	LV-EN-GE-RU	1319
Motor Transport Machinery	LV-RU	1318
Power Engineering	LV-EN-GE-RU	990

Civil Law	LV-EN	932
Sports Science	LV-EN-GE-RU	906
Nature Preservation, Nature Protection; Environmental Protection	LV-EN-RU	784
Legislation of LR	LV-EN	661
Social Security	LV-EN	652
Motor Vehicle	LV-EN-GE-RU	526
Pedagogy	LV-EN-GE-RU	417
Government Control	LV-EN	398
Zoology (Latvian Ornithological Terms)	LV-RU-LA	377
Textile Industry	LV-EN-GE-RU	275
Hydro Engineering	LV-EN-GE-RU	220
Mechanics	LV-EN-GE-RU	109
Forestry Science	LV-EN-GE	56

7.2.6 Hungary – an example of subject areas

In the following an example of subject areas appearing in the EU terminology prepared by the Hungarian Ministry of Justice is presented.

The tag <domPAR> means top-level classification, the tag <domHUN> means second level classification in Hungarian. The tags <domENPAR> and <domENG> are their counterparts in English.

```

<domPAR>Belügyek, rendészet</domPAR>
<domENPAR>Internal affairs, law enforcement</domENPAR>
<domPAR>Biztonságpolitika, hadászat</domPAR>
<domENPAR>Security policy, warfare</domENPAR>
<domPAR>Egészségügy</domPAR>
<domENPAR>Health and recreation</domENPAR>
<domPAR>Elnevezések</domPAR>
<domENPAR>Proper names</domENPAR>
<domPAR>Energia</domPAR>
<domENPAR>Energy</domENPAR>
<domPAR>Hírközlés</domPAR>
<domENPAR>Communication</domENPAR>
<domPAR>Ipar</domPAR>
<domENPAR>Industry</domENPAR>
<domPAR>Jog</domPAR>
<domENPAR>Law</domENPAR>
<domPAR>Környezet</domPAR>
<domENPAR>Environment</domENPAR>
<domPAR>Közgazdaság</domPAR>
<domENPAR>Business and economics</domENPAR>
<domPAR>Közlekedés, szállítás</domPAR>
<domENPAR>Transportation</domENPAR>
<domPAR>Kereskedelem</domPAR>
<domENPAR>Commerce</domENPAR>
<domPAR>Mezőgazdaság</domPAR>
<domENPAR>Agriculture</domENPAR>
<domPAR>Pénzügyek</domPAR>
<domENPAR>Finances</domENPAR>
<domPAR>Regionális politika, strukturális eszközök</domPAR>
<domENPAR>Regional policy, structural instruments</domENPAR>

```


<domPAR>Statisztika</domPAR>
 <domENPAR>Statistics</domENPAR>
 <domPAR>Szellemi élet</domPAR>
 <domENPAR>Intellectual life</domENPAR>
 <domPAR>Számítástechnika</domPAR>
 <domENPAR>Computing</domENPAR>

<domHUN>Élelmiszeripar</domHUN>
 <domENG>Food processing</domENG>
 <domHUN>Építőipar</domHUN>
 <domENG>Construction</domENG>
 <domHUN>Állattenyésztés, állategészségügy</domHUN>
 <domENG>Livestock production, animal health</domENG>
 <domHUN>Acélipar, kohászat</domHUN>
 <domENG>Steel industry, metallurgy</domHUN>
 <domHUN>Adó</domHUN>
 <domENG>Taxation</domENG>
 <domHUN>Alapjogok</domHUN>
 <domENG>Fundamental freedoms</domENG>
 <domHUN>Atomenergia</domHUN>
 <domENG>Nuclear energy</domENG>
 <domHUN>Bőrparipar</domHUN>
 <domENG>Leather and skin industry</domENG>
 <domHUN>Belső piac</domHUN>
 <domENG>Internal market</domENG>
 <domHUN>Biztosítás</domHUN>
 <domENG>Insurance</domENG>
 <domHUN>Büntetőjog</domHUN>
 <domENG>Penal law</domENG>
 <domHUN>Bányászat</domHUN>
 <domENG>Mining</domENG>
 <domHUN>Elektronika</domHUN>
 <domENG>Electronics</domENG>
 <domHUN>Eljárásjog</domHUN>
 <domENG>Legal procedures</domENG>
 <domHUN>Erdészeti, fafeldolgozás</domHUN>
 <domENG>Forestry and timber production</domENG>
 <domHUN>Földrajzi nevek</domHUN>
 <domENG>Geographical names</domENG>
 <domHUN>Fogyasztóvédelem</domHUN>
 <domENG>Consumer protection</domENG>
 <domHUN>Gépgyártás, gépészet</domHUN>
 <domENG>Production of machinery and mechanical engineering</domENG>
 <domHUN>Gyógyszerészet</domHUN>
 <domENG>Pharmacology</domENG>
 <domHUN>Hajózás</domHUN>
 <domENG>Navigation</domENG>
 <domHUN>Halászat</domHUN>
 <domENG>Fishery</domENG>
 <domHUN>Hulladékgazdálkodás</domHUN>
 <domENG>Waste management</domENG>
 <domHUN>Intézmények, szervezetek, testületek</domHUN>
 <domENG>Institutions, organisations, corporate bodies</domENG>
 <domHUN>Jogforrások címei</domHUN>
 <domENG>Titles of legal sources</domENG>
 <domHUN>Költségvetés</domHUN>
 <domENG>Budget</domENG>
 <domHUN>Közút, gépjárművek</domHUN>
 <domENG>Public roads, motor vehicles</domENG>

<domHUN>Közbeszerzés</domHUN>
<domENG>Public procurement</domENG>
<domHUN>Külkereskedelem, vámjog</domHUN>
<domENG>External trade, customs law</domENG>
<domHUN>Kultúra</domHUN>
<domENG>Culture</domENG>
<domHUN>Légiközlekedés, űrrepülés</domHUN>
<domENG>Air transport, space flight</domENG>
<domHUN>Menekültügy, idegenrendészet</domHUN>
<domENG>Asylum</domENG>
<domHUN>Meteorológia</domHUN>
<domENG>Meteorology</domENG>
<domHUN>Monetáris politika</domHUN>
<domENG>Monetary policy</domENG>
<domHUN>Méréstechnika</domHUN>
<domENG>Measuring technology</domENG>
<domHUN>Növénytermesztés, növényegészségügy</domHUN>
<domENG>Plant cultivation, plant health</domENG>
<domHUN>Nemzetközi jog, diplomácia</domHUN>
<domENG>International law, diplomacy</domENG>
<domHUN>Normatív szövegekben szereplő fordulatok</domHUN>
<domENG>General phrases appearing in normative texts</domENG>
<domHUN>Oktatásügy, tudományos élet</domHUN>
<domENG>Education, scientific life</domENG>
<domHUN>Orvostudomány</domHUN>
<domENG>Medical science</domENG>
<domHUN>Polgári jog</domHUN>
<domENG>Civil law</domENG>
<domHUN>Pénzintézetek, tőkepiac</domHUN>
<domENG>Financial providers, capital markets</domENG>
<domHUN>Rendőrség</domHUN>
<domENG>Police</domENG>
<domHUN>Szőlészet, borászat</domHUN>
<domENG>Viticulture</domENG>
<domHUN>Szellemi alkotások joga</domHUN>
<domENG>Intellectual property rights</domENG>
<domHUN>Szociális jog, foglalkoztatás</domHUN>
<domENG>Social law, employment</domENG>
<domHUN>Számvitel</domHUN>
<domENG>Accountancy</domENG>
<domHUN>Természetvédelem</domHUN>
<domENG>Environmental protection</domENG>
<domHUN>Textilipar</domHUN>
<domENG>Textile industry</domENG>
<domHUN>Vasút</domHUN>
<domENG>Railways</domENG>
<domHUN>Vegyipar</domHUN>
<domENG>Chemical industry</domENG>
<domHUN>Versenyjog</domHUN>
<domENG>Competition law</domENG>
<domHUN>Veszélyes anyagok</domHUN>
<domENG>Hazardous materials</domENG>
<domHUN>Vállalkozások, gazdasági társaságok</domHUN>
<domENG>Enterprises, business companies</domENG>
<domHUN>Vízgazdálkodás</domHUN>
<domENG>Water management</domENG>

7.3 Source identification

Terminology work (research and development) is usually carried out on the basis of documentary sources describing the domain in question for identification of concepts, terms, relationships, etc.

In each environment where terminology work is taking place, a method or a routine of document categorization is developed. Document categorization is an important task for proper registration, recovery and for a possible development of a system of source document retrieval.

In order to establish headers appropriate for source identification, the following questions of primary importance are considered:

- How does the organization categorize documentation; which basic criteria of categorization are applied, which header types and which labels are used? (Examples header types: publication type, publication size, approval status, etc. Examples of labels for the header 'publication type': encyclopedia, dictionary, book, article, brochure, technical manual, proceedings, internet sites, etc.)
- How does the organization prioritize; which are the basic preference criteria? (Example of priority list: Proceedings preferred to brochures, documents in original language preferred to translations into second language)
- How does the organization evaluate the source documentation; which are the basic evaluation criteria? (Examples of criteria: author's expertise, publication date, publication medium, bibliography, table of contents).

7.3.1 IATE – Source identification

All sources should be original documents and only translations that are legally authentic or have recognized authority in the domain in question may be used. References to sources should be precise and comprehensible in order for users to consult them.

A source can be an official body only if it is a recognized authority in the subject field. In cases of no specific source, the terminology institution should be given as default source.

Sources accessible via the Internet can have the URL added along with the date of consultation.

Documents from the same institution should be presented in the same way.

In IATE the request for information on the different kinds of sources is as follows:

- Official and legal documents like agreements, conventions, case laws, secondary legislation etc.: References should be concise and include relevant short forms, type and number of documents, key words, place and date of signing, article number and possibly hyperlink
- Books and other publications and articles from such publications: References should contain author/editor, title, publisher, place, year, ISBN, page number and other relevant bibliographical details.

- Organisations, institutions or private companies that are responsible for the authenticity of a term: References should include name of the institution and appropriate details of department, location, URL, contact person etc.

7.3.2 *TERMIUM – Source identification*

A broad variety of specialised written and oral knowledge source types is employed in the development of TERMIUM ® such as federal and governmental documents, terminology standardisation acts, various technical writings, scientific encyclopedias, oral communications with subject-field experts, specialist's statements, etc. Original and translated texts in English, French and Spanish are considered (priority is given to documents in original language).

In a terminology database, the documents selected as sources may be processed so that they meet the management requirements of that database. The method of source identification adopted in TERMIUM is not explained in detail, but the general principles sketched in the PTT are mainly based on the method of TERMIUM (as cited below from Section 2.3.5 Selection and Evaluation of Documentation)

- Source codes may be created according to source coding rules that apply to the entire database (alternatively, source titles can be recorded in full in such a way that they can still be recognized and processed by the software application)
- Source information is entered into the sources field of the terminology record according to the established record completion rules.
- Source references are given for any quotation or any document consulted, in conformance with copyright laws.

7.3.3 *TSK – Source identification*

The source identification is mainly based on suggestions from domain experts, internet search, public libraries, and using material from the document collection in TSK's library.

7.3.4 *Estonian Legal Language Centre – Source identification*

The research starts from the four sources:

1. Database of the Centre
2. Eurodicatom
3. Specialised dictionaries
4. Experts

The following sources are also used:

1. Databases on CD-ROM (e.g. Britannica, Celex, Ontario Legislation Infobases etc)
2. General language dictionaries and encyclopaedias
3. Estonian or European legislative acts of a similar subject
4. The translated documents of the Centre
5. Handbooks and textbooks
6. Articles in journals or newspapers

The criteria for deciding whether a source is trustworthy:

1. The source should not be a translation
2. Specialised dictionaries are more trustworthy than general language dictionaries
3. Specialised journals are more trustworthy than general journals or newspapers, except when the author of the corresponding article is an expert
4. Tabloid newspapers are not trustworthy at all
5. A journal dedicated to the same subject field that the terminologist's scanned document, is more trustworthy than a journal dedicated to some other subject field

7.3.5 *University of Tartu – Source identification*

Due to the limited volume of the dictionaries, only basic terms can be included. The terms are identified from the textbooks of the right complexity (basic school, pupils' age being 14-16). A term may be found in a glossary at the back of the textbook, in the text as an expression in boldface, as the heading.

The source of the definition may be a specialised dictionary, encyclopaedia, textbook, general language explanatory dictionary. Usually the author has to simplify the definition.

7.3.6 *Lithuanian language term base - Source identification*

Approbated term dictionaries, vocabularies and glossaries are the main source of the term base. At this moment in the term base there are nearly 274 000 terms from 25 various term vocabularies. Bibliography of each term dictionary is placed in separate field.

7.3.7 *Latvia – Source identification*

The reliability of the source is important. The main sources for terms and definitions are term standards, subject-field dictionaries, technical manuals, official bulletins, etc., approved by appropriate (recognized) authority (with approval status). Different LSP books, articles and the like are out-of-status of the official documents.

We can use different books (original and translated), articles, different dictionaries, etc., but only part of them (and only original ones) has approval status. Among the evaluation criteria of the source should be: the competence (prestige) level of the author (authors), the official status of the institution, the publication date (every 10 years the subject-field terminology should be up-dated).

7.4 Use of the internet as a resource

The internet has become the primary information source for many language professionals, such as translators, technical writers, terminologists, etc. because the world wide web presents a huge repository of texts from all domains, in all languages. The role of the terminologist is to identify relevant sites and texts and exploit the material for the purpose and task defined.

This section deals with the use of internet as a resource and discusses this topic in some detail, including tools and methods

- the method for identification of available resources on the internet
- the criteria for identification of language(s), domain, source, text type, etc.

7.4.1 *Types of search engines*

There exists a number of general, widely used web crawler/spiders fetching URL pages, commercial portals (e.g. Google, Alltheweb, Yahoo, etc.) language identifiers (based on identification of character sequences), domain identifiers (based on the presence of domain specific words), etc. Further, project-specific, in-house developed tools might be used which, being targeted at the particular purpose, satisfy individual requirements.

This subsection deals with the description of the search engines used for fetching relevant sources for terminology work within the selected subject field(s).

7.4.1.1 TERMIUM – Types of search engines

Examples of currently used online search engines are:

Google (www) locates Web pages containing a particular term and ranks them.

Other online search engines used are AltaVista (www) and Vivísimo (www) and the desktop metasearch software Copernic Agent (www).

The search engine Vivísimo performs queries by key words combined with operators such as + and -, selects documents and automatically sorts them into folders by topic and subtopic.

7.4.1.2 TSK – Types of search engines

TSK is mostly using Google for internet searches.

7.4.1.3 PolTerm – Types of search engines

PolTerm contributors also use Google when searching the Internet.

7.4.2 *Search methodologies*

Depending on the working task to be carried out various strategies can be adopted, such as identification of key words (e.g. KWIC), search for synonyms and broader terms; domain search; title search and search by concepts relevant to the domain in question (based on an onomasiological structure of the domain), etc.

This subsection gives an overview of the methodologies employed in searching for relevant documents on the internet.

7.4.2.1 TERMIUM – Search methodologies

The description of methodologies of searching the Internet is based on the information found in the PTT (mainly in Section 4.4.3 Search Engines).

The location of relevant home pages and documents can be done by using various search and navigation engines on the Internet, indexing them, and retrieving them with such tools as Isys Desktop 6 or AltaVista Discovery.

7.4.2.2 TSK – Search methodologies

At TSK, the Internet is regarded as an inexhaustible source of information both for terminography and translation purposes.

Only ten years ago it was much more difficult to translate technical texts, when the Internet could not be used on the ordinary translator's computer. If there was an unknown term and an equivalent could not be found in dictionaries, this meant time-consuming information search in libraries or from available experts. Now, the translator can use a search service and can usually get numerous references to sources dealing with the theme of the translation.

Searching the internet is today a matter of standard procedure in both terminology and translation work, it can be used e.g. to find out which term candidate or way of spelling is the most common.

The terminologist's and translator's searches on the Internet differ from each other. The terminologist defines concepts and looks for concept relations in order to create something more permanent, e.g. a term entry in a terminology database, whereas the translator wants to know which words to use in order to convey the message of a single text to the reader and may need the term only once.

Dictionaries or term collections can only include few examples of usage, but on the Internet terms can be found in their authentic contexts. The translator can also see what kind of style a certain product manufacturer uses on their own web pages, and can adapt the translation into the same format.

In this sense, the Internet is used as a large scale text source for terminological and linguistic observations supporting the work carried out at TSK.

7.4.2.3 PolTerm – Search methodologies

The Internet English-equivalent search with a view to add new entries in PolTerm whenever they appear is strictly term- and, whenever possible, text-type based. Thus a specific Polish term or, even better, a potential equivalent is entered into a search engine with a view to find it in an English-language legislative Act. It is crucial that such contexts are looked for that enable to delineate the meaning of the term and thus conduct a comparative inter-language concept analysis. Statutory definitions would be the best comparative material, but if impossible to find, several meaningful contexts could also be very helpful.

7.4.3 *Evaluation of web site information*

The evaluation of web site information comprises several tasks carried out (semi-)automatically and manually.

This subsection provides a brief description of the most important tasks and steps of the process and the tools and methods used. Main tasks are e.g. content analysis by a comparison with existing approved material; checking the linguistic quality of the document (re. grammar, vocabulary, style of an original-language text or of a translation); approval of the timeliness and completeness of the contents relative to the subject field. Examples of evaluation criteria are relevance to the subject field (domain), reliability of the source, content, actuality, language/usage of terms, etc.

7.4.3.1 TERMIUM – Evaluation of web site information

In general, all documentation and information sources including the web site information are evaluated before they are selected for terminology work. The set of general evaluation criteria is listed in PTT (Section 2.3.5. Selection and Evaluation of Documentation), these basic criteria may also apply in TERMIUM:

- Relevance and quality of the terminology and number of defining elements present in the text
- Adequacy of the document content to the project objectives
- Quality requirements (reliability of author, source/homepage, provider)
- Type of internal knowledge organisation within the document (including glossaries and indexes of concepts appearing in the document)
- Timeliness and completeness of the contents
- Linguistic quality

Virtual libraries (e.g. Bibliotheca Universalis) and searchable text databases (e.g. New York Times; Wired) are mentioned as examples of appropriate source types which can be accessed via the Internet.

7.4.3.2 TSK – Evaluation of web site information

In the evaluation process TSK relies preferably on in-house terminological expertise.

7.4.3.3 PolTerm – Evaluation of web site information

Any terminological findings are first verified for consistency with the PolTerm terminology database, i.e. we have to make sure that so far there is no such term in our collection, so as to avoid confusions within the terminology system. Secondly, a finding undergoes a thorough LSP dictionary analysis and, to approve the term finally, experts cooperating with TEPIS are consulted.

7.4.3.4 Estonian Legal Language Centre – Evaluation of web site information

The criteria for trustworthiness:

1. Authorship
2. Properties of the web page, like:
 - 2.1 When was the page updated? Are the links pointing to existing pages?
 - 2.2. Are there spelling errors?
 - 2.3. Are there links to the page from other pages?
 - 2.4. Who is the intended audience of the page?

2.5. What sources has the author of the page used?

See also <http://lib.nmsu.edu/instruction/evalcrit.html>

7.4.4 *Creation of subject specific corpus*

Once relevant sites, www-pages and documents are identified, fetched and evaluated, a subject specific corpus can be created. This subsection reports on the following questions:

- **What kind of corpus is compiled**, listing the relevant features of subject field(s), size, main sources and text types, etc.
- **How the creation of subject specific corpus is carried out**, including a brief description of the most important steps and tools and methods used as well.

7.4.4.1 PolTerm – creation of subject specific corpus

Today the PolTerm electronic corpus contains 42 consolidated texts of Polish legislative Acts and their English translations, where the translations. Some more important Acts are: the Civil Code Act, the Code of Commercial Partnerships and Companies Act, the Freedom of Economic Activity Act, the Natural Persons' Income Tax Act, the Legal Persons' Income Tax Act, the Excise Duty Act, the Banking Law Act. The corpus amounts to approximately 30,000 bilingual translation units.

The creation of the PolTerm subject specific corpus relies on a TRADOS TRANSLATOR'S WORKBENCH – a translation memory (TM) tool. In order to compile it, the WINALIGN tool, an interface between a word editor and the Workbench, has been used, so that source sentences and corresponding translations have been grouped in segments covering a specific chunk of a text according to predefined segmentation rules and automatically aligned side by side: an Article, a paragraph, or a subparagraph. Following the check of a correctness of the automatic alignment, appropriate Act-specific identification codes have been chosen and automatically added to all bilingual translation units created by automatic alignment of segments. Then a WINALIGN file is exported to a .txt format and then imported to the Translator's Workbench tool, and more specifically to the PolTerm Translation Memory file, according to our labels. In visual terms, the translation memory screen shows a Polish segment and a corresponding English segment beneath it, both of them bearing the relevant Act-specific identification codes identical to the ones included in the PolTerm terminological database.

Such operation is carried out on a regular basis, i.e. every two months a new Polish Law Collection is to be published. The methodology is as follows: all new segments are added to the PolTerm TM, segments made obsolete by the legislator are deleted, and relevant amendments to existing ones are done, making it possible to output any required subset of segments via either entire-TM-level- or specific-identification-code-restricted concordances.

7.5 **Compilation of terms and conceptual analysis**

This section describes the two main initial steps of terminology work and the tools used in these steps.

The current state of terminological data processing benefits from various information processing techniques, such as structuring information in a databank and from natural language processing methodologies, such as extracting information from machine readable texts. Knowledge management by using computers is a fast developing area of science which also influences terminology work.

Terminology compilation is generally carried out in two basic working situations

- Collection of single terms in specific cases requested by translators, technical writers, subject field experts, etc.
- Systematic development of terminology for documentation purposes e.g. in form a large-scale national term databank, a technical dictionary of a particular subject field, or a smaller term databank for company/concern language, etc.

The automation of terminology compilation, storage/management and retrieval opens up the prospect of large-scale high quality term banks with easy access to their systematic data collections.

7.5.1 *Term extraction*

In this process, a corpus or other collection of documentation of the subject field is systematically scanned for terms. The delimitation of the subject field (or domain) is highly dependent on the theoretical background and the practical motivation (e.g. users' request) of the work.

The aim of the process is to record and structure relevant information found in the text sources, this can be regarded as the linguistic dimension of terminology work with the following main tasks to be dealt with:

- Identification of concepts belonging to the subject field, and registration of their designations (terms) and definitions
- Observation of typical linguistic contexts and term usages.

This subsection specifies some relevant practical aspects and details of the term extraction procedure as it is carried out by various organisations and working groups. These details are intended to illustrate the working conditions and approaches taken.

Firstly, the initial position of the terminologists work is defined by the languages(s) to be dealt with i.e. whether the terminologist is working in one language at a time or in two - or more - languages simultaneously. There are both similarities and differences between monolingual and bi- or multilingual terminology work, also in respect to term extraction.

Secondly, although the possibilities offered by automatic analysis of texts are extensively integrated into term extraction, it still can be performed manually (e.g. identification of terms and their immediate contexts by high-lighting in running texts).

In many projects the process is carried out automatically (including recognition by using a term identification tool, recording term usages based on statistical analysis of large volumes of texts, etc.) or semi-automatically with a certain degree of human intervention. In semi-automatic term extraction there are several areas of manual actions which require particular knowledge of the

domain and the language(s) involved, e.g. checking the relevance, reliability and/or validity of the term extracted compared to the scope of the subject field, checking the textual information related to the term esp. in case of rarely used terms where the frequency information is insufficient as a basis for systematic term extraction.

Finally, the main steps of the workflow are outlined, such as how terms are collected and recorded, e.g. extract terms from each source and set up a list of term candidates, compare these lists with the aim of observing relevant term characteristics and merge the candidate list into a single final list, etc.

7.5.1.1 TERMIUM – Term extraction

The method of computerized term extraction is recommended by PTT in case of machine-readable documents e.g. downloaded from the Internet.

Different approaches and steps can be employed when using the Internet as a resource. It is evident that the more steps of the term extraction process are automated the more effective is the whole process as such.

Terms are searched and the evidences found are ranked by using Google, it provides the searched term highlighted in the context. For term documentation and textual support purposes the standard copy-and –paste function of Windows is appropriately used.

In printed documents, terms can be manually highlighted and marked for extraction.

7.5.1.2 TSK – Term extraction

The term extraction is usually carried out manually, because a significant part of the source material is not accessible in electronic format. For automatic or semi-automatic processes also term extracting programs are tested.

7.5.1.3 INTERA – Term extraction

In the Intera project, a term is identified merely on statistical grounds, on the basis of its relative frequency in a corpus. Although not theoretically correct (as the status of “termhood” is in principle independent on the number of occurrences), this practice is based on practical considerations and on the particular procedure adopted.

Although the procedure adopted is not necessarily always best practice, the procedure is however a viable and fruitful one given the following conditions:

- The produced terminology is not to be intended as a reference or normative terminology for a given field or domain. Instead, it is to be conceived as an organized collection of the terms occurring in a particular corpus.
- Data are sparse
- No NLP tools are available for the target languages
- No reference corpora are available for the target languages

But:

- Many NLP tools and reference corpora are available for one language (the *pivot*)

The task of automatic term extraction is organized around two main phases:

- Automatic extraction of terms from the English corpus.
- Automatic identification of candidate translations in the target languages

Both phases are followed by an editorial intervention involving manual verification of the candidate translations found with the automatic procedure

The Intera project adopted two different procedures, one for the automatic identification of single terms (i.e. one-word expressions) and another for automatic identification of compounds (i.e. multi-word expressions).

Single Terms

In order to determine the domain specificity of single terms for a given linguistic domain a frequency general lexicon for the pivot language was used. Since this lexicon was not available inside the project, a lemma-based frequency lexicon of the *British National Corpus* was generated.

A parallel frequency lexicon was then generated for each specialized language.

In more detail, the comparison between the frequency distributions of terms in the general lexicon and that of the different domain-specific lexicons was performed adopting a mathematical function evaluating the distance of the frequency of domain-specific terms from the frequency which was expected on the basis of the general lexicon.

The procedure yields a file containing the ordered list of frequencies of all candidate terms in a corpus.

Compounds

Multi-word terms are represented by patterns of grammatical categories. To perform the analysis the procedures developed for the system *Linguistic Miner* developed at ILC, was applied. An initial phase of work concerned the definition of “Rules”, i.e. linguistic patterns to be identified in the texts. As for single-term techniques, lemmas are used instead of inflected forms. After a testing phase, the following patterns proved significant:

Adjective + Noun: e.g. *administrative procedure*

Noun + (and/or) Preposition (and/or)=OF + Noun: e.g. *merger regulation*

Adjective + Noun + Noun: e.g. *national road haulage*

Adjective + Adjective + Noun: e.g. *chronic idiopathic urticaria*

Noun + Preposition (and/or)=OF + Adjective + Noun: e.g. *service of general interest*

Noun + Preposition (and/or)=OF + Noun + Noun: e.g. *driver of motor vehicle*

Adjective + Noun + Preposition (and/or)=OF + Noun: e.g. *documentary letter of credit*

Adjective + Noun + Preposition (and/or)=OF + Adjective + Noun: e.g., *common organisation of agricultural market*

Once the search patterns have been created, text analysis procedures retrieve and rank the data. The final product of this procedure is, for each rule, an ASCII file containing the compound candidate terms for the domain under inspection, in inverse frequency order.

7.5.1.4 PolTerm – Term extraction

The extraction of new terms for the PolTerm terminological database is carried out on a regular basis together with preparation of a set of new amendments to relevant Polish legislative acts to be translated into English and then published in a subsequent edition of the Polish Law Collection (see 3.5.5 above for the PLC format details). As both the amendments and the whole corpus of bilingual consolidated versions of laws (the PolTerm translation memory - TM) are stored in an electronic format, the extraction of new Polish terms is semi-automatic and involves checking the presumed novelty of a term noticed by the editor against both the PolTerm terminological database and the PolTerm TM. Presumed novelty occurs when neither the PolTerm term database nor the PolTerm TM, which together form the so-called “PolTerm platform”, identified the term at the stage of automatic processing (automatic processing through the platform is carried out in respect of a new Polish text to be translated with a view to preservation of terminology consistency: if the platform finds any matches, either of terms or of segments, they are automatically inserted in the word-editor text as proposed ready-made translations). If such a term still proves to be new in the “PolTerm platform” following the manual checking of occurrence thereof, it is analyzed as regards the syntactic structures in which it usually occurs and the most frequent construction – together with the term itself, if it is not a syntactic structure, or a phrase, but a multiword term that can be extracted from such a construction – is entered into a preliminary term list. Afterwards the term is subject to conceptual analysis in view of finding its best possible English-language equivalent.

7.5.1.5 Estonian Legal Language Centre – Term extraction

The aim of term extraction as the first step in document translation process is:

- to ensure that terminology is used consistently in the translations, no matter who the translator is
- to ensure that the translator uses correct terms
- to avoid unnecessary duplication of work by different translators

The terms are selected manually.

A terminologist scans the source text, in order to identify all the terms in it that should be included in the Centre’s database of terms. During this process, the terminologist groups the terms into 1) the terms of the subject field the document deals with (e.g., chemistry); 2) legal terms; 3) other terms. The legal terms are redirected to the person, responsible for legal terminology. If the terms and the phraseology of the text indicate that it would require a specialist to understand and translate it, then the Centre will try to find one from outside. Otherwise, the terminologist will start with the concept analysis of the extracted terms. In all cases, the terms are added to the term database of the Centre, no matter how incomplete their entries are, so that other translators and terminologists would know whether it is being dealt with, and by whom.

7.5.1.6 University of Tartu – Term extraction

The aim of term extraction is to arrive at collections of terms, suitable for including in specialized dictionaries, meant for 14-16-year-old pupils. That is, every collection of terms must cover the teaching subject in appropriate detail and complexity. Terms that are not used in the curricula, should not be included, even if they would otherwise definitely belong to the whole conceptual system of the subject field (unless they are absolutely necessary for explaining the included terms).

The terms are selected manually – from textbooks and supplementary materials of the curricula.

7.5.1.7 Latvian partner – Term extraction

Term extraction from the text during translation process is practiced mainly with source language texts. National (in our case – Latvian) term equivalents are taken mainly from term dictionaries, or LGP dictionaries and, if necessary, creating new words or new word combinations (adjective or participle + noun; noun + noun, etc.).

Term extraction tools as mentioned in chapter 6.3.2.3 are usually not used. Translation centre TTC usually uses its own translation memory and term database and other databases.

7.5.1.8 Morphologic – Term extraction in Hungary

Due to the fact that the Hungarian language is an agglutinative language, and every word can take up millions of forms, current off-the-shelf terminology extraction software does not provide satisfactory results for this language. Therefore, only translators could use terminology extraction to extract source language terms (English, German, French usually), but so far, the Association of Hungarian Translation Companies does not know about any translation agency applying this method. Some translators, however, know about the existence of such tools.

Due to the fact that in Hungary terminology is usually the servant of translation, MorphoLogic and Kilgray are developing a term extraction tool with a strong emphasis on translational terminology – all those lexical elements the inconsistent translation of which can spoil the comprehensibility of a text. The term extraction engine employs both rule-based and statistical methods. Rule-based methods require a basic dictionary, a morphological engine and a shallow syntax parser, while some statistical methods work with a lemmatized input. Therefore, the engine to be developed only works for English, German, French, Czech, Polish, and Hungarian at the moment.

The engine is based on six extraction methods, and a term candidate is more likely to be a term candidate if it appears in the results given by several methods. A term candidate does not necessarily have to repeat in a given text.

The methods include the following:

1. Finding lexical items not appearing in the basic dictionary (high recall)
2. Identifying the internal morpho-syntactical structure of terms and searching for matching patterns (high recall, low precision)
3. Identifying terms by their environments (e.g. appearing in a definition, etc.) (high precision)
4. Looking for unusually frequent expressions (high precision)
5. Identifying association measures, for the extraction of multiword expressions (high recall, low precision)
6. Inductive methods (departing from a base glossary, identifying SL collocations using the former methods, and identifying similar SL elements to those included in the glossary)

At the moment, the methods are oriented towards monolingual (source language) term extraction, but the engine also contains some bilingual statistical methods increasing the precision of the source language methods by performing sub-sentence alignment or checking possible translations included in dictionaries.

Hitherto, this engine has not been released on the market. Currently only one company (SZAK Publishers, a publishing house active in translating IT books) has integrated term extraction into its workflow. During the translation of two books, preparation was provided with the term extraction engine. Due to the fact that it is translation-oriented, high recall is expected. Unfortunately, it is

very hard (if not impossible) to define a percentage of the recall, but translators only reported a few other terms, less than 3% of the terms included in the final list. Therefore, we can estimate that its recall is about 97%. Precision was lower, an experiment was conducted, which resulted in a precision value of 77,08% for English, and 67,08% for Hungarian as source language on texts of ca. 175000 words.

See also references from Hungary.

7.5.2 *Concept analysis*

The basic element of terminology work is the concept. (cf. Guide to Terminology, p.8.) The process of concept analysis is **closely related to term extraction**; it **represents the cognitive dimension of terminology**. The information extracted from the textual sources need to be analysed from the point of view of domain knowledge structure. “The analysis required to identify and determine the scope of a concept designated by a given term as it is used in a particular subject field.” (Pavel Tutorial, TERMIUM).

Concept analysis of terms of a subject provides the knowledge basis for organising these concepts into a concept system of the field in question.

A concept analysis of extracted terms reveals

- The types of relationships that hold between concepts in one language, viz. generic, partitive and associative relations
- Synonyms (including abbreviations, not preferred, but still valid term variants, etc.)
- Equivalent terms across languages (in bi- or multilingual terminology).

The aim of this subsection is to give a brief overview of the methodologies employed by terminologists in this process. The main questions discussed here concern the method of identifying and recording relations between concepts, e.g. in concept systems, by informing of narrower terms and broader terms, with hyperlinks where relations are unspecified or by other methods and means.

(Although concept analysis always comprises the identification of relationships between concepts, in the very initial phases of a terminology project, it might be the case that the work done so far does not include this process yet.)

7.5.2.1 TSK – Concept analysis

Conceptual analysis: based on ISO/TC 37 standards; a concept-oriented approach is applied to the work (instead of word-oriented or context-oriented approach) to ensure that the degree of terminological quality of the work is as high as possible.

7.5.2.2 PolTerm – Concept analysis

So far the concept analysis within PolTerm has been limited to a secondary term formation, i.e. the search for English-language equivalents of Polish legal terms, once they are extracted (for extraction methodology details see 7.5.1.2).

The basis of the methodology underlying the secondary term formation within PolTerm: a model of pragmatic translation of legal terms (Kierzkowska 2002) based on modern terminology and translation theory as regards transposition of the contents of a language-specific legal terminology into another language.

The model's principles: the translation strategy is determined by the type of the recipient (“the recipient’s imperative”) and by a possibly existing English-language prescriptive terminology with regard to a given source term (“the imperative of terminology usage”). With respect to the PolTerm terminology system a “close recipient” is assumed, meaning one who has the motivation to acquire knowledge of the source language culture and is not necessarily an English-speaking-country citizen, being at the same time “an international/multicultural recipient”, i.e. an “international standard English” speaker. Communicative needs of such recipient require the use of a translation strategy based on seeking denotative equivalence, i.e. source-language-culture oriented (*overt* translation – House 1997) and possibly “neutral” (not Anglo-Saxon legal systems-bound) English-language equivalents.

There are three main practices in PolTerm aimed at retaining the specificity of Polish legal concepts and simultaneously at “neutrality” of terms, taking into account the interdependence of denotative-equivalence strategy and the terminology usage imperative in the context of increasingly relevant harmonization and unification of Polish legal system mainly with *acquis communautaire*, but also with international provisions, including initiatives and reports issued by internationally recognized organizations (e.g. International Monetary Fund, the World Bank, UNCITRAL etc):

1. *In case of finding no identical or highly similar concept* expressed by an English term in any of the systems a PolTerm term may be coined by means of naturalized borrowing (e.g. “*voivodeship*” for “*województwo*” – a unit of local government);

2. *In case of a partially similar concept* in one or more of the foreign contexts a PolTerm term may be “seminatural” equivalent which evokes an association with a relevant system, but at the same time a difference is indicated by, for example, adding an element (e.g. “*registered partnership*” for “*spółka jawna*” – a type of partnership, the concept being similar in both British and American contexts, which, unlike partnerships in these systems, requires registration);

3. *In case of a highly similar or identical concept* expressed by a given term in one or more systems, or a given term covers essential characteristics of the Polish concept collectively in many systems, the examples here being “*podpis elektroniczny*” and “*electronic signature*” (EU law) – of the former, and “*upadłość*” and “*bankruptcy*” (British law, United States law, international documents) – of the latter. The case of “*electronic signature*” is an ideal situation as regards “neutrality” as here we “take over” or retranslate a “neutral” and yet widely recognized term.

Consultation with PolTerm specialists

The essential issue in respect of grounds for formation, or selection of terms in such cross-system comparative analyses of concepts and their denotations is evidently the degree of equivalence of concepts. One source of such information is the legislature itself: the statutory definitions of and the grounds for an Act or issued by the legislators and containing the elucidation of their decisions, often including a direct reference to a particular foreign context with regard to many new Polish legal terms, even with one-to-one Polish and English term citation cases, which often serves as a supplement to definition-based comparisons. There has also been made an extensive use of the doctrine – published comments and analyses of specialists concerning the legal provisions. But the overall success of the decision-making process determining the form of PolTerm terminology has been possible due to active cooperation between terminologists and subject specialists: legal and economic consultants, who are able to ascertain the degree of similarity of a source-language term and a presupposed equivalent.

A recent PolTerm methodological innovation: the notion of semantic and normative micronetworks (Borkowski 2005) in which a given source-language legal term can be placed and which reflect its legal and semantic context by putting it in relationships with other relevant terms, thus recreating the legal norm concerning the term. This enables to identify its legal, and not only semantic, meaning, within the micronetwork. Said micronetworks also serve as matrixes enabling to conduct cross-system comparative terminological analyses more efficiently, as they allow to identify compatible, even if partially, networks of terms in other legal systems, or to exclude any similarities.

7.5.2.3 UKIE – Concept analysis

The terms that are qualified to be entered into the data base are either new or such that may prove difficult for the verifiers. These terms are entered into the provisional data base where they are supplemented, checked in the professional literature and consulted by subject specialists. Terminology checks often require a lot of time-consuming and meticulous research; it is usually the case that our Department is provided with the support of experts in central administration bodies and scientific institutions that are competent for considering given texts.

Also other sources of terminology are useful in terminology work – data base of EC legal acts Celex, mentioned above Eurodicautom, Lex, the Polish law data base, and numerous special texts in the Polish language.

7.5.2.4 Estonian Legal Language Center - Concept analysis

The most difficult problem for a terminologist is to deal with a set of closely related concepts that are expressed by homonyms or synonyms, occurring in sometimes different and sometimes closely related subject fields.

The terminologists should clear up such sets of concepts only if they are directly related to the terms from the documents they are currently dealing with, that is, if the translator would otherwise not know which term she should use.

7.5.3 Tools

As mentioned in the above sections, the **compilation of terminology** highly benefits from the development of computational tools. Especially, larger projects make use of both commercial tools (e.g. for corpus extraction) and in-house developed term extraction and knowledge structuring tools.

This subsection provides a brief overview of the tool types used for particular tasks in term compilation, such as term extraction tools (e.g. browser-based such as Yahoo's tool, XML based tools), and tools for structuring (term clustering), storing and representing concept systems (e.g. LEXTER, FASTR,..)

7.5.3.1 TERMIUM - Term extraction tools

In case of texts being downloaded from the Internet and other machine-readable documents, it is evident to use a semi-automatic term-extraction software, PTT mentions YVANHOË[©] and an automatic term-extraction software Nomino.

A further account for possibilities of automating term extraction processes to a high degree:

“For example, with the professional version of Copernic Agent, it is possible to explore 120 subject categories using over 1,000 different search engines as well as extract key concepts from the results using the summarizer. The product also features automated search operations such as downloading, validating and refining, and offers a spell checker for queries and a search wizard to facilitate query formulation. With it you can organize searches into folders and track them, and even set the system to automatically relaunch searches at regular intervals and e-mail you new results.”

7.5.3.2 TSK - Term extraction tools

TSK has tested automatic technical term extraction programs from Connexor Oy. (Finland). The recognition of term candidates is based on automatic content analysis which produces an estimation of essential information in a document and the term-likelihood of the term candidates. The term-likelihood is measured by weights; a high weight indicates that a term candidate is likely to be a term. The tool is working for Finnish and English.

Further information on the Machine Metadata tool: <http://www.connexor.com>.

7.5.3.3 INTERA - Term extraction tools

The Intera project applied the system *Linguistic Miner* developed at ILC. *Linguistic Miner* was developed for the creation of an automatic system for the extraction of linguistic knowledge and to be used for a variety of goals, including didactic, publishing, and cultural ones.

7.5.3.4 PolTerm – Term extraction tools

There are no direct term extraction tools, but an indirect semi-automatic extraction method worth considering is mentioned above (7.5.1.2). In brief: the presumed novelty of a term noticed by the editor and deriving from it not having been identified either by the PolTerm term database or the PolTerm TM during an automatic source text processing stage, is verified by the editor against the entire PolTerm platform. If proved new, a term is extracted by the editor.

PolTerm – term compilation tools

The Trados Multiterm 95 plus tool is used for terminology compilation.

7.6 Creation of terminological entries

The information type coverage of a given entry reflects the intellectual and practical work done in previous phases of a given compilation of a terminological vocabulary.

As a consequence of this correlation, the information types (or data categories) of entries will by definition differ depending on the context in which the terminological work is embedded. Not only is the content of an entry a variable element in an entry, also the choice of data category names and their definitions can be used differently. To give an example: The information type called 'context' in one term collection is perhaps called 'definition' in another, but both do belong to the data category in ISO 12620 called 'context' (a context that provides a summary explanation of a concept).

7.6.1 Data categories in terminological entries

The kind of information that is presented in this section will be the following: Which and how many data categories (information types) are used in the terminological entries (e.g. term, definition, context, note). Do the category names and definitions comply to the ISO 12620?

7.6.1.1 IATE – Data categories in terminological entries

IATE Data Fields:	DOMAIN_NOTE
	ORIGIN
	ORIGIN_NOTE
	PROBLEM_LANG_CODE
Language independent level	COLLECTION
	CROSS_REFERENCE
	GRAPHICS
LIL_RECORD	Language Level
INSTITUTION	
AUTHOR	
PROPOSER	
MARKED_FOR_DELETION_MERGING	LIL_RECORD
CONFIDENTIALITY	AUTHOR
DATE_MADE_CONF	INSTITUTION
MADE_CONF_BY_USER	CREATION_DATE
	CHANGED_BY
CREATION_DATE	CHANGE_DATE
CHANGED_BY	CHANGED_IN_FIELDS
CHANGE_DATE	MARKED_FOR_DELETION_MERGING
CHANGED_IN_FIELDS	
DOMAIN	LANGUAGE_CODE

DEFINITION	TERM
DEFINITION_NOTE	TERM_TYPE
DEFINITION_REF	LOOKUP_FORM
DEFN_REF_CONF	OBSOLETE
LL_COMMENT	TL_COMMENT
COMMENT_CONF	COMMENT_CONF
COMMENT_MADE_CONF_BY_USER	DATE_COMMENT_MADE_CONF
DATE_COMMENT_MADE_CONF	COMMENT_MADE_CONF_BY_USER
RELATED_MATERIAL	RELIABILITY_VALUE
RELMAT_CONF	
GRAPHICS	TERM_REF
	TERM_REF_CONF
Term Level (includes word level information)	LANGUAGE_USAGE
	LANG_USAGE_REF
	LANGUSE_REF_CONF
TL_RECORD	REGIONAL_USAGE
AUTHOR	REG_USAGE_REF
PROPOSER	REGUSE_REF_CONF
INSTITUTION	
CREATION_DATE	CONTEXT
CHANGED_BY	CONTEXT_REF
CHANGE_DATE	CONTEXT_REF_CONF
CHANGED_IN_FIELDS	GENDER
MARKED_FOR_DELETION_MERGING	PART_OF_SPEECH
INITIAL_SOURCE	
VALIDATION_STATUS	
STAGE	
CYCLE	

7.6.1.2 INTERA - Data categories in terminological entries

A list of the INTERA data categories and their equivalents in the ISO 12620

INTERA	ISO 12620	Explanation
Id	entry identifier	
Domain code	subject field	
Language	language identifier	
Term	term	
grammatical info	PoS	
context	context	
component	-	indication of the components of a mwu entry
rank	-	indication of the position of the components

inside the term

7.6.1.3 TSK - Data categories in terminological entries

The structure of the entries is based on the Nordic Terminological Record Format (NTRF) and the information is recorded, modified and converted with the help of an own xml-based terminology management system. A typical NTRF record contains information about a single concept.

The NTRF is introduced in 1995 as the common format for the central terminology institutions in Finland, Norway and Sweden. The last HTML version is dated 1999-04-28. (Source: <http://www.rtt.org/ntrf/ntrf.htm>)

Below a brief outline of the main characteristics of NTRF based on the information on home page mentioned above.

In NTRF the text is broken down into fields according to the information types to be included, the grouping and nesting of fields are permitted while overlapping between fields is prohibited, this latter is a simplification and it makes NTRF different from SGML, although NTRF is compatible with SGML.

Field types in NTRF

Top level (may be also subordinate to another top level field, but not to an embedded field)

Embedded (nested into another field to any depth)

An NTRF record may look like this (Source: cf. above)

enTE English term

POS noun

frTE terme <GEND m> anglais

enDEF typical definitions frequently contain

cross-references to <RCON another term>

CX A context where the English term is found

SOURF Source-document-1

CREA 1996-10-24 HHj

APPR 1996-11-03 CBE

Fields and field type tags

Field tags consist of three parts: a language symbol in accordance with ISO 639, a field type tag, and an optional number. The field type tags are used for data exchange.

The list of field type tags is divided into six categories of information relating to:

Terms (Examples: TE: primary term, SY: synonym term, ACRO: acronym, GRAM: grammatical information; PHR: phrase: phraseological context to a term);

Concepts (Examples: DEF: definition, EXPLAN: explanation, CX: context);

Concept relations (Examples: BCON: broader concept, NCON: narrower concept, TYPR: type of concept relation, EQUI: equivalence between terms in different languages);

Classification (Examples: SUBJ: subject area, SCOPE: specific area of use, CLAS: classification, viz. the position within the main classification system used, CLASYS: classification system);

Presentation (Examples: HEAD: heading for use in a systematic presentation, SORT: sorting form for production of proper sorting sequence, ILLU: illustration);

Administration (Examples: NUMB: running serial number, SOURC: source of record, CRDAT: date of creation, UPDAT: date of updating, APPR: approved date and by viz. date and initials, STAT: status).

Further information on the Nordic Terminological Record Format: <http://www.rtt.org/ntrf/ntrf.htm>

7.6.1.4 PolTerm – Data categories in terminological entries

Currently, a typical terminological record layout in PolTerm has the following form:

Creation Date
Created By
Change Date
Changed By
Entry Number
Polish
English
Code
Dawniej
Retusz
Drobne

DESCRIPTION:

Fields <creation date>, <created by>, <change date>, <changed by>, <entry number> are called entry header fields; they are created by the program by default and their contents display important house-keeping (called also “administrative”) data concerning a concept. The <Change date> field allows to reflect the changes in conceptual structure of the lexicon covered by the PolTerm TM as we can show, for example, when exactly a concept was removed by the legislator within the collection, i.e. when a term became obsolete, by entering a note “invalid” in <Drobne> (“misc.”) field.

Index fields <Polish> and <English> are used to store terms, i.e. language-specific linguistic data.

Text fields <Code> through <Dawniej> (“previously”) refer to additional descriptive data relating to a concept and the term. <Code> contains an Act-specific identification code attributed to all PolTerm terms and being usually an abbreviation of the full English name of an Act, so that, for example, ACCOUNT is the code of the ”Accounting Act”. So it refers to conceptual classification of the term and its equivalent as it points to the particular domain of law where a term is used. In <Drobne> various notes can be stored, like the term-invalidation information. <Drobne> field is also used by us to enter source information, such as a document or terminological resource on the basis of which the English term was, for example, retranslated. However, not all English terms can be attributed any such type of source, as the source-oriented legal translation philosophy of TEPIS often results in coining a term. But even in such cases sometimes the English version of a term is based on linguistic decisions of Polish authorities. This was the case with all local government nomenclature, such as: *gmina*, *powiat*, *voivodeship* etc., the forms of which were approved by Konwent Marszałków RP (the Convent of Heads of Local Self-Government) in 2000.

The consistency-maintenance policy of TEPIS is so essential, that we have long strived at documenting the consistency-building record within the PolTerm term base microstructure. The <Dawniej> and <Retusz> (“Correction”) fields serve to input the information concerning any errors with respect to an English term form, such as typing errors or insufficiently concise formulation of a term. The term is corrected in the index field, in <Dawniej> the previous form is entered, and in <Retusz> the appropriate form is inputted to be globally changed on the level of the entire PolTerm platform.

Prospects

Due to increasing harmonization of the Polish law with that of the European Union resulting in many PLC-included Polish Acts implementing the EU provisions, we have decided to improve the conceptual-structure-reflection balance of our bi-lingual corpus within the database microstructures, so that it would look like that (the following restructuring to be applied first to the ”basic” terms subset placed on the TEPIS website):

[house-keeping data categories]:

Creation Date
Created By
Change Date
Changed By
Entry Number

[linguistic data categories]

Polish
English

[conceptual data categories]

Subject field
Code
Definition

[pragmatic data category]

Context

[misc.]

Previously
Correction
Source

Apart from that, work is also underway to provide cross-entry links conveying important information about the relationships between terms within the terminological system. It refers mainly to homonyms and apparent synonyms (here information is provided to differentiate between such two terms).

7.6.1.5 UKIE – Data categories in terminological entries

The data stored in the base can be divided into:

- house-keeping information data that help to control the data. They include: entry number, creation date and change date, as well as data on the person who made the change. These fields are filled in by the program automatically;

- linguistic information reflected by index fields containing entry terms and equivalent terms;
- term-related information: text fields containing the source of the term, its definition, the source of the definition, remarks and the source of remarks. Additional information included in attribute fields, such as the subject fields or personal notes. Said data serve the users and administrators of the data base to communicate with each other; they may also have a warning function if there are any doubts as regards the quality of the data stored in the record.

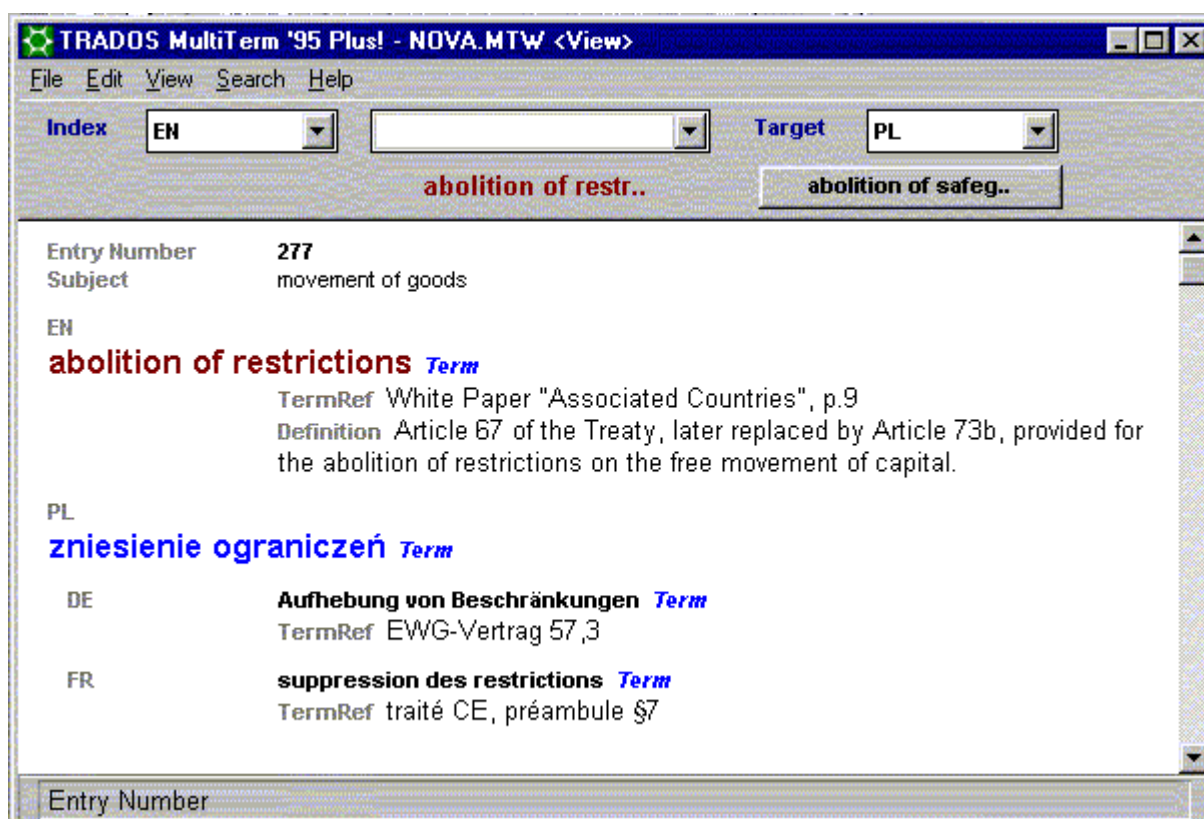


Figure 7.1. A sample record of the UKIE Translation Department's data base

The first stage of the terminology work consists in the separating terminology from four language versions that are equivalent to each other. Such situation is rather rare when we deal with several original versions of the same texts.

The data base structure is based on a model recommended to all bodies dealing with the translation process coordination in all EU newly-admitted states. However, some modifications have been made, as its original form did not fulfil all desired functions.

The data collected in the data base are entered as four-language terminological records, which make the Polish translation and verification sub-unit the only one among sub-units of other newly-admitted states that applies as many languages in its work. Usually, the record structure is based on one foreign language chosen originally. Creation of records in four languages enables also translators and verifiers who use a language different than English to make use of the *acquis communautaire*.

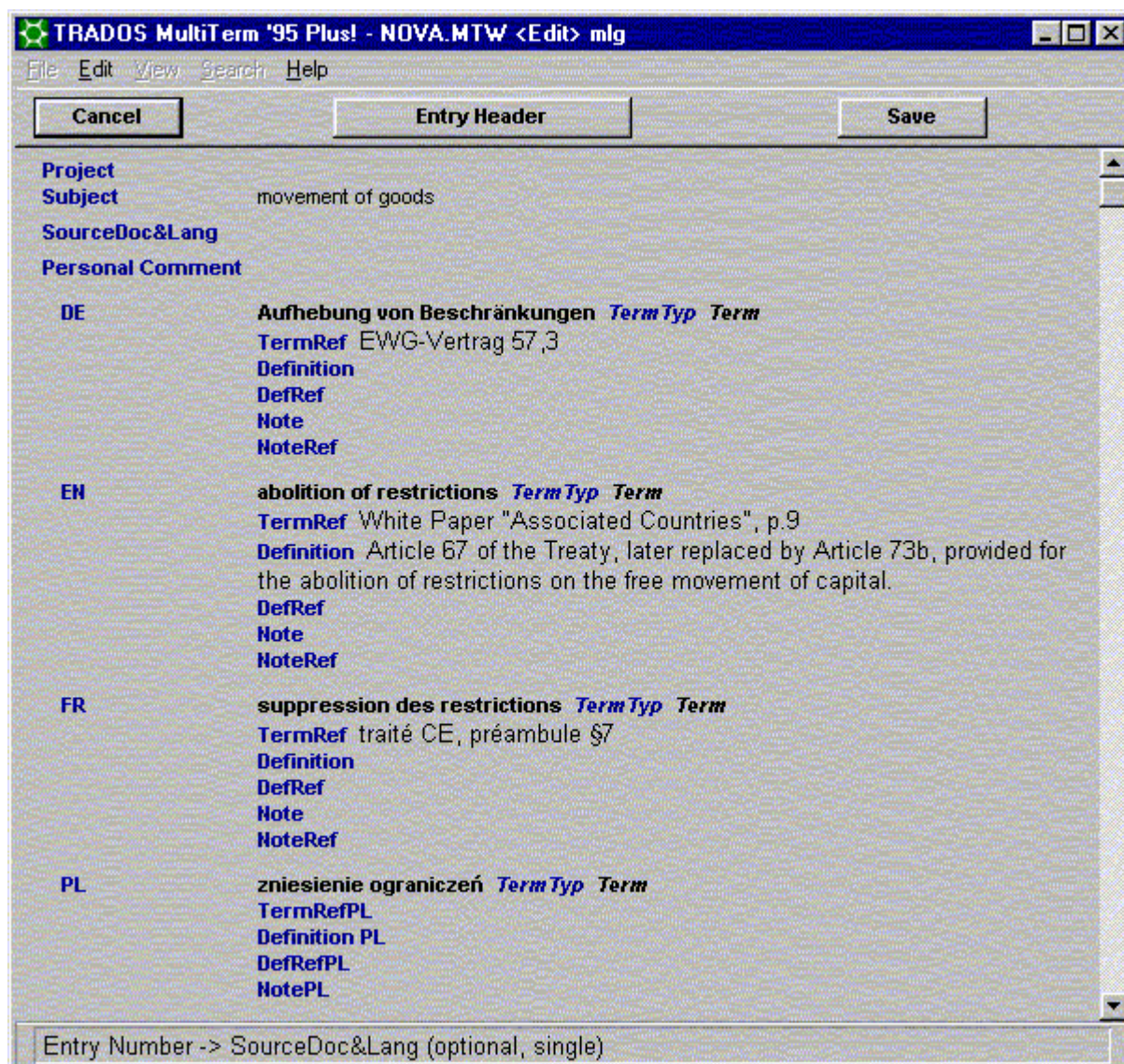


Figure 7.2. The record from fig. 1 in the edit mode where all fields can be filled in with relevant data

7.6.1.6 Estonian Legal Language Centre - Data categories in terminological entries

An entry of the publicly consultable term base ESTERM (<http://www.legaltext.ee/en/esterm/findex.htm>) contains information about a single concept. The data categories should be self-explanatory:

Language independent level:

- Record_no
- Subject field according to Lenocho classification (Universal Classification System)
- Last_modified
- Translated_From

Language dependent level:

- .. Term
- .. Term_Reference
- .. Definition
- .. Context
- .. Abbreviation
- .. Synonym
- .. Variant

A detailed description of the data categories is available at:
<http://www.legaltext.ee/en/esterm/findex.htm>

In addition to the data categories of the publicly viewable information, the in-house database contains project-related data categories, restricted for in-house use:

- Reference to the folder that contains unstructured draft information about the concept and term
- Origin of the term
- Type of the term (real term, translation problem, name of an institution, or profession)
- ID of the translation project that is dealing with the term
- Translation direction
- Is the term fit for public display in Internet? (yes/no)
- Previously used term
- Synonym that should be avoided
- Comments

7.6.1.7 University of Tartu - Data categories in terminological entries

The terminological databases that will serve as the basis for printed dictionaries contain the following data categories:

1. Term
3. Subject field
4. Synonym
5. Definition
6. Example
7. Reference to a related concept
8. Russian translation (and synonyms) of the term

7.6.2 Valid entries vs. concise, complementary and exhaustive entries

Important information that should be stated in this context is the definition of what maximum number of information types are allowed in a given entry and what is required in order to pass for constituting a valid entry.

So it is necessary to be given the information of the maximum set of information types in the terminological entries of the organisation and how many information types are being defined as obligatory.

7.6.2.1 IATE - Valid entries vs. concise, complementary and exhaustive entries

A terminological entry must meet certain criteria with relation to credibility, relevance and minimum amount of information.

A terminological entry should give translators and terminologists as much information as possible in order to judge whether the information given is credible and appropriate, if necessary by consulting the sources and references used. The minimum information of a terminological entry should enable translators and other terminologists to clearly delimit the concept in question which means that definition together with references and contexts should have high priority. An ideal entry should have information on at least one of these three categories in order to be useful and credible. An entry should also have an added value over what can be found in databases and on the internet. The added value could be different kinds of terminological information like e.g. definitions, references. Multilingual information is also an added value.

The IATE database should only contain relevant entries. A relevant entry concerns a drafting, translational or interpreting problem in an area of relevance for the Community. And the term must correspond to the correct sense in the relevant context. A relevant entry should not be an everyday term since the inclusion of such might lead to problems with duplication of information and the definition of the boundaries of the entries. Only everyday terms that have an added value in Community documents compared to their definitions in language dictionaries should be entered.

Entries in IATE are by definition valid and correct unless they are marked ‘deprecated’.

A term should be as concise as possible, i.e. it should be the smallest unit that can designate a given concept exactly. Complex terms or expressions can often be broken down into various concepts and lack of conciseness should be avoided for the following reasons:

- It is difficult to add other languages to complex expressions since such expressions are seldom found in other languages.
- It is also difficult to check for duplicates if a term covers more than one concept.
- Finally the relevance of an inconcise entry is reduced and the lack of conciseness might be an obstruction for its use in computer-assisted translation.

All entries in IATE are based on consultation of sources and the inclusion of references to such sources has high priority. However, copyright laws might prevent the inclusion of longer segments of reference material unless copyright owners have granted permission for such use.

A concise definition is very important in order to enable users and terminologists from other languages to delimit the concept and estimate the solution proposed.

Context is a practical way to illustrate the term and its use but it is important to follow the rules for conciseness.

7.6.2.2 PolTerm - Valid entries vs. concise, complementary and exhaustive entries

For an entry to be valid, the following mandatory terminological data categories apply in translation-oriented terminography (ISO/DIS 12 616.2: 8): term; language symbol; source. In the

case of the PolTerm terminological database said categories translate into: <English>/<Polish> fields' content; <English>/<Polish> fields' names; <Code> field's content as regards both a Polish term and its English equivalent in relation to their Act-specific origin within the PolTerm Translation Memory; <Source> field's content: a crucial data category in connection with English equivalents of given Polish source terms. Whenever possible, such information is contained, so far in the <Drobne> field. Under this category there is currently a number of terms of the EU origin which we have decided to use as retranslations (the imperative of terminology usage) on the basis of comparative interlingual linguistic and legal analysis, likewise there are many other equivalents. As described under 7.6.1.4, also other highly relevant data categories are included in entries and intended to be included in the nearest future with regard to the "basic" terms subset.

The conciseness requirement has been complied with ever since the PolTerm term base was created, as it results from the theory of terminology (e.g. Sager 1990: 89-90). The approach is reflected in avoiding the use of certain parts of speech, especially in English terms, e.g. the preposition "of", if deemed redundant in a given example, so that even three- or four-word compound nouns are obtained and yet are perfectly comprehensible.

7.6.2.3 Estonian Legal Language Centre – Valid entries

The minimum set of data categories that form a valid entry (in addition to categories that are necessary for the formal consistency of a database) is the following:

- .. Term in the source language
- .. Subject field
- .. Definition in the source language
- .. Context of the source language

7.6.2.4 University of Tartu – Valid entries

The minimum set of data categories that form a valid entry is the following:

- Term
- Definition
- Russian translation of the term

7.6.2.5 TC of LAS and LLI of LU – Valid entries vs. concise, complementary and exhaustive entries

When picking out a model for a term database entry, we could compare it with an entry of a term dictionary. The best model for an entry of a printed term dictionary according to the needs of nowadays society in member-countries of EU is the term with its definition and term equivalents at least in one, but more preferred – in three or more main languages of EU.

For term and definition standards a model is accepted where definitions are given in two or more languages too (at least, in English and/or French).

The e-format of term databases allows us to store rich information, although the model for definitions should be the same as for standardized definitions where the characteristic features of definitions are systematicity (systemic), conciseness, etc. (see in specified version under the title “Main principles for term definitions”). Contexts are different and variable. In fact, it is impossible to collect all possible contexts for future translations. It is and will be necessary that a translator has a good skill both in the source and target language.

7.6.3 Writing rules

In major terminological institutions in which many people are involved in the creation of terminological entries, the making of *Writing rules* often ease meeting the goal of achieving coherent and consistent entries.

Such *Writing rules* describe i.a. the purpose of each information type and how to fill them in the various information types. Another important issue in this Writing rules document is to provide guidelines for how to fill in the fields of the entry seen from a formal point of view: Where to use capital and lower case letters? How to separate multiple information in e.g. reference fields? Below, descriptions of the organisations' writing rules can be found.

7.6.3.1 IATE – Writing rules

The writing rules of IATE cover all fields of the data creation interface except fields that are filled in automatically or that are filled in by choosing a single item from a pick-list.

The rules are categorized according to the three levels (see IATE data structures 7.7 and IATE data categories 7.6.1)

Entry level (Language Independent Level - LIL)

Domains and domain note

An entry should only belong to one or few domains. Several domains in one entry may indicate that the entry covers more than one concept. It should also be checked when adding new data to an existing entry whether the context and domain are the same as the ones already there. Narrower descriptors of domains can be added in domain note.

Origin and origin note

A country name can be selected here to indicate the geographical origin of a concept, if necessary supplied in the origin note by a more specific indication of the political, cultural, ethnic or religious origin.

Problem language

The problem language is the language of the term in which it was created and serves as the basis for addition of other languages.

Proposed by

Name of person who proposed the entry if this is not the creator of the entry.

Cross-references

Cross-references are links to related concepts (note that synonyms should be included in the entry itself). The type of relation that holds between the concepts should be indicated (selected from a list).

Collections

The names of term collections should include an institution ID and a short description of the collection.

Graphics

Relevant graphic files can be included with a short description.

Language level:

Definition and note

A definition should follow the principle of substitution and should be broad enough to identify a concept in a general context. Further explanation that are not part of the definition must be placed in the note.

Related material

A list of relevant material apart from the main references.

Graphics

Relevant graphic files can be included with a short description

Term Level:

Term

The term should be concise, i.e. should consist of the smallest indivisible part that designates the concept. Complex expressions should be separated into their constituent parts and an entry should be created for each of the concepts involved.

All terms should be correct, not recommended terms should be marked as ‘deprecated’ followed by an explanation in the note field.

Specific rules for how to write for each part of speech and more general rules that hold for all word classes. The field for grammatical information can be used for exception or specific indications.

In the case of the lack of a definitive term (or title etc.) a provisional solution should be proposed by the terminologist and explained in a note. The term should be updated as soon as possible.

Short form

The short form of a name or title should be included where appropriate.

Phrase

“Phrases” which are frequently found in Community documents, have a standard translation and pose a problem for translation should be included.

Abbreviation

Should be written according to the language specific rules.

Formula

Chemical formulae, mathematical and other scientific expressions should be written according to international standards.

Term Group

Term group numbers indicate synonyms that are morphologically related.

Context

The purpose of this is to demonstrate how a term is used in context.

Language usage

Provides information on usage and style or level of language.

Regional usage

Indicates whether a term reflects regional usage.

Customers

Customer names should ideally include an 'institution ID'.

Lookup forms

Forms of the terms like spelling variations or inflected forms that should be made searchable.

Proposed by

User name of the proposer of the term level information if different from the creator.

References

See above on Source Identification (7.3)

7.6.3.2 TERMIUM – Writing Rules

The writing rules employed in TERMIUM are formulated in the document *Record-completion guide*, which is a reference document that systematically presents the rules to be followed when entering terminological data on records. Example: TERMIUM® record-completion *[insert hyphen]* guide. This means that all information on the record must be entered in accordance with the rules laid out in the Record-completion guide that is applied to the entire database guide in question. (Source of information in this section is PTT, Section 3.4.2 Entering Terms on Records).

The essential elements which must be present in a terminology record are the following:

- The subject field to which the concept belongs
- The languages dealt with
- The terms
- The terms' usage labels
- The textual supports (definition, context)

Examples of commonly accepted writing rules for recording terms:

- Enter terms in singular unless the plural is required to properly designate the concept.
- Enter terms in lowercase, except when uppercase is required in order for the term to be correct.

- Enter terms without elements in parentheses unless they are required in order for the term to be correct (avoid indicating subject fields or optional words in parentheses).

7.6.3.3 TSK – Writing Rules

The writing rules applied by TSK are based on ISO/TC 37 standards and language specific writing rules relevant for the Finnish language.

7.6.3.4 PolTerm – Writing Rules

The writing rules in PolTerm have been drawn up and apply to a kind of consistent “metalanguage” serving to convey some important information on particular entries in descriptive fields (Dawniej; Retusz; Drobne). The point is that the same types of information within these fields have to be indicated in the same way on the entire-PolTerm base level, so as to enable an easy retrieval of a given set of entries containing the same type of information when a need arises. For example, all instances where a Polish term has an EU English equivalent are indicated by entering the words: “SOURCE: EU LEG: (here a particular legislative act is cited)” in the <Drobne> field. If an English equivalent is based on a certain English-language source, but is intentionally different due to the source-language cultural specificity-retention requirement underlying the secondary term formation in PolTerm, a note entered in the <Drobne> field reads: “BASED ON: ... (source cited); MAVERICK GROUNDS: ... (reasons justifying a different form of an English term than the original one)”. There are also many other writing rules.

7.6.3.5 University of Tartu – Writing Rules

The writing rules reflect the aim of the terminological dictionaries – to help Russian-speaking pupils to understand the subject and at the same time learn Estonian.

Subject field

If indicating a subject field inside a broader field (like “geometry” in “mathematics”) makes the definition easier to understand, it should be added to the entry, otherwise not.

Synonym

It is advisable to list all the synonyms, including deprecated ones, which are marked accordingly.

Definition

A definition should be short, using simple language and simple words.

7.7 **Structures of terminological entries**

Different terminology databases have different data structures as approaches and requirements vary from one company to another. Some companies just have a two-column spreadsheet containing source term and target term whereas others have terminological entries with several levels and many information types.

The data structure of one terminology database could for example have 3 levels: a concept level, a language level and a term level. The concept level contains all the information that relates to the entire concept, for example the ‘Domain’ of the particular concept. The language level contains the

information that relates to the particular language, and that could for example be the ‘Definition’ (allowing one definition per language). The term level contains information that relates to the term, for example ‘Abbreviation’, ‘Context example’ and ‘Comment’ about use of the particular term. The data structure of another terminology database might only have 2 levels: a concept level and a term level. The ‘Definition’ would then probably be in the concept level allowing only one definition in one language for the entire concept.

In this chapter we will give an overview of the data structures identified in the various terminology collections and we will determine whether they are consistent with ISO 12200.

7.7.1 IATE - Structures of terminological entries

The IATE data structure took into consideration the evolving standards in the field (SALT/MARTIF, GENETER). It adopted a concept-oriented approach; the mono- and multilingual information on each aspect of a concept can be expressed on four inter-related levels of the data structure of the terminological entries:

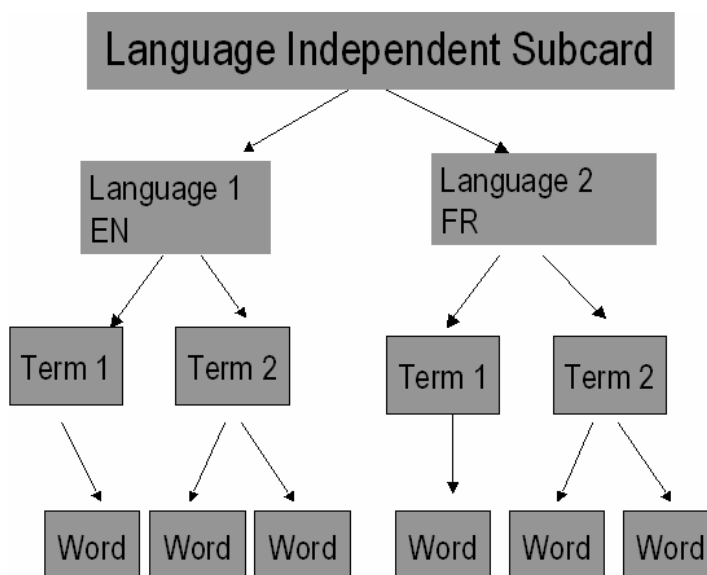


Figure 7.3. IATE data structure

- The language-independent level can contain all information that relates to the entire concept. “Domain” is the classic example of that type of information. But the database also makes it possible to be more exhaustive: the user can add a domain note in cases when the classification system for domains does not contain a suitable descriptor; collection, problem language, cross references to other entries, origin of the concept and – as we are living in an age of multimedia – links to images complete the language independent level.
- Beneath this top level, information like definition and comments can be stored in and for each of the languages the entry contains (language level). This level is enriched by the possibility to add notes on several fields, references to source documents and, again, multimedia files.

- Each language level may refer to several terms – synonyms of the same concept or abbreviations (term level). A large variety of information can be associated with each of the terms: term type, reference, regional usage, context, customers, links to homonyms etc.
- Finally the system includes the option to add linguistic information, like part of speech or gender, for each term or each of the words constituting a term.

7.7.2 *INTERA*– Structure of terminological entries

The **TMF**² family of formats is taken as the reference model to encode the *INTERA* terminological entries.

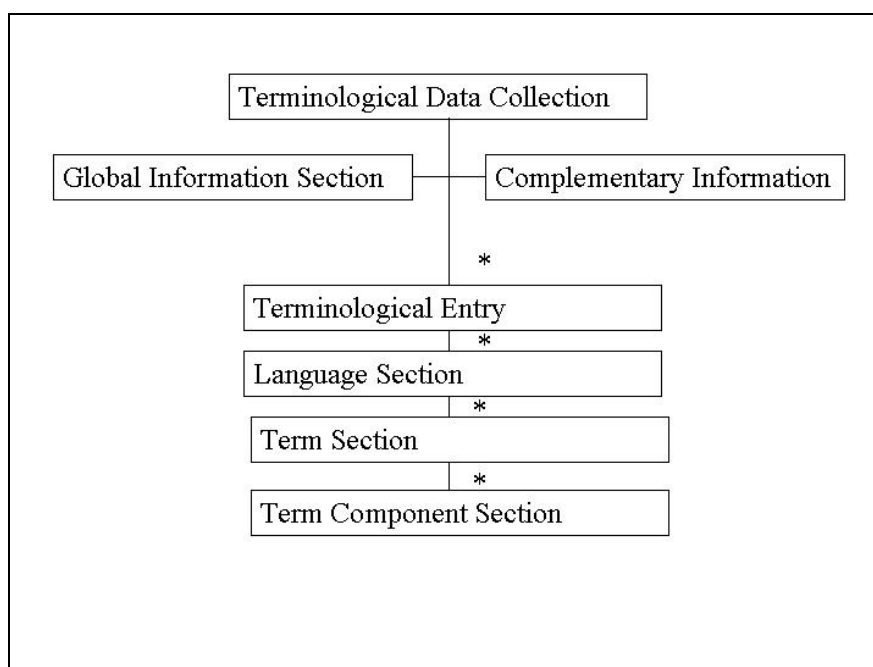


Figure 7.4. *INTERA* structure

7.7.3 *PolTerm* – Structure of terminological entries

The current terminological entry structure in the *PolTerm* term database has already been shown in 7.6.1.4. At the language-independent level the microstructure informs us both about administrative data relating to a particular concept (entry header fields: Creation Date; Created By; Change Date; Changed By; Entry Number), and about the domain of law where it is used (text field: Code). The language level obviously contains data on the linguistic forms of a concept (index fields: Polish; English), whereas the other text fields: *Dawniej*, *Retusz* and *Drobne*, include the information relating first and foremost to the English term (the term level). We shall not elaborate on the prospective structure of terminological entries covered in 7.6.1.4 until it becomes a reality.

² TMF stands for Terminological Markup Framework (ISO 16642 2001), an international standard designed in the framework of the ISO initiatives to support the creation and use of computer applications for terminological data and exchange of such data between different applications. TMF can be described as a meta-model consisting of two levels of abstraction (see chapter 7.8.1).

7.7.4 Estonian Legal Language Centre - Structure of terminological entries

The current terminological entry structure has already been shown in 7.6.1.6

7.7.5 University of Tartu - Structure of terminological entries

The current terminological entry structure has already been shown in 7.6.1.7

7.7.6 Lithuanian language term base – Structures of terminological entries

A terminological entry consists of several parts: term, synonym, definition, equivalents and note. Structure of terminological entry depends on source. The entry of a term on the screen is not always identical to that presented in a vocabulary. The entries of the Lithuanian language term base are represented according to its own rules. Every part of terminological entry is marked. Separate term meanings are placed into separate entries. Singular or plural forms of terms are placed into definition and have separate reference. Terms with facultative part are considered as term variants. Authentic and adapted names are also recognized as variants.

Looking through the results of search, the user himself is rendered an opportunity to comprise entries of mentioned parts, to choose the ordering of terms and the form of entry presentation. Terms can be ordered: a) alphabetically; b) according nest alphabetically. Also it is possible to choose request data form: a) text, b) table.

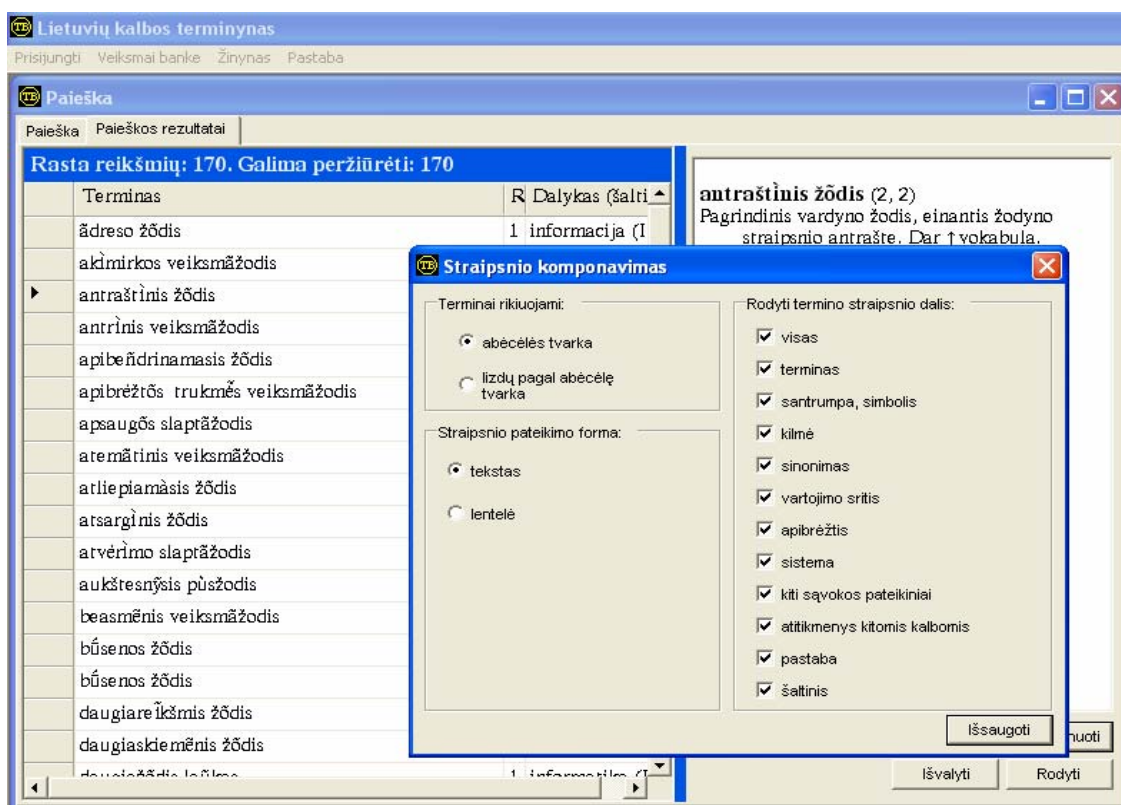


Figure 7.5. Structure of terminological entries - Lithuania

There is a separate section to display a full terminological entry or separate parts of a terminological entry:

1. Term
2. Abbreviation, symbol
3. Derivation
4. Synonym
5. Sphere of usage
6. Definition
7. System
8. Other accessories of notion
9. Equivalent
10. Note
11. Source.

7.7.7 *TermNet.lv Structure of terminological entries*

See section 7.8.7.

7.8 Exchange formats

Information is not directly transferable between databases with different data structures, but there are some exchange formats, that facilitate exchange of information between databases. Exchange formats should be consistent with ISO 12200. It should be noted though that if the structure of the source database is richer than that of the target database, a potential loss of information can only be avoided by re-structuring the target database.

7.8.1 *TBX (TermBase eXchange)*

TBX (TermBase eXchange) is an open XML-based standard format for terminological data. It provides a number of benefits as long as TBX files can be imported into and exported from most software packages that include a terminological database. This capability facilitates the flow of terminological information throughout the information cycle both inside an organization and with outside service providers. In addition, terminology that is made available to the general public should become much more accessible to humans and more easily integrated into existing terminological resources.

For various types of machine processing, including transmission over the Internet, terminological data can be represented using XML. The TBX format is a standard-based XML application designed to support machine processing of terminological data in various computer environments, including standalone computers, the Internet, and intranets.

The terminological framework for TBX is provided by three established international standards: ISO 12620, ISO 12200, and ISO 16642 (see chapter 9).

TBX is designed to support the analysis, representation, dissemination, and exchange of information from human-oriented terminological databases (termbases). TBX is a format that qualifies as a TML (Terminology Markup Language) by complying with the requirements of the Terminology Markup Framework (TMF) according to ISO 16642. It is based on the TMF structural metamodel; it specifies a set of data categories from ISO 12620 and adopts an XML style compatible with ISO 12200.

Each variant of TBX is a TML within TMF. Since each TML is interoperable with every other TML, limited only by incompatibilities in the choice of data categories, TBX XML documents can be converted to XML documents in other formats within TMF. However, interoperability between TBX and formats that do not qualify as TMLs is not guaranteed. Nevertheless, limited interoperability is possible between non-TML formats such as OLIF and TMX.

Even though TBX supports customization according to user needs, there are limits to what variations can be defined by an XCS file; otherwise, certain variations would not qualify as TMLs according to 16642. All acceptable variations on TBX have the same core structure. They differ mainly with respect to the data categories from ISO 12620 that are allowed by a particular user group.

According to the hierarchy of a TBX document, the highest-level XML element is the *martif* element, which contains a *<martifHeader>* element and a *<text>* element. The *<martifHeader>* element provides a description of the file, on the applicable XCS file and unusual character encoding, and a history of major revisions to the collection.

The *<text>* element contains the terminological data. It includes in the *<body>* the actual terminological entries – one entry per concept – enclosed in *<termEntry>* tags, as well as complementary information, e. g. bibliographical data, in the *<front>* and *<back>* elements, to which can be referred from the *<body>* entries. Within the terminological concept entries various data categories allow to provide different kinds of information, either in free text or chosen form a pick list, as well as cross-references that points to either somewhere inside the *martif* element or to an external object using a URL. The terminological concept entries (*<termEntry>*) can be multi- or monolingual.

```
<?xml version='1.0'?>
<!DOCTYPE martif SYSTEM "./TBXcoreStructureDTD-v-1-0.DTD">
<martif type='TBX' xml:lang='en' >
<martifHeader>
  <fileDesc><sourceDesc><p>from an Oracle corporation termBase</p></sourceDesc></fileDesc>
  <encodingDesc><p type='DCSName'>TBXdefaultXCS-v-1-0.XML</p></encodingDesc>
</martifHeader>
<text> <body>
  <termEntry id='eid-Oracle-67'>
    <descrip type='subjectField'>manufacturing</descrip>
    <descrip type='definition'>A value between 0 and 1 used in ...</descrip>
    <langSet xml:lang='en'>
      <tig>
        <term tid='tid-Oracle-67-en1'>alpha smoothing factor</term>
        <termNote type='termType'>fullForm</termNote>
      </tig>
    </langSet>
  </termEntry>
</body>
</text>
</martif>
</langSet xml:lang='hu'>
```

```

        <tig>
        <term tid='tid-Oracle-67-hu1'>Alfa sim&#x00ED;t&#x00E1;si t&#x00E9;nyez&#x00F5;
        </term>
        </tig>
    </langSet>
</termEntry>
</body> </text>
</martif>

```

Example of a TBX document

TBX includes meta-markup tags for distinguishing embedded non-TBX markup from text. They allow TBX elements to contain various kinds of other markup, e. g. html or text processing markup that needs to be retained but should not necessarily be processed during terminology management functions.

In the Annex and separate files of the TBX specifications, important information and examples for the encoding of data according to TBX is provided. This includes a formal XML representation of the core structure containing the basic data categories, element groups, attribute lists and comments associated with meta data, tables defining the TBX master XCS (data constraint specification), and guidelines for encoding particular data categories in TBX (e.g., as XML elements).

The TBX format is a principle result of the SALT project (see chapter 6.2.2). The LISA³ OSCAR⁴ group has adopted TBX as its terminology exchange format and continues the development of the specifications.

7.8.2 *TERMIUM – Exchange formats*

In the development of TERMIUM both standardisation and exchange requirements have a high priority. The use of exchange standards ensures that globalisation and localisation, and term exchange between databases as well can be carried out safely. The following standards produced by ISO Technical Committee 37 and the LISA/OSCAR standards group are mentioned in PTT and presumably also relevant to TERMIUM:

- ISO 12620 Data Categories
- ISO 12200 Computer applications in terminology — Machine-readable terminology interchange format (MARTIF) - Negotiated interchange
- ISO/DIS 16642 Terminological Markup Framework (an abstract conceptual data model to facilitate the exchange of terminological data)
- TBX - TermBase Exchange format (TBX is an open XML-based standard format for terminological data.
- TMX - Translation Memory eXchange (TMX is a vendor-neutral, open standard for storing and exchanging translation memories created by Computer Aided Translation (CAT) and localization tools.

(A more detailed description of the last two can be found at <http://www.lisa.org/tbx/> and <http://www.lisa.org/tmx/>)

³ Localization Industry Standards Association

⁴ Open Standards for Container/Content Allowing Re-use

7.8.3 *TSK – Exchange formats*

The Nordic Terminological Record Format is developed for facilitating the exchange of terminological data between central terminology institutions in Northern Europe, primarily in Finland, Norway and Sweden. The NTRF is described more in detail in subsection 7.6.1.3.

7.8.4 *PolTerm – Exchange formats*

The PolTerm term database can be converted into the .txt format without a slightest data loss. The PolTerm TM (the bilingual corpus database) in turn has both .txt and .tmx format conversion capability, rendering the data, including the language codes, interchangeable between various CAT (computer-aided translation) tools.

7.8.5 *Estonian Legal Language Centre – Exchange formats*

XML, Trados Multiterm.

7.8.6 *University of Tartu – Exchange formats*

MS Access, XML

7.8.7 *TermNet.lv Exchange format*

Internal XML based exchange format. Contains root element which is named **root**.

Root element can contain one or more child elements named **entry**. Each **entry** has its identifier, which is stored in the attribute named **id**.

“**entry**” contains several sections, which contains meaning information:

- **subject** element contains sphere or field in which this term is used. It has three attributes:
 - **lenoch** – Lench subject heading for this meaning;
 - **lenochexpl** – full explanation of lenoch heading in English;
 - **id** – identifier of the subject;
- **date** element contains the date (or only year) when this meaning was approved or published;
- **subcom** element contains the name of the subcommission which has approved the term and its definition. It has two attributes:
 - **url** – internet address of subcommission;
 - **id** – identifier of subcommission;
- **source** element contains the source from where comes current term. It has two attributes:
 - **id** – identifier of source;
 - **visible** – should this source be or not be visible for normal user;
- **mcomment** element contains comments referring to whole meaning;
- **type** element contains the type of the term, it can be “approved”, “suggested” or something else. It has one attribute **id** – identifier of the type;

“**entry**” element has two or more (it depends of the number of languages term is translated to) sub elements named **term**; each of them has further described structure:

- **lang** element contains language information of the term;
- **definition** (optional) element contains definition for the term in current language;
- **link** (optional, but may contain more than one) element contains terms which are related to the current term;

- **langid** element contains LCID identifier of the current language;
- **langorder** element contains numeric language priority in which terms should be displayed;
- **word** element contains term itself in current language;

Example:

```
<?xml version="1.0" ?>
<root>
<entry id="960474">
  <subject lenoch="AU" lenochexpl="Automation (includes telecommunications and
computers)" id="44">Informācijas tehnoloģija un telekomunikācija</subject>
  <date>1998</date>
  <subcom url="http://www.vvk.lv/index.php?sadala=207&id=810"
id="3">Informācijas tehnoloģijas un telekomunikācijas terminoloģijas
apakškomisija</subcom>
  <source id="43" visible="1">LZA TK Informācijas tehnoloģijas un
telekomunikācijas terminoloģijas apakškomisijas apstiprinātie termini</source>
  <mcomment />
  <type id="1"/>
  <term>
    <lang>EN</lang>
    <langid>1033</langid>
    <langorder>2</langorder>
    <word>computer</word>
  </term>
  <term>
    <lang>LV</lang>
    <definition>Tehniska sistēma (ierīču komplekts), kas saskaņā ar uzdotu
programmu veic automātisku datu apstrādi un ievadizvadi. Sk. arī</definition>
    <link>computer system</link>
    <link>desktop computer</link>
    <link>home computer</link>
    <link>laptop computer</link>
    <link>microcomputer</link>
    <link>notebook computer</link>
    <link>palmtop computer</link>
    <link>personal computer</link>
    <link>portable computer</link>
    <langid>1062</langid>
    <langorder>1</langorder>
    <word>dators</word>
  </term>
  <term>
    <lang>RU</lang>
    <langid>1049</langid>
    <langorder>3</langorder>
    <word>ЭВМ</word>
  </term>
</entry>
</root>
```

7.9 Management of terminological entries

Management of records may for example include *adding, updating, deleting, importing, exporting, extracting, converting etc.* In a term database it is usually possible to add new records, update existing records and delete incorrect records. In some databases it is also possible to import/export

data for example to facilitate modifications, to extract subsets of the database perhaps for out-of-house translation jobs and to convert data that are used in more than one application.

Management of terminological entries includes many other types of operations that require certain facilities from the database system. In connection with a large updating task a terminologist may want to start with the latest records. This of course requires that the system contains a mechanism to sort entries by creation date. Furthermore, a terminologist may have certain responsibilities to check the correctness of certain entries. This requires that the system contains a facility that will draw the terminologist's attention to the particular entries.

This chapter describes facilities available in the various terminological databases to manage data.

7.9.1 Process requirements

In connection with management of terminological entries one company's primary process requirements may vary considerably from those of another company. Examples of primary process requirements could for example be that the database should be technically simple, efficient (no unnecessary steps), easy to use, easy to maintain, up-to-date (new info should be added fast), complete (entries should be as exhaustive as possible) and require minimum effort from language professionals. Some of these requirements do not exactly go hand in hand. A complete entry (or exhaustive entry) will for example seldom require a minimum effort from the language professionals. Priorities in relation to these requirements will heavily influence management facilities.

In this chapter we will shortly describe priorities of the various organisations.

7.9.1.1 TSK - Process requirements

Based on ISO/TC 37 standards.

7.9.1.2 PolTerm – Process requirements

The PolTerm terminological database is stored in a computerized terminology processing tool, which presupposes computer-assisted terminology management, i.e. compilation, storage and retrieval of terminology. The management is a two-way process: the necessity to maintain consistency with regard to a new material for translation triggers a need of terminology retrieval, which in turn leads to recording of new data, and sometimes updating (supplementing or correcting) or deleting of the existing data in the database. Since terminology retrieval is covered in 7.9.3.2, we will take up the compilation issue.

Term compilation, update, export and deletion

As a part of consistency-oriented translation and terminology management model in a CAT tool framework, all new source and target term pairs approved in a given Polish Law Collection edition (for extraction of source terms and equivalent approval information see previous points) are collected and assigned relevant Act-specific identification codes within one text file in a word-editor. The file is formatted in conformity with the PolTerm terminological record layout specifying the fields containing important terminological information relating to an entry term (a Polish term)

and an equivalent term. Then the terminology-processing tool imports the records to the PolTerm terminological database in the automatic mode. For occasional record additions working in the manual mode within the database, without the interface file, is also possible, such user-friendliness contributing to the fact that every single new term finding in-between successive PLC editions is appropriately recorded on an immediate basis.

As regards updating and exporting of entries, both of these operations can also be carried out automatically on the basis of any pre-defined set of criteria, such as specification indicating which field contents are to be changed in what way in the entire database, or which entries are to be exported into a .txt file.

Deletion of entries and entry fields, in turn, may be performed only manually within the database, for example if an entry turns out to be no more valid due to a change of a term within the entire Polish legal system (it is often the case with names of ministries and institutions).

7.9.1.3 Estonian Legal Language Centre – Process requirements

A priority is that a new term should be included in the term database as soon as possible after the terminologist has spotted it. The new entry need not be complete at that moment.

Changing existing terms is not encouraged, as this would lead to incompatibility of new and old translations, or to revising previous translations.

7.9.1.4 University of Tartu – Process requirements

The management of a term database is the responsibility of the author of the whole dictionary. There is no need for a complicated management system, only a need to periodically transform the original for reviewing, and adding linguistic, formatting and structuring information automatically.

A priority is that a term included in the final term database should not contain terms in its definition part that are missing from the database. That is, the term database represents a “closed world” in its subject field.

7.9.2 *Responsibilities*

Management of a terminology collection is crucial to ensure that the collection is up-to-date and contains correct data and therefore it is also important that management tasks are well structured and distributed among the right employees. In this chapter we will describe the types of employees involved in the different types of management tasks.

7.9.2.1 IATE - Responsibilities

The Data Management Group is responsible for making the decisions. Each institution must appoint a person or a body who is responsible of implementing these decisions and guaranteeing that all terminological work is carried out in compliance with the writing rules.

The responsible terminologist for each language should validate all entries to ensure that terms are reliable and comply with the writing rules.

7.9.2.2 TERMIUM - Responsibilities

The management of terminology entries in the database comprises the following actions:

- Add/Update/Delete
- Link
- Manage
- Import/Export
- Share
- Extract
- Convert

Each activity can be assigned to a person or group to carry out having the responsibility for that particular task, the responsibility assignment ensures that all work is performed consistently. Depending on the task type, the requested qualifications may vary, thus a responsible person may be a terminologist, database designer/developer or analyst, etc.

7.9.2.3 TSK - Responsibilities

In order to ensure a proper management of work and information flow the responsibility lies with one terminologist per project and one terminologist per term bank.

7.9.2.4 PolTerm - Responsibilities

It is the editor responsible for managing the working procedures who is also responsible for the reliability of the database.

7.9.2.5 Estonian Legal Language Centre – Responsibilities

There is one terminologist per translation project plus one terminologist for legal terminology.

7.9.2.6 University of Tartu – Responsibilities

There is one or two authors per one subject field (and consequently, dictionary), who are responsible for the content of the dictionaries. Computational linguists are responsible for automatic transformation of the source dictionaries into formatted, nearly ready-to-print versions.

7.9.3 *Search and retrieval options*

According to the size and purpose of the term collection and the organisation's process requirements management tasks may be more or less extensive and the types of user requirements may be more or less numerous. The complexity and number of search and retrieval options will reflect the complexity of management tasks and the number of different user requirements.

7.9.3.1 TERMIUM - Search and retrieval options

There are developed several linguistic and terminology tools that access information from the TERMIUM database, such as TERMIUM ® on CD-ROM, TERMIUM Plus ®. The search and retrieval options differ accordingly, e.g. as regards which language or language combination (English, French, Spanish) can be searched, which subject fields and/or information types than can

be retrieved, etc. Further, the user type (e.g. translators, terminology vendors, standardisation or federal bodies, etc.) is also a defining factor as regards the access level to search and retrieval options.

7.9.3.2 TSK - Search and retrieval options

Search and retrieval can be performed at two levels as follows. In working files can any word or part of word be used, whereas in term banks: words which are entered in term fields viz. preferred term and synonyms (abbreviations are treated as synonyms).

7.9.3.3 PolTerm – Search and retrieval options

Basing on consistency-oriented translation and terminology management model in a CAT tool framework underlying the working procedures in TEPIS, there are several stages of said procedures at which relevant terminology search and retrieval options are exercised:

1. The pre-translation stage.

During this stage all amendments to the existing Acts or entirely new Acts are prepared to be translated into English. To this end, first they have to be automatically processed in the PolTerm platform. The terminology database looks for any matches between the terms in the new texts and those stored in the database, irrespective of their identification codes. If it finds any matches, they are automatically inserted in the word-editor text as proposed ready-made translations.

2. The PolTerm database update stage.

Once added (see 7.1.9.2 above), the new terms can be automatically retrieved in future translations on the basis of any criterion attributable to recorded terminological information relating to them and grouping one or a certain number of records, for example the Act-identification code, or the “basic” feature assigned to many terms. Obviously, any subset of previously entered terminological records is also retrievable.

7.9.3.4 Lithuanian language term base - Search and retrieval options

Overall in the term base one can seek terms according to 21 characteristics of the term. After subject field selection terminological data can be found according to 8 characteristics:

1. to word or sequence of letters
2. to abbreviation
3. to symbol
4. to definition words
5. to synonym
6. to directive synonym
7. to variant
8. to equivalent of other language term.

Advanced search can be performed according other 4 characteristics of terms:

1. Sphere of usage
2. Derivation

3. Availability
4. Accentuation.

In separate field data retrieval is possible according to 3 characteristics of complex terms. Primarily complex terms can be found according to the head component or the secondary component. It is also possible to pick complex terms with:

- a) synonym
- b) directive synonym
- c) formula.

Finally it is possible to find: a) double-component term, triple-component term or any number-component term; b) single-letter term, double-letter term or any number-letter term.

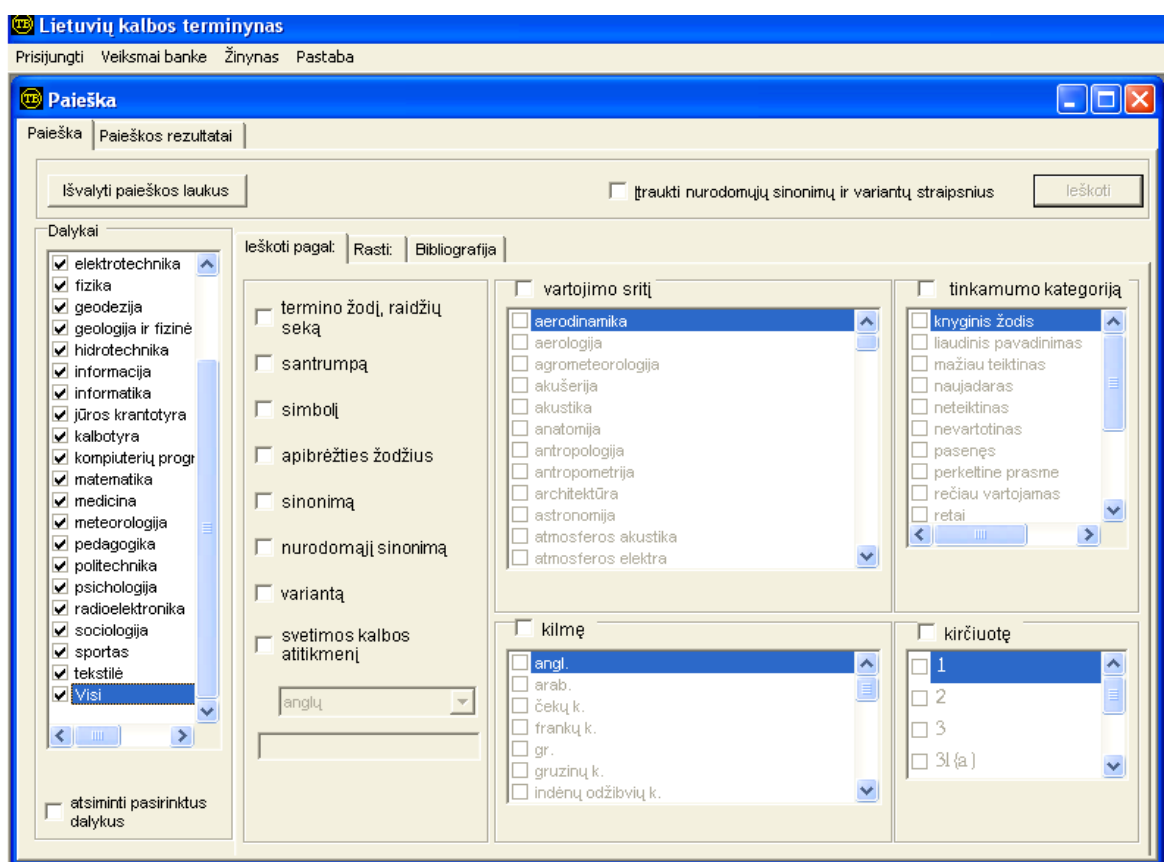


Figure 7.6. Data retrieval

7.10 Validation workflow

Proofreading is often necessary to ensure correctness of a terminology collection. A validation workflow is in some organisations a very informal process just involving proofreading by a colleague. In other organisations validation is a process with several specific steps in a fixed framework where different types of information are checked and validated by different types of employees

7.10.1 Process requirements

An organisation's process requirements in relation to validation will determine the type and complexity of a validation workflow. Examples of process requirements could for example be 'fast and simple', 'adequate and accurate' or 'efficient'. If an organisation's primary process requirements are fast and simple the validation will probably be very informal or there will be no validation at all. Conversely, if the primary requirements are adequate and accurate the validation stages will probably be well defined.

7.10.1.1 IATE - Process requirements

Bearing in mind that the IATE term database is a union set of term collections provided from 10 different translation institutions and bodies which in size vary profoundly, it is obvious that in order to meet the different institutional specific needs, a very flexible validation system is required. So the implementation of the IATE validation system allows each institution to decide how complex its validation process shall be. In IATE the validation process can consist of up to nine stages but if necessary one single phase can constitute the validation system of an institution. An example of the latter case is the European Investment Bank who due to limited human resources in the terminological work is forced to perform a very simple in-house validation routine in which a colleague will make a first validation before the entry will be moved on to an inter-institutional term expert group for further examination. In order to make the users aware of entries not being validated fully, these are tagged with a status information indicating the assessment of the entry's reliability

The Council, on the other hand, having a large terminological division will be able to meet the more resource demanding adequate and accurate criteria cf. above by including five steps in the validation process and thus having the possibility of providing the basis for an in-depth review of data.

7.10.1.2 TSK - Process requirements

The steps of the validation process carried out at TSK are based on the relevant ISO/TC 37 standards.

7.10.1.3 INTERA - Process requirements

The validation of automatically extracted terms was carried out by native speakers who were experts within the field and structured as follows:

The lists of candidate English terms and their corresponding translations in the other languages (lists of single-word terms and multiword terms as produced by the tool) were presented to the human validators.

The validators' task consisted in examining the lists and marking bilingual pairs of terms (English – their mother tongue) as correct, based on the following criteria:

(a) the pair is indeed a term of the specific domain (and not a general vocabulary word) AND

(b) the translational equivalence is also correct.

The terms identified as correct pairs were further lemmatized (single word terms and multiword terms), and the final lists produced by the validators were used for the production of the multilingual terminological entries.

7.10.1.4 PolTerm – Process requirements

A partial validation process is carried out simultaneously with preparation of a new set of texts to be translated into English as any flaws in the data retrieved during the automatic processing stage are corrected. The validation process concerning the entire database is yet to be carried out.

7.10.1.5 Estonian Legal Language Centre – Process requirements

Validation is carried out during the process of translating the document where the new term has first appeared.

7.10.1.6 University of Tartu – Process requirements

The dictionaries are systematically validated at the following stages:

1. After the selection of the terms has been completed
2. After the authors have finished writing the definitions
3. After the final formatting has been completed (before sending the files for printing)

When half of the terms have been selected, when half of the definitions have been written, the project leader occasionally looks at the dictionaries and points to deficiencies he noticed. During the whole project, a terminologist keeps an eye on the dictionaries.

7.10.2 *Actors*

Different actors can be involved in a validation workflow, for example translators, terminologists and subject field experts. Dependent on the purpose of the term collection, the types of employees and the process requirements the expertise of these people can be utilized differently. In this section we will define the types of actors involved in validation in the various organisations and describe their responsibilities.

7.10.2.1 IATE - Actors

A key information of the IATE validation procedure is the usage of the 'user profile' information type, containing the information a.o. about e-mail and postal address but above all in this context, the competencies of the people involved in the IATE terminological work. These competencies are represented as roles such as translator, expert translator, domain expert and terminologist.

The roles assigned to the user are used to manage the validation process in IATE. The validation stages mentioned above are thus associated with specific roles. The role of terminologist is thus linked to the final consolidation of the validation process and after this phase the entry in question will be assigned the highest possible reliability status - finally validated.

In order to illustrate the IATE concept of various validation cycles consider the following very simple validation workflow:

- An entry is created or modified in the termbase (e.g. by a translator)

STAGE 1: Formal correctness (e.g. spelling) of each field is checked (e.g. by an expert translator)

STAGE 2: Contents of each field are checked and new information is possibly added (e.g. by a subject field expert)

STAGE 3: The entire entry is checked in terms of contents and coherence (e.g. by a terminologist)

7.10.2.2 TERMIUM - Actors

Terminological records of TERMIUM are continuously validated and updated (approx. 100.000 pro year) in cooperation between translators, terminologists, proofreaders subject-field specialists. Professional users' comments are also part of the validation of the material. Standardised terminology in English and French is progressively entered into the termbank by experts of term standardisation esp. as regards federal government terminology.

7.10.2.3 TSK - Actors

The validation work involves assistance of terminologists and subject field specialists.

7.10.2.4 PolTerm: Actors

The occasional validations are the responsibility of the editor in charge of the database management.

7.10.2.5 Estonian Legal Language Centre – Actors

Validation is carried out mainly by terminologists and translators. The publicly available version of the term database in Internet contains a feedback form for users who would like to report on the terms.

7.10.2.6 University of Tartu – Actors

Systematic validation is carried out by external reviewers. Occasional validation is carried out by the project leader and a professional terminologist.

7.10.3 *Feedback mechanisms*

Newly created, updated or even old entries in a terminology collection often give rise to comments and suggestions from all the users of the terminology collection. These comments can be very useful, but it is of course a prerequisite that they are directed to the right people. In some organisations feedback mechanisms are very informal and a user will just drop a hint when convenient or send an email. In other organisations feedback mechanisms are fairly formalised and may even constitute part of the database system.

7.10.3.1 IATE - Feedback mechanisms

In order to exploit the possibility of getting feedback from public users of the IATE term base, a communication channel has been implemented in IATE. The database users may come across

entries that they think should be commented on. Realising that such comments can prove very beneficial when directed to the right people, IATE offers the possibility to the external data base users that can 'mark' an entry. These marks can be attached to each entry and sent to individual users or a terminology group of an institution specific division. The information could be that two entries represent the same concept and therefore should be merged.

7.10.3.2 TSK - Feedback mechanisms

The term entries are circulated previous to their approval for comments by subject field specialists. The normal procedure followed is consultation by e-mail.

7.10.3.3 Estonian Legal Language Centre – Feedback mechanisms

The publicly available version of the term database in Internet contains a feedback form for users who would like to report on the terms.

7.10.3.4 University of Tartu – Feedback mechanisms

The reviewers write their comments on paper print-outs of the manuscripts they receive.

8 Conclusions

The structure of this report is the result of a cooperative effort conducted by all project partners. In an iterative process the outline of the report has been discussed and agreed on. An example of this refinement process was the inclusion of chapter 3 describing organizations that operate within the area of terminology work in the new EU-member countries involved in the project.

The contents of this report are also based on common efforts. Each project partner has thus given input to the relevant parts of the project. In broad terms, the project partners representing the new EU-countries have primarily provided information on terminology work in their own countries while the project partners from Germany and Denmark have contributed by describing more internationally and language independently oriented terminology work.

As described in the technical specifications of the EuroTermBank project, the aim of this report was to establish a knowledge platform on both a theoretical and practical level in order to pave the way for aligning methodologies and standards also with respect to designing and implementing data exchange mechanisms and procedures.

We consider this report as a very practical reference work which will undoubtedly prove very beneficial in terms of being a solid knowledge platform in future terminology work.

Not only for the next report in Work package 1 in which recommendations for how to conceive, collect and harmonize terminology material will be given, but also seen in a more long term perspective. When the information of this report is eventually released to a broader audience it is foreseen that it will constitute a well-founded knowledge basis for improving and harmonizing terminology work and infrastructure in other new EU member states.

9 Appendices

9.1 ISO standards

9.1.1 ISO 704: Terminology work – Principles and methods

This standard gives definitions of all the pivot concepts within terminology such as *objects*, *concepts*, *concept relations*, *concept systems*, *definitions* and *designations*, i.e. it constitutes the foundation of all basic terminology work.

Firstly, *objects* and *concepts* are defined. *Concepts* are divided into *individual* and *general*. Then the *characteristics* (essential, delimiting) of a concept are explained, defining *intension* and *extension*. *Concept relations* are described thoroughly, divided into *hierarchical*, of which there are two kinds: *generic* and *partitive*, and *associative*, and the nature of *concept systems* including advice on how to develop them is also treated.

Subsequently, *definitions* are handled. They are divided into *intensional* and *extensional* and there is an exposition of the principles for writing definitions. Key words are *reflecting the concept system*, *conciseness* and *avoiding deficient definitions* (*circular*, *incomplete*, *negative*). *Notes* for secondary information and *graphic representations* are introduced.

Designations are the representation of a concept in natural language. Designations can be *terms* (one or more words) that designate general concepts, *appellations* that designate individual concepts and *symbols* that designate both individual and general concepts. The standard explains homonymy, synonymy and treats the general principles for term formation.

Finally, there is a brief introduction to some *standardization* issues, explaining *preferred*, *admitted* and *deprecated* terms and the reasons for *deprecation* of terms.

An appendix gives examples of *term-formation methods* in English, i.e. new forms formed e.g. by derivation, compounding or abbreviation, and existing forms transformed into terms e.g. by changing syntactic category.

9.1.2 ISO 860:1996 Terminology work – Harmonization of concepts and terms

Harmonization of terminology is a highly relevant issue for both monolingual and bi- or multilingual communication since the aim is to minimize the terminological difficulties of communication. This standard handles the process of harmonization of terms and concepts from the preliminary analyses of subject fields and concept systems to the construction of harmonised concept systems, definitions of harmonised concepts and harmonization of the individual terms.

Before a *concept harmonization* can be carried out, several analyses have to take place. An analysis of the *subject field* is obligatory since subject fields that are well established, with a tradition of standardization or which deal with concrete objects are more likely to result in successful harmonization than subject fields under development, or subject fields within humanities or social

sciences and with no tradition for standardization. If the chances for a successful harmonization seem good, a preliminary *analysis of the concepts and their characteristics* has to be carried out.

The harmonization process is as follows:

- It starts with a *comparison of the involved concept systems* in terms of number of concepts, relations between concepts, depth of structure and type of characteristics leading to the *construction of harmonised concept systems*.
- All the concepts must then be analysed by *comparing the definitions*. If the definitions differ, it must be decided whether the difference is relevant or irrelevant. If relevant, it means that there are indeed two or more different concepts involved that must be defined and placed in the concept system.
- The *defining characteristics* for the harmonized concepts have to be established.
- When the concepts are harmonized, the *terms can be harmonised* taking into account the differences and similarities between languages, the tradition of term formation in the subject field and in a given language as well as the already established terminology.

The standard includes a flow chart of the preliminary analyses and the harmonization process.

9.1.3 ISO 10241: International terminology standards – Preparation and layout

This standard gives a practical introduction to how to write an international standard for terminology. The aim of an international standard on terminology is the harmonization of concepts, concept systems and terms in a given subject field in different languages. This standard deals with the procedure for developing standards whereupon some principles for terminography are given.

The preliminary work consists of defining the target group, delimiting the subject field defining the scope and the sub-fields and analyzing the terminological usage of the selected sources (incl. evaluation of sources). The work should be carried out simultaneously for all languages involved and the number of terms dealt with should be limited.

All possible terminological data should be collected and recorded, extracting all relevant material for term lists, concepts and definitions in one operation from the source data. From this material, first the term lists for each language should be established after which the concepts and their relations should be specified establishing concept systems. Finally a comparison and harmonization of the language specific concept systems should be performed. The last part of the working procedure is to write the definitions in accordance with ISO 704.

The terminography part goes through the important data categories. The essential information is entry number, preferred term and definition; additional information comprises among other things symbols, pronunciation, grammatical information, subject field, references, examples of usage and term equivalents. The preferred order of the entries is systematic order, but alphabetical order as well as mixed order is possible. Alphabetical standards shall contain systematic indexes and vice versa.

9.1.4 ISO 12200:1999 – Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) — Negotiated interchange

ISO 12200 is an international standard for the interchange of terminological data allowing the distinct identification of separate data sets and data categories as well as its' dependencies and relations. MARTIF has been elaborated in order to facilitate more universal, less costly exchange of data collections containing concept-oriented terminology entries. The format relies heavily on the data category names and definitions contained in the companion standard ISO 12620 (see chapter 9.5). MARTIF is based on Standard Generalized Markup Language (ISO 8879, SGML) and was originally developed in close cooperation with the Text Encoding Initiative (TEI) and the Localisation Industry Standards Association (LISA).

As an SGML-based solution, MARTIF has the additional advantage that terminological data can be easily processed like any other SGML document, e.g. for the publication of printed terminological glossaries. Due to its high degree of flexibility MARTIF is able to adequately represent all forms and structures of terminology resources.

MARTIF not only provides an open, flexible mechanism for exchanging data with other potential users employing different terminology management systems. It can also be used when companies need to change or upgrade software from one database format to another.

The main body of the MARTIF standard specifies the formalism to be used in preparing terminology data collections for interchange by defining the SGML Document Type Definition (DTD) and listing the appropriate tags (markup) used to structure the data. Normative Annex A of the standard specifies the markup for the individual terminological data categories to be used in the MARTIF environment, based on ISO 12620.

A complete MARTIF document consists of a prolog, followed by a document instance of type MARTIF. The document instance consists of a `<martifHeader>` followed by the text, which in turn consists of optional front matter, the `<body>` (a sequence of terminological entries), and optional `<front>` and `<back>` matters.

Basic components of a MARTIF document:

- | |
|---|
| <ul style="list-style-type: none">I. PrologII. Document instance (<code><martif lang=en></code>)<ul style="list-style-type: none">A. header (<code><martifHeader></code>)B. text<ul style="list-style-type: none">1. front (optional)2. body<ul style="list-style-type: none">a. first terminological entry <code><termEntry></code>
(minimum of one)b. second terminological entry <code><termEntry></code>c. etc. (additional terminological entries)3. back (optional) |
|---|

Structure of the document instance:

```
<martif lang=en>
<martifHeader>
... (The header goes here.)
</martifHeader>
<text>
  <body>
    ... (The terminological entries go here.)
  </body>
  <back>
    ... (Included bibliographic entries go here.)
    ... (Any external references (<xref>s) also go here.)
  </back>
</text>
</martif>
```

Example for a full MARTIF term entry:

```
1 <termEntry id='ID000073578'>
2   <descrip type='subjectFieldLevel1'> appearance of materials </descrip>
3
4   <ntig lang=en>
5     <termGrp><term>opacity</term>
6     <termNote type='partOfSpeech'>n</termNote></termGrp>
7     <descripGrp><descrip type='definition'>degree of obstruction to the transmission of
8     visible light</descrip><ptr type='sourceIdentifier' target='ASTM.E284'></descripGrp>
9     <adminGrp><admin type='responsibility'>E12</admin> </adminGrp>
10    </ntig>
11
12    <ntig lang=de>
13      <termGrp><term> Opazit&auml;t</term>
14      <termNote type='partOfSpeech'>n</termNote>
15      <termNote type='gender'>f</termNote></termGrp>
16      <descripGrp><descrip type='definition'>Ma&szlig; f&uuml;r die
17      Lichtundurchl&auml;ssigkeit </descrip><ref type='sourceIdentifier' target='DIN-
18      6730.1996-05'>p. 383</ref></descripGrp>
19      <adminGrp><admin type='responsibility'>Normenausschu&szlig; Papier und Pappe
20      (NPa) im DIN Deutsches Institut f&uuml;r Normung e. V. </admin></adminGrp>
21    </ntig>
22    <ntig lang=fr>
23      <termGrp><term>opacit&eacute;</term>
24      <termNote type='partOfSpeech'>n</termNote>
25      <termNote type='gender'>f</termNote> </termGrp>
26      <descripGrp><descrip type='definition'>rapport du flux lumineux incident au flux
27      lumineux transmis ou r&eacute;fl&eacute;chi par un noircissement
28      photographique</descrip>
29      <ptr type='sourceIdentifier' target='HJdi1986'></descripGrp>
30      <adminGrp><admin type='responsibility'>C.I.R.A.D.</admin> </adminGrp>
31    </ntig>
32  </termEntry>
```

As noted above, MARTIF was originally designed for so-called *negotiated interchange*, where partners examine each other's data before interchange and make decisions about preconditioning

the data before importing it from the interchange format. This approach allows a high degree of flexibility in individual applications.

For a more “blind” interchange, specific MARTIF-compatible formats can be defined on the basis of ISO 16642 (see chapter 9.6). Following the structure of MARTIF the XML-based standard format for terminological data TermBase eXchange (TBX) has been developed (see chapter 7.8.1).

9.1.5 ISO 12620:1999 – Computer applications in terminology – Data categories

Terminological data are collected, managed, and stored in a wide variety of environments. For purposes of storage and retrieval, these data are organized into terminological entries, each of which traditionally treats information associated with a single concept. Data items appearing in individual terminological entries are themselves identified according to data category. Differences in approach and individual system objectives inevitably lead to variations in data category definition and in the assignment of data category names. The use of uniform data category names and definitions, at least at the interchange level, contributes to system coherence and enhances the reusability of data.

The International Standard ISO 12620:1999 specifies data categories for recording terminological information in both computerized and non-computerized environments and for the interchange and retrieval of terminological information independent of the local software applications or hardware environments in which these data categories are used.

The data category specifications are divided into three major groups: data categories for terms and term-related information, descriptive data, and administrative data. The groups are further subdivided into ten sub-groups.

1. Term and term-related data categories:

- Subgroup 1 consists of the data category term and contains a term or other information treated as if it were a term (e.g., phraseological units and standard text).
- Subgroup 2 specifies data categories for term-related information.
- Subgroup 3 specifies data categories for information relating to equivalence between or among terms assigned to the same or very similar concepts.

2. Descriptive data categories:

- Subgroup 4 specifies data categories for the classification of concepts into subject fields and subfields, along with other classification-related information.
- Subgroup 5 specifies data categories for concept-related description, i.e., different kinds of definitions, explanations and contextual material provided to define or otherwise determine the subject field and concept to which a term is assigned.
- Subgroup 6 specifies data categories for indicating relations between pairs of concepts.

- Subgroup 7 specifies data categories used to express the position of concepts within concept systems.
 - Subgroup 8 specifies the data category *note*. This category stands alone because it can be associated with any one of the other categories and therefore cannot be subordinated to any other specific subgroup.
3. Administrative data categories:
- Subgroup 9 specifies data categories for documentary languages and thesauri.
 - Subgroup 10 specifies data categories for all other strictly administrative information.

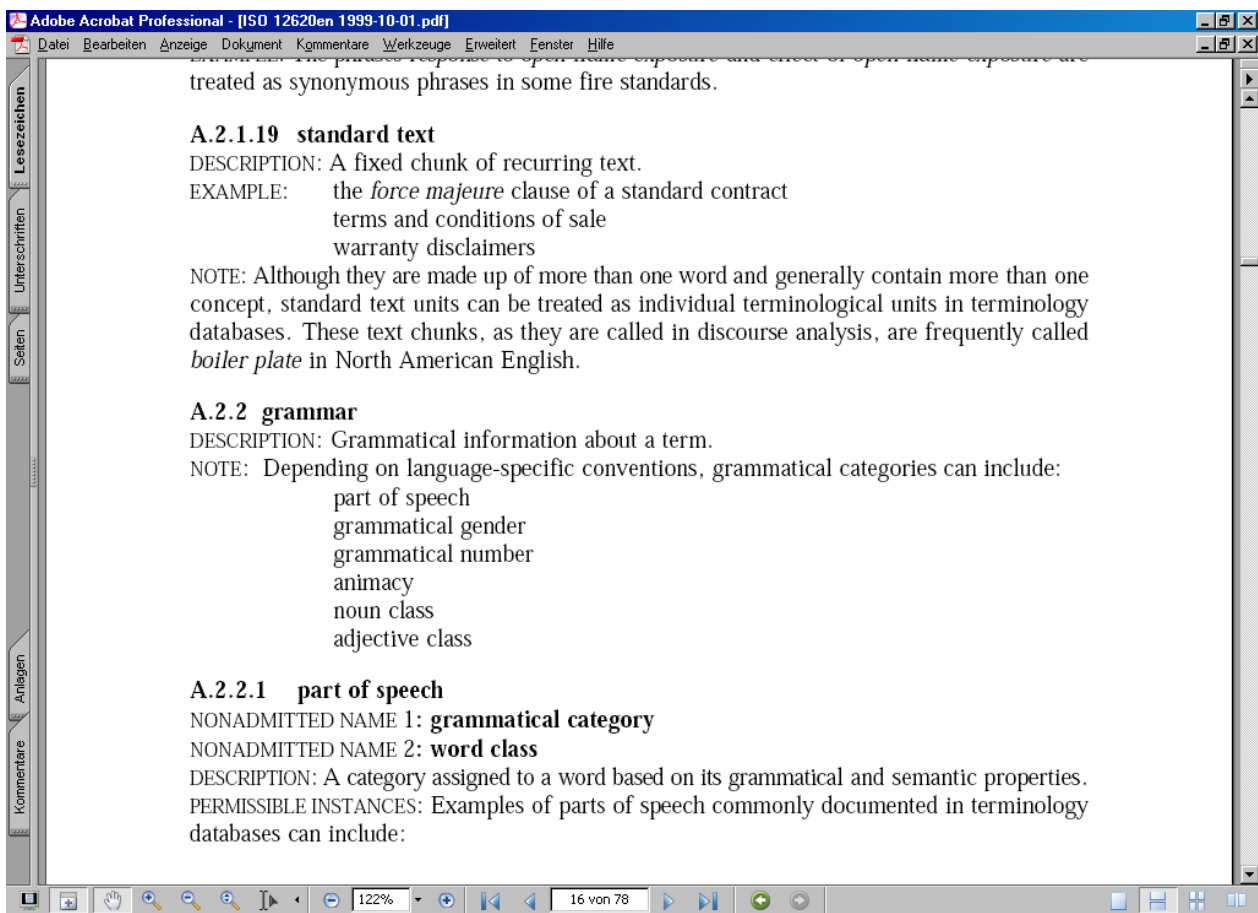


Figure 9.1. Example for the description of data categories in ISO 12620

If applied for the purpose of interchanging machine-readable terminology, it is recommended that this standard is used in conjunction with ISO 12200 (see chapter 9.4), although it can also be used for modelling terminological information independently of computer applications.

The data categories specified in ISO 12620 constitute the basis for various other standards dealing with the processing of terminological data, e.g. ISO 12200. The data categories correspond to data element concepts in the ISO/IEC 11179 series of standards.

International Standard ISO 12620 was prepared by Technical Committee ISO/TC 37, *Terminology (principles and coordination)*, now called *Terminology and other language resources*, Subcommittee SC 3, *Computer applications for terminology*.

At present TC 37 is revising ISO 12620 and planning to publish the new version entitled *Computer applications in terminology -- Data categories -- Model for description and procedures for maintenance of data category registries for language resources* in November 2006. This revision implicates quite fundamental changes in the handling of data categories, meaning that in the future, the data categories will be maintained in a universal Meta Data Registry available in the Internet, whereas ISO 12620 will provide all the specifications indispensable for the description of data categories, as well as the admission procedure for new data categories to be added to the Data Category Registry. In this context, the foundation of a new subcommittee of ISO/TC 37, SC 4 *Language resource management*, will grant access to collections of data categories beyond those defined by SC 3, e.g. the language codes defined in ISO 639-1 and ISO 639-2.

9.1.6 ISO 16642:2003 – Computer applications in terminology - Terminological markup framework (TMF)

This international standard has been developed to facilitate the use and re-use of terminological data collections, taking into account the real-live conditions of different formats, database environments and term-bank systems as well as the various data models the collections are based on. The standard is also motivated by the need to provide better connections between terminological databases and other lexical resources dedicated, for instance, to machine translation or natural language processing.

Core element is a single high-level meta model representing a unique information structure shared by all terminology markup languages (TML), which decomposes the organisation of a terminological database into basic components – the structural skeleton, defined as a “set of XML elements which, in a given TML, results from the expansion of the meta-model”, and the elementary units of information (i.e. data categories) that can be attached to the structural skeleton.

For mapping any given format, or TML, onto the abstract components of TMF, a simplified XML application have been defined. This format called GMT (Generic Mapping Tool) is based on a reduced set of XML elements and attributes, which serve as containers for nodes of the structural skeleton (identified by <struct> tags) and data categories (identified by <feat> tags).

Thus, the data of a terminology data base expressed in any format is mapped onto a given data model using GMT, by

- 1) decomposing every entry into the three structural levels of the meta-model, the Terminological Entry (TE), the Language Section (LS) and the Term Section (TS) (<struct> element); and
- 2) expressing each information unit by means of the <feat> element where the type signifies the data category name.

To illustrate how a terminological data collection can be analysed as an abstract structure, a simple terminological entry expressed as an XML document conformant to MSC (MARTIF with Specified

Constraints, a variant of ISO 12200) specifications is mapped via the TMF meta-model onto data categories defined in ISO 12620 (see chapter 9.5).

Terminological entry as expressed in MSC:

```

<?xml version="1.0"?>
<martif type="MSC" lang="en">
  <text>
    <body>
      <termEntry id="ID67">
        <descrip type="subjectField">manufacturing</descrip>
        <descrip type="definition">A value between 0 and 1 used
        in ...</descrip>
        <langSet lang="en">
          <tig>
            <term>alpha smoothing factor</term>
            <termNote type="termType"
            datatype="picklistVal">fullForm</termNote>
          </tig>
        </langSet>
        <langSet lang="hu">
          <tig><term>Alfa ...</term></tig>
        </langSet>
      </termEntry>
    </body>
  </text>
</martif>

```

This document can be mapped to the abstract model by identifying a structural skeleton corresponding to the meta-model and by associating the corresponding information units with each structural node in the structural skeleton, as shown below.

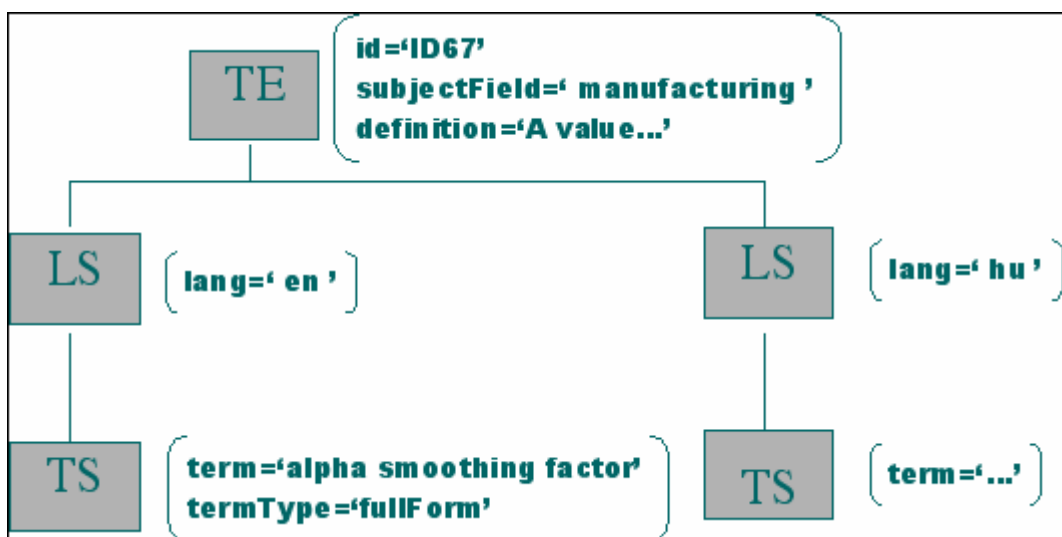


Figure 9.2. Mapping a document to the abstract model

The data categories correspond to the data categories specified in ISO 12620 as follows:

Data category	ISO 12620 number	ISO 12620 name
id	A10.15	entry identifier
subjectField	A04	subject field
definition	A05.01	definition
lang	A10.07.01	language identifier
Term	A01	term
termType=fullForm	A02.01.07	full form

One possible encoding in the GMT format:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<tmf>
  <struct type="TE">
    <feat type="entry identifier">ID67</feat>
    <feat type="subject field">manufacturing</feat>
    <feat type="definition">A value between 0 and 1 used in ...</feat>
    <struct type="LS">
      <feat type="language identifier">en</feat>
      <struct type="TS">
        <feat type="term">alpha smoothing factor</feat>
        <feat type="term type">fullForm</feat>
      </struct>
    </struct>
  </struct>
  <struct type="LS">
    <feat type="language identifier">hu</feat>
    <struct type="TS">
      <feat type="term">Alfa ...</feat>
    </struct>
  </struct>
</tmf>
```

The combination of the meta-model and a given Data category specification (DCS) is enough to define the degree of interoperability of two TMLs, encompassing its full informational properties from a terminological point of view. Any information structure that corresponds to such conditions has a canonical expression as an XML document using the GMT (Generic Mapping Tool) representation. The interoperability between two different TMLs depends solely on their compatibility at that level.

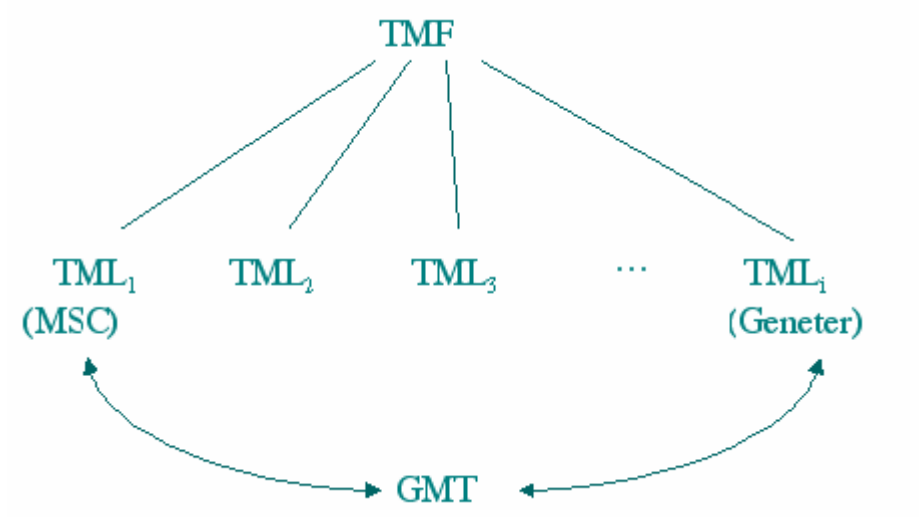


Figure 9.3. Interoperability between two TMLs using GMT

When two TMLs are based upon two different DCSs, GMT provides a framework for identifying what information can be transformed between one format and another and what will be lost during the transformation.

The comparison between two TMLs is only possible if there is a central repository of data categories, associated with a consistent model for these, which can act as a broker between any two formats. For the application of this standard, ISO 12620 forms a reference Data Category Registry (DCR) for any information unit to be used in the definition of a TM.

9.2 Principles in Latvian terminology work

In this section we describe first some theoretical principles for the terminology work in Latvia, and then the experience of applying these principles in the development of Latvian terminology.

9.2.1 Theoretical principles for terminology work in Latvia

The basic theoretical principle of terminology development is the consideration of the requirements placed upon a scientific term: consistency, the exactness of meaning, the brevity of form, monosemy, mononimity, contextual independence, emotional neutrality, etc. (see works of E. Wüster, D. Lotte, H. Felber, H. Picht, etc.).

In Latvian terminology theoretical principles are divided into three main groups (Skujina, V. The Principles of Formation of Latvian Terminology. In Latvian. Riga, 2002:213-215).

9.2.1.1 Intralinguistic principles

Taking into account that terminology has close links with linguistics, lingual (intralinguistic) principles are dominating in the formation of terminology. Namely, these are principles with respect to the lexical aspect (the place of a term in a word-stock system of LGP), the semantic aspect (the

differentiation of the meaning of terms and elements of terms, the separation of the semantic functions of terms of different structures, undesirability of synonyms, etc.), the phonological and morphological aspect (the use of spelling and form variants for differentiation of term content, role of structural semantic analogy for creating new terms, etc.), the word-building aspect (with respect to the formation of terms according to the general rules for word-building, atomization of word-elements of small productivity and non-productive elements, and making new patterns, for instance, using international word-elements for creation of new terms in a national language, etc.) and the syntactic aspect (with respect to patterns of phrase-terms built from two or more words, role of abbreviations in the function of term, etc.).

Some specific principles can be mentioned in connection with the linguistic approach.

Principle of semantic correspondence

The aim of the semantic correspondence is to make a term suitable in the terminological system according to its meaning. Thus one must be aware that in word formation each affix has a particular semantic meaning, characteristic of definite language system. For example, meanings of the Latvian suffixes -*šana* and -*ums* have to be distinguished: *apliecinājums*, *apgalvojums*, *pasvītrojums* denote a result of a process, while the same derivatives with the suffix -*šana* denote an active part of the process: *apliecināšana*, *apgalvošana*, *pasvītrosana*. Derivatives ending on -*nis* denote tools or means, for example, *dzinis*, *veltnis*, *vednis*, *slīdnis*.

Semantic correspondence from the literary point of view is connected with the criterion of back translation. For example, *mouse* - *pele*, *worm* - *tārps* etc. Transfer of such words must be performed very carefully, since it involves semantic calques – borrowing of the meaning according to the pattern of another language.

The question of metaphors also has to be considered. Metaphors differ from one language to another. They can coincide in several languages, but rather often due to different cultures they do not. That burdens the development of terms, since de-metaphorisation of a term in Latvian makes it to lose the link to the corresponding term in the language it comes from. Then it is difficult to recognise the term and that in its turn complicates the job of translation. That is the reason why one of the guidelines of creating IT&T terminology is the criterion of back translation when the back translated term must coincide with the initial term in the original language. It can be achieved by linking meanings of common words, for example, *pele* will undoubtedly be *mouse* and vice versa. These words have long existed in Latvian, but new meanings have appeared which, depending on the semantic system, can enrich the language or on the contrary – pollute it.

The Calque problem is a very frequent one in terminology. Nevertheless in creating Latvian terminology the main requirements are the correspondence between the term and the concept it conveys, and the consideration of word formation rules and structurally semantic models existing in the language. If a particular English metaphor does not convey the main concept of the term in Latvian then the concept should be evaluated and incorporated in the Latvian term. E.g. the term *cookie* has two different Latvian equivalents *sīkdatne* and *ripa*, none of them corresponding to the meaning of the English word *cookie*.

Principle of functional correspondence

A term is functionally qualitative if it meets the requirements, rules and models of definite language system. New forms, new words and new syntactic units are formed on the basis of stable models. Terminology observes spelling and pronunciation norms, rules of morphology, word formation, syntax and semantics of literary language.

To ensure the requirements of functionality of a term its brevity, convenience of usage and euphony have to be considered.

Brevity and convenience of usage

Short terms are more favourable as they are easy to use and to fit the system. Such terms also permit attached elements, thus allowing creation of subconcepts. Then the usage of suffixal endings will give as a result new proper term, for example, *aizķere* (*hook*), *aizkave* (*delay*), *apraide* (*broadcast*).

One should avoid long terms, for such words do not fit easily either in text or speech and they rarely permit further derivatives. Short word forms as *datne*, *atteice*, *aizķere* are preferred in the terminology.

The most productive method in the IT&T terminology is formation of compounds. From the functional point of view they are created to shorten a denotation of multicomponents and instead of a prepositional phrase a possessive phrase is obtained which is better context–fit. Many new terms are created as compounds, for example, *atpakaļsaderīgs* (*backward compatible*), *augšupsaderīgs* (*upward compatible*), *darbderīgs* (*off-the-shelf*), *skārienjutīgs* (*touch-sensitive*), *datorkopne* (*common bus*), *dublētājdatne* (*backup file*) etc.

Euphony

When borrowing a foreign word or introducing an internationalism euphony is an essential aspect, and the IT&T terminology commission is paying special attention to their euphony. The Latvian language is considered to be a melodic language where vowels and consonants stand in sufficient balance. This balance should not be destroyed by polluting the language with uncharacteristic sound combinations as previously mentioned combination *pjū* in the attempt to translate the term *computer* in Latvian as *kompjūters*.

A part of borrowings are used successfully, and their implementation is not conflicting with the guidelines, for example, the terms *baits* (*byte*), *komparators* (*comparator*), *translators* (*translator*) fit easily in the created system of terminology.

However, the perception of euphony is quite a subjective quality and as such should not play the primary role.

Principle of term dissemination

Before the creation of a new term its usage is evaluated. Due to the appearance of the term in common language IT&T terminology commission determines to which usability level this term will correspond.

Special attention is paid to creation of terms which appear frequently in common language – broadly used terms. They have to be short, explicit and euphonious and have to meet all the set of guidelines. If the term is used seldom the requirements can be reduced as these terms are often used only by professionals of this specific branch.

Principle of tradition

If the term is already broadly used and is approved by its users it should not be changed without sufficient reason. The history of IT&T terminology compared with terminology of other branches is not very long, nevertheless there are terms which have been transferred to Latvian by transliteration, for example, *interfeiss* (*interface*), *fails* (*file*), and changing them into proper Latvian terms *saskarne* and *datne* is a rather difficult task. As we had mentioned before some of the terms which had had a long history and frequent usage we did not change because they were not also contradictory to some of the guidelines.

9.2.1.2 Interlinguistic principles

Nowadays terminology of any language develops in close contact with terminology of other languages. It causes the necessity to achieve a solution on two levels:

- conceptual level;
- form level.

On the conceptual level, the main principle is the harmonization of the notional content of corresponding term-equivalents in contact languages. The concept expressed with a corresponding term has to be the same.

On the form level, in many cases it is necessary to solve issues connected with the recommendable and non-recommendable borrowings. Borrowings are different regarding their origin and dissemination. Some borrowings, especially those of Latin or Greek origin, are the so called internationalisms (with the same or similar spelling, pronunciation and meaning in different languages). They are widely used in a number of different European languages. Other terms are borrowed from a particular language and are individual borrowings from one language into another.

In Latvian terminology practice estimating the usefulness of borrowings:

- preference is given to terms borrowed from Latin or Greek, or formed (in different languages) on the basis of Latin and/or Greek word-elements;
- more strict criteria are used for borrowings from a particular national language with specific phonetic, morphological or other peculiarities of their own.

It is recommended to avoid non-motivated borrowings.

9.2.1.3 Extralinguistic principles

The success of the terminology work depends on a number of extralinguistic subjective and objective factors:

- the personality and the linguistic, terminological and respective subject-field competence of a term-creator (developer);
- the sociolinguistic attitude towards new terms, both neologisms and new borrowings;
- the sociopolitical processes in a country;
- the attitude of the government to terminology planning, education and development;
- the funding of terminology work, etc.

Many of these factors and the rapid changes of the circumstances including the prevalence of needs for coordinated term-development against the limited possibilities are not favourable for successful terminology work in Latvia.

The main extralinguistic principles being more or less implemented in terminology sphere are:

- the scientific approach to terminology development with respect to the achieved level on the terminology science and research on a world-wide scale;
- the use of high technology for development of unified term systems and information services in some branches;
- the cooperation between subject-field experts and linguists;
- the academic courses on terminology theory and practice in some higher education professional study programs;

- the rendering of methodological and term-advising help free of charge;
- the publication of the decisions of TC of LAS in term dictionaries, mass media including internet portals, a special quarterly brochure „Terminoloģijas Jaunumi” („Terminology News”), etc.

Thus, according to the current model of terminology work in Latvia, in general, an asset is the large number of subject-field experts involved in implementation of uniform terminology work principles. This ensures the inheritance of the terminology already formed. On the other hand, a negative aspect is the insufficient funding, shortcomings in the informational support while the common harmonized database is still in the stage of development, as well as drawbacks in the linguistic and terminology education process, etc.

System or exception, innovation or tradition, coordination between different branches or maintaining the peculiarities of each branch, – those are, on one hand, the natural phenomena and, on the other hand, contradictory ones that have an influence upon the terminology work and further development in Latvia.

9.2.1.4 Supplement: Ernest Drezen’s theoretical conclusions

- Names of concepts have to be unified. Otherwise communication will be burdened.
- If words in different languages have the same meaning and only differing national form (“diverse in form, identical in meaning”), such terms are considered as positive.
- The deeper knowledge have been developed in an appropriate subject-field, the more successful is the formation of terminology .
- The absolute stable classification and unchanging term system is not possible due to continuous development of science.
- Experts of every subject-field are concerned in the terminology development, therefore terminology problems are widely disseminated both on the national and international level.
- Unification of terms is impossible in a society where personal, group or class interests are prevailing common, national ones. Necessity of terminology unification is recognized more distinctively in countries with higher developed engineering.

9.2.2 *Specific requirements for terms – Latvian experience*

The development of Latvian terminology is proceeding in conformity with the world-recognized basic principles of terminological work. The term is interpreted as a functional unit (not a specific substance) with a definite sphere of application and functional meaning.

A term is the result of the process of termination and by “term” we denote the unit of termination – a word or a combination of words that expresses (names and designates) a definite scientific concept in the subject field’s terminology system.

The term is the synthesis of the definition. A term should be as concise as possible. Complex concepts and terms or expressions which can be broken down into various concepts and terms should be avoided. Each term should cover only one concept.

The basic principle of terminology formation (creation) is to consider the requirements placed upon the scientific term: systematicity, the exactness of meaning, the brevity of form, monosemy, mononimity (one concept – one term), contextual independence, emotional neutrality, etc.

The substantive nature of a term and differentiated understanding of the terminological vocabulary have a particular role. A concept of termeme gives for the words of various parts of speech a possibility to allocate a definite place within terminological systems and terminological vocabulary in general. The termeme is a conceptually united aggregate of terminological units with a term as the central member, and words of other parts of speech as secondary members. (More details in: V. Skujiņa “The Specificity of the Term and the Concept of the Termeme” – IITF Infoterm Multilingualism in Specialist Communication. – Vienna, TermNet, 1996. – Vol. 2. – P. 1123–1130.)

Creation of Latvian terminology is based on:

- Latvian vocabulary (word-stock);
- new derivations and compounds;
- borrowings from other languages.

9.2.2.1 Main principles for term definitions

(Compiled on the base of different ISO standards and Latvian practice)

The term and the definition are the main items of the content of the terminological entry. The definition contains the essence of the concept expressed with the term. The essence of the concept – that means the essential – i.e. necessary and at the same time sufficient – characteristic features of the concept. The quality of terminological products, terminological databases including, are determined by the quality of the definitions.

According to the ISO 1087-1:2000(E/F) 3. 3. 1 the definition is: ‘*representation of a concept by a descriptive statement which serves to differentiate it from related concepts*’.

In terminology, different types of definitions are recognized, usually divided into two groups: intensional definitions and extensional definitions. “The greater the extension, the fewer the characteristics which distinguish it; the number of objects included is large. The higher the degree of specialization of concepts, the lower their extension, i.e. the number of objects is small” (Picht, Heribert; Draskau, Jennifer. *Terminology: An introduction*. Copenhagen, 1985, p. 42).

In practice, intensional definitions are preferable to other concept descriptions. Intensional definitions should be used whenever possible as they most clearly reveal the essential characteristics of a concept within a concept system. The intensional definition should be based on the concept relations determined during analysis.

Extensional definitions are to be used only when intensional definitions are difficult to elaborate (ISO 704:2000 (E) 6. 2. 1).

Some terms are so long and complex that they could almost serve as definitions; some definitions are so short they could almost be thought of as terms. In spite of this, the definition should not be confused with the designation.

For **multilingual** terminological databases it is necessary to elaborate **harmonized** definitions. N. B.! (See ISO 860:1996 (E) 5)

Before formulating the harmonized definition, consensus shall be reached on:

- the characteristics that are essential to the intension of the concept;
- the characteristics that are essential for inclusion in the definition.

All versions of a definition in the different languages shall include the same characteristics. The formulation of definitions, however, depends on the rules of the individual languages.

A definition for a terminological entry – according to the ISO 704:2000 (E) 6. 3. 1

A terminological entry shall be composed of a statement explaining what the concept is. The statement is made up of a subject, copula and predicate. The subject is the designation, the copula is understood to be the verb “is” and the predicate constitutes the definition. The verb “is” usually (in term dictionaries, terminological databases) is expressed by typographical means, such as a colon, a dash or by starting a new line of text. Such typographical conventions introduce the beginning of the predicate, i.e. the definition.

A definition shall describe a concept, not the words that make up a designation.

Before drafting a definition for a given concept, it is necessary to determine the relations between the concept and its related concepts and to model a concept system within which the concept is situated.

References to standardized definitions should be used whenever possible, and non-standardized definitions should be avoided (ISO 10241:1992 (E) 5. 2. 5).

If a definition already exists, in an International Standard for example, it shall be adopted as it stands only if it reflects the concept system in question. Otherwise, it shall be adapted. (ISO 704:2000 (E) 6. 3. 1).

Main **characteristic features** (and at the same time **requirements**) for definitions according to the ISO 704:2000 (E) 6. 3. 2; 6. 3. 3.

1. Systemic nature of definitions

A definition shall reflect the concept system. Definitions shall be coordinated so as to be able to reconstruct the concept system.

2. Conciseness

Ideally, definitions shall be as brief as possible and as complex as necessary. Complex definitions can contain several dependent clauses, but carefully written definitions contain only that information which makes concept unique. Any additional descriptive information deemed necessary should be included in a note.

3. A definition shall describe only one concept

The definition shall not include hidden definitions for any concepts used to identify characteristics. The definition should not contain characteristics that belong logically to superordinate or subordinate concepts. Any characteristic that requires an explanation shall be defined separately as a concept or given in a note.

4. Subject field appropriateness – according to the ISO 704:2000 (E) 6. 3. 4

The extension and the characteristics reflected in a definition shall be appropriate to the concept system in a given subject field.

In the specific field of the concept is not clearly indicated in the designation or is not generally understood, it shall be added to the beginning of the definition (comment: in Latvian in this case, the dash is used after the subject field indication).

When restricting the definition to a specific subject field, the extension of the concept should not be narrowed incorrectly.

5. Principle of substitution

The principle of substitution shall be used as the method for testing the validity of a definition according as the ISO 704:2000 (E) 6. 3. 5. A definition is valid if it can replace a designation in a text without loss or change in meaning.

Notes

Notes are that part of terminological entries where all secondary and extra information on a concept and its designations shall be given (ISO 704:2000 (E) 6. 5).

The function of a note is to complement a definition. Notes may include nonessential characteristics or optional parts often associated with the concept, typical elements that make up the extension of the concept, or explanatory information that complements the definition but is not essential for understanding the concept.

Graphic representations may be used as explanations, if necessary. They do not, however, replace a verbal definition.

Some typical errors and inconsistencies

When taking over existing definitions, special care shall be taken to avoid errors and inconsistencies. The following basic principles shall apply to the drafting of definitions.

- The definition shall have the same grammatical form as the term. Thus, to define a verb, a verbal phrase shall be used; to define a singular noun, the singular shall be used.
- The preferred structure of a definition is: a basic part stating the class to which the concept belongs, and another part enumerating the characteristics that distinguish the concept from other members of the class.
- The definition shall not begin with an expression such as “term used to describe” or “term denoting”; neither shall it take the form “[term] is...” or “[term] means...”
- Unless there is a specific reason, the definition shall not begin with an article.
- The definition of a quantity shall be formulated in accordance with the provisions of ISO 31-0:1992, sub-clause 2. 2. This means that a derived quantity may be defined by means of other quantities only. No unit shall be used in the definition of a quantity.

Deficient definitions – ISO 704:2000 (E) 6. 4. 1–3

Common types of deficient definitions are:

- **circular;**

If one concept is defined using a second concept, and second concept is defined using the term or elements of the term designating the first concept, the resulting definitions are said to be circular. Circular definitions do not add our understanding of the concept and shall be avoided as much as possible.

- **incomplete;**

A definition shall describe the content of the concept precisely. It shall be neither too narrow nor too broad. Otherwise, the definition is considered incomplete.

In adapting an existing definition to a specific subject field or context, care should be taken not to change the extension of the concept. A change to the extension leads to a new unit and a different concept.

- **negative;**

A definition shall describe what a concept is, not what it is not.

However, when the absence or non-existence of a characteristic is essential to the understanding of a concept, a negative definition may be required.

A particular context rarely refers to all the objects making up the extension of a concept. Definitions in laws and regulations tend to be interpretive rather than defining. Definitions in International Standards should be defining rather than interpretive. If a concept is restricted to a particular interpretation for a given text, it shall be explained in the body of the International Standard rather than by creating a new concept with a narrower extension. If specification information is associated with the concept, then this should be given in an appropriate specification clause rather than in a definition.

10 References

10.1 References from Estonia

<http://www.eter.ee> 3.1 *Estonia*

<http://www.legaltext.ee> 3.1 *Estonia*

<http://www.eki.ee> 3.1 *Estonia*

Project website <http://www.ttc.ee/~triin/phare/> (password protected) - 6.2.4 PHARE projekt „Terminological Dictionaries for Russian-Language Basic Schools"; - 7.1.2 *University of Tartu – Workflow*; 7.3.4 *University of Tartu – Source Identification*; 7.5.1.4. *University of Tartu – term extraction*; 7.6.1.7 *University of Tartu - data categories in terminological entries*; 7.6.2.4. *University of Tartu – valid entries*; 7.6.3.4 *University of Tartu – Writing Rules*; 7.8.4 *University of Tartu – exchange formats*; 7.8.4 *University of Tartu – Process requirements*; 7.9.2.5 *University of Tartu –*

Responsibilities; 7.10.1.5 University of Tartu – Process requirements; 7.10.2.5 University of Tartu – Actors; 7.10.3.4 University of Tartu – Feedback mechanisms

In-house document of the *Estonian Legal Language Centre* “Guidelines for Terminology research” – 7.1.1 *Estonian Legal Language Centre – Workflow*; 7.2.2 *Estonian Legal Language Centre*; 7.3.3 *Estonian Legal Language Centre – Source identification*; 7.4.3.3 *Estonian Legal Language Centre – Evaluation of web site information*; 7.5.1.3 *Estonian Legal Language Centre – term extraction*; 7.5.2.4 *Estonian Legal Language Centre - Concept analysis*; 7.6.1.6 *Estonian Legal Language Centre - data categories in terminological entries*; 7.6.2.3. *Estonian Legal Language Centre – valid entries*; 7.8.3. *Estonian Legal Language Centre – exchange formats*; 7.8.3. *Estonian Legal Language Centre – Process requirements*; 7.9.2.4 *Estonian Legal Language Centre – Responsibilities*; 7.10.1.4 *Estonian Legal Language Centre – Process requirements*; 7.10.2.4 *Estonian Legal Language Centre – Actors*

<http://www.legaltext.ee/en/esterm/findex.htm> - 7.6.1.6 *Estonian Legal Language Centre - data categories in terminological entries*; 7.10.3.3 *Estonian Legal Language Centre – Feedback mechanisms*

10.2 References from Latvia

1. BAUMS, A., BORZOVŠ, J., GOBZEMIS, A., FRICNOVIČS, G., ILZIŅA, I., SKUJIŅA V. Angļu-latviešu-krievu informātikas vārdnīca – datori, datu apstrāde un pārraide, 2001, 660 lpp. (English-Latvian-Russian Dictionary of Informatics – Data Processing and Transmission, Riga, 2001, 660 p.).
2. BORZOVŠ, J. *Mazas tautas lielais papilddarbs*. Referāta tēzes II Pasaules latviešu zinātnieku kongresam, Rīga, 2001, 569.lpp.(*Extra-work of a small nation*. Abstract of a paper of II Worldwide Congress of Latvian Scientists, Riga, 2001, p.569).
3. BORZOVŠ, J., ILZIŅA, I., SKUJIŅA, V., VANCĀNE, I. *Sistēmiska latviešu datorterminoloģijas izstrāde*. LZA Vēstis, Rīga, 2001, 55.sēj. 1./2. num., lpp.83-91. (*Systemic Development of Latvian Computer Terminology*, Proceedings of Latvian Academy of Sciences, Riga, 2001, vol. 55, N 1/2 pp. 79-83).
4. DREZEN, E. *Internationalization of Scientific-technical Terminology*. (Translated from Russian.) Rīga, LU LVI, 2002. – 72 p.
5. GALINSKI, CHR. The economics of terminology in the age of the multilingual information society. // *Terminology and Technology Transfer in the Multilingual Information Society*. Proceedings of the 2nd International Conference on Terminology in commemoration of E. Drezen’s 110th Anniversary. Vienna/Riga, IITF-Infoterm/LLI of LU, 2003. – P.33-42.
6. ILZIŅA, I. *Latviešu termini Eiropas Savienības datortīklos*. Referāta tēzes starptautiskai konferencei *Latviešu grāmata un bibliotēka: 1525.g – 2000.g*, Rīga, 2000, lpp. 202 - 205. (*Latvian terms in European Union networks*. Abstract of a paper of Intern. Conf. *Latvian Book and Library from 1525 - 2000*, Riga, 2000, pp.202 - 205).
7. *Konferences „Zinātnes valoda”materiāli*. (*Proceedings of the conference „Language of Science”*.) Rīga, National Language Commission, 2003 – 51 lpp.
8. KRŪMIŅA, V., SKUJIŅA, V. *Normatīvo aktu izstrādes rokasgrāmata*. (*Manual for creating normative acts*.) Rīga, Valsts kanceleja, 2002. – 116 lpp.

9. *Latviešu valoda – robežu paplašināšana. (The Latvian Language – Extending the borders.)* Rīga, National Language Commission, 2005. – 216 lpp.
10. PICHT, H., DRASKAU, J. *Terminology: An introduction.* Copenhagen, School of Economics; University of Surrey England, 1985. – 265 p.
11. *Situācijas izpēte latviešu terminoloģijas izstrādes, saskaņošanas un apstiprināšanas jomā: problēmu identifikācija un to risinājumi. (Investigation of the situation in the terminology elaboration (output?), harmonization and approval of the Latvian Terminology: Problem identification and solution.)* TTC. Rīga, 2004.
12. SKUJIŅA, V. Gramatikas kategoriju loma terminu sistēmas veidošanā. (The role of grammar categories for the developing of term system.) // *Baltu Filoloģija, XI (2).* Rīga, LU, 2002. – P. 67-78.
13. SKUJIŅA, V. *Latīņu un grieķu cilmes vārddāļu vārdnīca. (Dictionary of the Latin and Greek elements.)* Rīga, Kamene, 1999. – 233 lpp.
14. SKUJIŅA, V. *Latviešu terminoloģijas izstrādes principi. (The Principles of Formation of Latvian Terminology.)* 2nd ed. Rīga, LZA/LU LVI, 2002. – 224 p.
15. SKUJIŅA, V. Latviešu valodas terminoloģijas attīstība nacionālo un internacionālo tendenču mijiedarbībā. (Development of Latvian terminology in the framework of interactive national and international tendencies.) // *Baltu Filoloģija, X.* Rīga, LU, 2001. – P. 113-125.
16. SKUJIŅA, V. The Specificity of the Term and the Concept of the Termeme // *IITF Infoterm Multilingualism in Specialist Communication.* Vienna, TermNet, 1996. – Vol.2. – P. 1123-1130
17. *Terminology and Technology Transfer in the Multilingual Information Society.* Proceedings of the 2nd International Conference on Terminology in commemoration of E. Drezen's 110th Anniversary. Vienna/Riga, IITF-Infoterm/LLI of LU, 2003. – 184 p.
18. *Translation Handbook for Latvian Legislation. (lv→en)* Rīga, TTC, 2000. – 80 p.
19. VANCĀNE, I., BORZOVŠ, J. *Development of the Latvian Terminology of Information Technologies.* Abstract of a paper of Intern. Conf. & Exhib. *Information Technologies and Telecommunications in the Baltic States*, Riga, 1999, pp. 98-102.
20. WEISSENHOFER, P. *Conceptology in terminology theory, semantics and word-formation.* Wien, TermNet, 1995. – 271 p.

10.3 References from Poland

Web Sites – references to items 2.5. and 3.5. of D.1.1.

[2.5.1. and 3.5.1.] Polish Standardization Committee 'PKN' (Polski Komitet Normalizacyjny)
<http://www.pkn.pl>

[2.5.2. and 3.5.2.] The Office of the European Integration Committee 'UKIE' (Urząd Komitetu Integracji Europejskiej)
<http://www.ukie.gov.pl>

[2.5.3. and 3.5.3.] Council for the Polish Language (Rada Języka Polskiego)
<http://www.rjp.pl>

[2.5.5. and 3.5.5.] Polish Society of Economic, Legal and Court Translators TEPIS (Polskie Towarzystwo Tłumaczy Ekonomicznych, Prawniczych i Sądowych TEPIS)

<http://www.tepis.org.pl>

[2.5.7. and 3.5.7.] On-line Dictionaries

Polska Terminologia Informatyczna – Biuro Tłumaczeń Informatycznych

www.btinfo.pl

Zintegrowany katalog anglistyczny POLANKA

www.bibang.uw.edu.pl/makar.htm

Słownik terminologii hutniczej

www.stalportal.pl/webapp/slownik/slo_index.jsp

Słownik Hydrogeologiczny

www.mos.gov.pl/dg/slownik/wstep.htm

Polska Klasyfikacja Tematyczna

www.kbn.gov.pl/pub/kbn/docs/pkt.html

www.nauka.pwr.wroc.pl/dn/oferta/zal2.pdf

Bank Terminów Znormalizowanych

www.pkn.pl/informacyjne4.htm

Międzynarodowy Tezaurus Techniczny – TIT

www.pkn.pl/informacyjne4.htm

Słownik terminologii z zakresu zapewnienia bezpieczeństwa zdrowotnego żywności

www.cbr.edu.pl/slownik.php

Pierwszy w Polsce Słownik Terminologii Dziennikarskiej

www.media.netpr.pl/notatka_4100.html

Słownik terminów ubezpieczeniowych

<http://www.ebroker.pl/pls/s!/slo.slownik>

Leksykon Górniczy

www.teberia.pl/encyklopedia.php

Słownik haseł ekonomicznych

www.nbportal.pl/common/Wiadomosc.jsp?nid=12575&page=1&active=3028

Leksykon Teleinformatyka

www.networld.pl/leksykon

Leksykon nowych nazw i pojęć w biochemii i biologii molekularnej

www.ptbioch.edu.pl/html/leksykon.htm

Słownik terminów biologicznych

www.aneksy.pwn.pl/biologia/

Słownik terminów finansowych

www.expander.pl/expander/index.jsp?place=Menu3&news_cat_id=45&layout

Słownik Medyczny

www.lexland.com.pl/produkt/medap.php

Słownik terminologii prawniczej

<http://www.mareno.pl/ksiazka.php?isbn=83-89073-46-3>

Wielka Encyklopedia Internetu i Telekomunikacji on-line *Wydawnictwo Robomatic s.c.*

<http://www.robomatic.pl/?id=enc&l=A>

Tezaurus Transportu – TEZ

http://www.gbk.mi.gov.pl/_sgt/m7_1.htm

Polski Tezaurus Medyczny

http://www.gbl.waw.pl/html/bazy_danych.htm

Słownik pojęć statystycznych

<http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm>

Tezaurus Prawa

<http://bib.sejm.gov.pl/ALEPH/300313/file/base-info-bis05>

Leksykon Biznesu

<http://www.placet.com.pl/leksykon.asp>

Słownik astronomiczny polsko-angielski i angielsko-polski

http://www.astronomia.pl/slownik/slownik_full.php

AstroWORDS – Internetowy Słownik Astronomiczny

<http://words.astronet.pl/>

Otwarty Słownik Terminologii Naukowej – OSTEN

<http://biot.ar.szczecin.pl/slownik.php>

Słownik terminologii logistycznej

<http://www.logistyka.net.pl/cgi-bin/index.cgi?portal=logistyka§ion=terminologia>

Internetowy Słownik Ekonomiczny

<http://isb.republika.pl/>

Słownik Terminów Technicznych

www.slownik.kargul.net/

Internetowy Słownik Fotograficzny

<http://www.fotograf.fir.pl/dodatki/sownik.php>

Słownik Astronomiczny Polskiego Towarzystwa Astronomicznego

<http://leksykon.pta.edu.pl/>

Słownik zwrotów branżowych

<http://www.gepol.com.pl/systemy/main.php>

Słownik sztuki starożytnej Grecji i Rzymu

<http://www.wiw.pl/kulturaantyczna/twardecki/>

Pierwszy polski "Tezaurus Nauk Teologicznych"

www.fides.org.pl

Baza terminów tezaury

http://www.biblos.pk.edu.pl/baza_terminow_tezaury.php3

Angielsko-polski słownik nowych polskich terminów fizycznych

<http://ptf.fuw.edu.pl/nazew.html>

Słownik historyczny

<http://www.polskiedzieje.pl/sownik.php?idslowa=251&calosc=1&PHPSESSID=57153cb9aa78f597e91c31286d7eeb9b>

Angielsko-polski i polsko-angielski słownik terminów i pojęć używanych w programach współpracy zagranicznej w dziedzinie edukacji

www.socrates.org.pl/erasmus/pub_sownik.html

Słownik terminów literackich

<http://wysylkowa.pl/ks259599.html>

Angielsko-polski słownik terminów z zakresu dokumentacji i informacji patentowej

<http://www.uprp.pl/NR/rdonlyres/18BAE901-5C48-4090-9A93-650A65D73662/135/sownikangielskopolski.pdf>

Słownik muzyczny

<http://www.ulalala.obywatel.pl/sownik.php?id=2>

PolTerm Polish-English Basic Legal Dictionaries

www.tepis.org.pl/terminologia (access restricted by memberships)

[2.5.7. and 3.5.7.] Publications

Bartnik, E. (1994), *English-Polish Dictionary – Genetic Engineering*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.

- Bień, J. S. (1994/95), *English-Polish Dictionary – Artificial Intelligence Vocabulary*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Dalek, J., S. Dalek (1994), *English-Polish Dictionary – Informatics (incl. Polish-English index)*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Dawidowicz, A. (1994/95), *English-Polish Dictionary – Food Additives*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Degler, A., J. Stoberski (1994), *English-Polish Dictionary – Pensions*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Degórska, M., M. Francke, D. Kubik, R. Tomaszenko (1993), *English-Polish Dictionary – Protection of Ozone Layer*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Domanowska, Z. (1994), *English-Polish Dictionary – Health Services*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Kierzkowska, D. (ed.) (1994-2005), *Lingua Legis. Journal of the Polish Society of Economic, Legal and Court Translators TEPIS*, No. 1-13. Warsaw: TEPIS Publishing House.
- Kujawa, J. (red.) (2001), *MARDIC – Słownik anglojęzycznych terminów i skrótów handlu morskiego*. Gdańsk: Wydawnictwo UG.
- Kuźmicki, T. (1994), *English-Polish Dictionary – Free Trade*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Nowicki, L. (1992), *Słownik angielsko-polski – Energia Nuklearna*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Nowosad, T. (1995), *English-Polish Businessman's Manual Dictionary*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Pawlicki T., H. Bubis (2001), *A Dictionary of Agricultural Engineering in Six Languages*. Poznań: Przemysłowy Instytut Maszyn Rolniczych.
- Różycka J. (1994), *English-Polish Dictionary – Statistics and Surveys*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Sokolski, J. (2000), *Słownik barokowej symboliki natury. Tom wstępny. Barokowa księga natury*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Stoberski, Z. (ed.) (1984-1997), *Neoterm – World Specialized Terminology. Journal of the International Organization for Unification of Terminological Neologisms – IOUTN*, No. 1-32. Warsaw: International Organization for Unification of Terminological Neologisms – IOUTN.
- Stoberski, Z. (ed.) (1993), *English-Polish Dictionary – Banking*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Stoberski, Z. (ed.) (1993), *English-Polish Dictionary – Solid Waste*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Stoberski, Z. (ed.) (1994), *English-Polish Dictionary – Employment*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Stoberski, Z. (ed.) (1994), *English-Polish Dictionary – Finances*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – PCIOUTN.
- Stoberski, Z. (ed.) (1994), *English-Polish Dictionary – Stock Market*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.

- Stoberski, Z. (ed.) (1994), *Słownik angielsko-polski – Ochrona środowiska w rolnictwie*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Stoberski, Z. (ed.) (1994/5), *English-Polish Dictionary – Environment*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Stoberski, Z. (ed.) (1995), *English-Polish Dictionary – Cell Engineering*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Wojtasiak, I. (red.) [et al.] (1999), *Słownictwo znormalizowane. Technika Informatyczna*. Warszawa: Biuro Polskiego Komitetu Normalizacyjnego.
- Wojtasiak, I. (red.) [et al.] (2001), *Słownictwo znormalizowane. Ochrona środowiska*. Warszawa: Biuro Polskiego Komitetu Normalizacyjnego.
- Wojtasiak, I. (red.) [et al.] (2001), *Słownictwo znormalizowane. Bezpieczeństwo pracy i technika przeciwpożarowa*. Warszawa: Biuro Polskiego Komitetu Normalizacyjnego.
- Woydt, M. (1995), *English-Polish Dictionary – Parliamentary Procedure*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.
- Wyderka, B. (red.) (2000-2002), *Słownik gwar śląskich, T. I-III*. Opole: Państwowy Instytut Naukowy, Instytut Śląski.
- Zamojski, J. (1994), *English-Polish Dictionary – Electrical Engineering*. Warsaw: Polish Committee of the International Organization for Unification of Terminological Neologisms – IOUTN.

[2.5.7. and 3.5.7.] **NON-PUBLISHED DICTIONARIES CREATED WITHIN THE FRAMEWORK OF UNIVERSITY RESEARCH ACTIVITIES ‘SYNABA’**

Toponimiczny słownik Krymu (SYNABA Database held by the Information Processing Centre ‘OPI’)

Podręczny słownik terminów literackich (SYNABA Database held by the Information Processing Centre ‘OPI’)

Słownik terminologii handlu zagranicznego (SYNABA Database held by the Information Processing Centre ‘OPI’)

Ludy i narody obu Ameryk. Słownik etnologiczny (SYNABA Database held by the Information Processing Centre ‘OPI’)

10.4 References from Hungary

Kis, B., Villada, B., Bouma, G., Bíró, T., Nerbonne, J., Ugray, G. and Pohl, G. 2004a. A new approach to the corpus-based statistical investigation of Hungarian multi-word lexemes, In: *Proceedings of LREC 2004*. Lisbon.

Kis, B., Villada Moirón, B., Bíró, T., Bouma, G., Pohl, G., Ugray, G., Nerbonne, J. 2004b. Methods for the Extraction of Hungarian Multi-Word Lexemes. In: Decadt, B. (ed.): *Proceedings of CLIN-2003*. Antwerp: University of Antwerp.

Castellví, M. T. C., Bagot, R. E., Palatresi, J. 2001. Automatic term detection: A review of current systems. In: Bourigault, D., Jacquemin, C. and L'Homme, M.-C. (eds.): *Recent Advances in Computational Terminology*. Amsterdam–Philadelphia: John Benjamins. 53–88.

10.5 References from Germany

4.2.6 Germany – DIN

- <http://www.din.de>
- <http://www.beuth.de>

6.1.3 Infoterm

- <http://www.infoterm.info>
- Flyer “30 Years of Infoterm” by Infoterm

6.2.2 Standards-based Access to multilingual Lexicons and Terminologies (SALT)

- <http://www.iim.fh-koeln.de/iim/salt.html>
- <http://www.loria.fr/projets/SALT/saltsite.html>
- <http://www.ttt.org/salt/description.html>

6.3.1 Terminology management tools

Schmitz, Klaus-Dirk: Systeme zur Terminologieverwaltung. In: Technische Kommunikation, No. 2/2001, p. 34-39.

Schmitz, Klaus-Dirk: Die neuen Terminologiedatenbanken: online statt offline. In: Felix Mayer, Klaus-Dirk Schmitz, Jutta Zeumer (eds.): Terminologie und Wissensmanagement. Akten des Symposiums, Köln, 12.-13. April 2002. Köln: Deutscher Terminologie-Tag e.V.

Deutsches Terminologie-Portal: <http://www.iim.fh-koeln.de/dtp>

7.8.1 TBX (TermBase eXchange)

- OSCAR Group: TBX specification draft 1j (2002-05-02)

9.4 ISO 12200:1999 – Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) — Negotiated interchange

- ISO 12200:1999(E). Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) — Negotiated interchange
- Melby, Alan K.; Schmitz, Klaus-Dirk; Wright, Sue Ellen: The Machine Readable Terminology Interchange Format (MARTIF) — Putting Complexity in Perspective. In: TermNet News, Nr. 54/55-1996, S. 11-21.

- Schmitz; Klaus-Dirk: MARTIF – A New ISO Standard for the Interchange of Terminological Data. In: TermNet News, Nr. 50/51-/1995, S. 4-5.
- Schmitz, Klaus-Dirk: MARTIF - Ein SGML-basiertes Austauschformat für terminologische Daten. In: Wiebke Möhr, Ingrid Schmidt (eds.) (1999): SGML und XML - Anwendungen und Perspektiven. Berlin: Springer, S. 109-121.

9.5 ISO 12620:1999 – Computer applications in terminology – Data categories

- ISO 12620:1999(E). Computer applications in terminology – Data categories

9.6 ISO 16642:2003 – Computer applications in terminology – Terminological markup framework (TMF)

- ISO 16642:2003(E). Computer applications in terminology – Terminological markup framework (TMF)
- Laurent Romary: An abstract model for the representation of multilingual Terminological data: TMF – Terminological Markup Framework. Published on: http://www.papillon-dictionary.org/info_media/1620632.pdf

10.6 References from Lithuania

3.4.1

A. Auksoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijas Jaunumi: Starptautiskais terminoloģijas seminārs „Terminoloģijas darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (16-17).

A. Auksoriūtė. *Terminology Work at the Institute of the Lithuanian Language // Programme of the International seminar Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.

www.lki.lt

3.4.2

A. Auksoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijas Jaunumi: Starptautiskais terminoloģijas seminārs „Terminoloģijas darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (18-19).

A. Ivanauskienė. *Lithuanian Bank of terms – a New Tool for the Regulation of Terminology // Programme of the International seminar Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.

www.vlkk.lt

3.4.3

N. Dudlauskienė. *Standardization of Lithuanian Technical Terms // Programme of the International seminar Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.

www.lsd.lt

4.2.4

- N. Dlauskienė. *Standardization of Lithuanian Technical Terms* // Programme of the International seminar *Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.
- R. Maslauskienė. Terminų standartizavimas / Standardization of Terms // Terminologijos istorijos ir dabarties problemos. Vilnius, 2004 (256-270).

www.lsd.lt

6.1.7

- A. Auksoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijos Jaunumi: Starptautiskasis terminologijos seminars „Terminologijos darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (21-23).
- S. Meškauskienė. *Presentation of Terminynas (Term base) of the Institute of Mathematics and Informatics* // Programme of the International seminar *Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.

www.terminynas.lt

7.2.4

- A. Auksoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijos Jaunumi: Starptautiskasis terminologijos seminars „Terminologijos darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (21-23).
- S. Meškauskienė. *Presentation of Terminynas (Term base) of the Institute of Mathematics and Informatics* // Programme of the International seminar *Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.

www.terminynas.lt

7.3.6

- A. Auksoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijos Jaunumi: Starptautiskasis terminologijos seminars „Terminologijos darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (21-23).
- S. Meškauskienė. *Presentation of Terminynas (Term base) of the Institute of Mathematics and Informatics* // Programme of the International seminar *Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.

www.terminynas.lt

7.7.6

- A. Auksoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijos Jaunumi: Starptautiskasis terminologijos seminars „Terminologijos darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (21-23).

S. Meškauskienė. *Presentation of Terminynas (Term base) of the Institute of Mathematics and Informatics // Programme of the International seminar Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.
www.terminynas.lt

7.9.3.4

A. Aukšoriūtė, J. Gaivenytė, A. Umbrasas *The State of Lithuanian Terminology // Terminologijas Jaunumi: Starptautiskais terminoloģijas seminārs „Terminoloģijas darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā”*. Riga, 2003 (21-23).

S. Meškauskienė. *Presentation of Terminynas (Term base) of the Institute of Mathematics and Informatics // Programme of the International seminar Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union*. Vilnius, 2004, October 5-6.
www.terminynas.lt

10.7 References from Denmark

Chapter 4.1.

<http://www.iso.org/iso/en/ISOOnline.frontpage>

Chapter 5.1

European Commission, Translation Service – Terminology and Language Support Services, 1996.
Lenoch Universal Classification System, LUC – vol. 2, Main part. 2. edition. Upd. H. Wellenstein.

Lenoch: <http://europa.eu.int/eurodicautom/Controller>

Lenoch: <http://www.admin.ch/ch/i/bk/termdat/d/termdat/class.htm>

Eurovoc:

http://europa.eu.int/celex/eurovoc/cgi/sga_doc?eurovoc_dif!SERVEUR/menu!prod!MENU&langue=EN

Chapter 6.1

The IATE Project, an overview, internal report.

Chapter 6.2.

Intera: <http://www.elda.org/rubrique22.html>

Gavrilidou, M., V. Giouli, E. Desipri, P. Labropoulo (ILSP), *Intera – Report on the Multilingual Resources Production*, Deliverable 5.2, e-content EDC-22076 INTERA / 27924

Enabler: <http://www.enabler-network.org/>

Enabler: <http://www.elda.org/article103.html>

Calzolari, N., K. Choukri, M. Gavrilidou, B. Maegaard, P. Baroni, H. Fersøe, A. Lenci, V. Mapelli, M. Monachini, S. Piperidis (2004): ENABLER Thematic Network of National Projects: Technical, Strategic and Political Issues of LRs. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, p.937-940, Lisboa.

Chapter 6.3

Arppe, A., (1995) : *Term Extraction from Unrestricted Text*, NoDaLiDa-95

Streiter, O., D. Zielinski, I. Ties, L. Voltmer (2003): *Term Extraction for Ladin: An Example-based Approach*, TALN 2003.

Thurmair, G. (2003): *Making Term Extraction Usable*, EAMT workshop 2003, Dublin (Powerpoint presentation)

Chapter 7

Ball, S., D. Rummel (2001) The IATE Project - towards a Single Terminology Database: In *Proceedings of the 23rd International Conference on Translating and the computer*, London, November 2001

Interinstitutional Committee for Translation and Interpretation, IATE Data Management Group (2003): *Writing rules for the IATE database*, internal report.

IATE, best practice, internal report, prefinal version.

Quality and Reliability SA (2000), *System analysis and design*, report of the IATE project

Sager, J.C. (1990), *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Suonuuti, H. (1997), *Guide to Terminology*. Helsinki:Tekniikan Sanastokeskus

The PAVEL Terminology Tutorial

http://www.termium.gc.ca/didacticiel_tutorial/english/

Chapter 9.1

ISO 704:2000 Terminology work – Principles and methods

ISO 860:1996(E) Terminology work – Harmonization of concepts and terms

ISO 10241:1992(E) International terminology standards – Preparation and layout