Terminology and Technical Documentation

Translation technology and workflowmanagement

Prof.dr.Frieda Steurs
Lessius/KULeuven

Structure of the presentation

- Some economic aspects of translation
- Source text control
- Issues in translation and multilingual document management :
  - Case 1 : Integrated Language Services (ILS)
  - Case 2 : Cummins Engine Company
  - Case 3 : Tweddle Litho Company
  - Case 4 : Yamagata Europe
  - Case 5 : Atlas Copco
- Conclusion

Economic aspects of translation

- The translation and interpreting market
- The market of multlingual document management
- The market of “language services”
  - 20-30% growth each year
  - 30 billion Euro turn around world wide
  - EU : 1,1 billion Euro/year on translation costs
  - Loss of markets because of monolingualism

Key factors

- Liberalisation and deregulation of key industries (telecom, power generation)
- Free trade, open borders between countries
- Emergence of regional economic and political communities
- Creation of worldwide technical and logistics infrastructure

Internet

- Low-cost worldwide advertising and marketing platform
- Low-cost, worldwide sales platform
- Low-cost, worldwide distribution platform
- Low-cost, worldwide support platform
  - Localisation as part of the documentation and marketing process : allows speakers of less common languages enjoy acces to the same products and resources that those in major markets use
- Manufacturers of automobiles, heavy equipment, consumer goods, retailers, media, entertainment, legal firms, pharmaceutical companies, financial institutions, food services companies, software companies,….

Economic interests

- Localisation industry
  - Adapting software, websites, etc. to local cultural and technical requirements of the target market
  - Interdisciplinarity :linguists, cultural specialists, project managers, webdesigners, software developers, marketingspecialists, etc.
  - Downturn in the global economy makes it much more important to communicate to customers effectively and in their own language

Some aspects of source text creation

- Management and usability of technical documentation
- The reader/user :
  - Unexperienced
  - Expert
- ‘Restructuring minimalism’
- The minimal manual

Source text management

- Case 1 : documentation for air line reservation system
  - Research into user behaviour :
    - 80% reduction in volume of the manual
    - 80% reduction of phonecalls to helpdesk
- Case 2 : documentation for parcel delivery
  - 74 billion $ losses due to the use of an inadequate Operational Manual

Controlled language

- Constrained terminology, syntax and/or semantics
- Clear and consistent style
- Enhance clarity, usability, transferability, retrievability, extractability, and translatability
- Some resistance when first using it
- The objectives are met

How to Provide

- the right information

  relevant
  accurate
  non-ambiguous

- at the right time

  retrievable
  up-to-date
  maintainable

- in the right form

  clear
  structured
  consistent
  translatable

Content Optimization Measures

- Controlled Language and CL Checker
- Spelling and Grammar Checker
- Text-graphics ratio
- Single-sourcing
- Meta-information
- Translation Memory
- Machine Translation
- Process Control
- ...

What is Controlled Language?

Controlled Language is a *subset* of conventional language.

The subset is characterized by

- a standardized terminology
- a restricted core vocabulary
- a restricted set of grammar and style rules

Motivation

- Readability and international ***understandability***
- ***Translatability***
- National and international ***standardization***
- ***Retrievability*** in the context of knowledge and content management systems

- The new language in International Business. Simplified English (Tedopres International) www.tedopres.com

Classification Scheme

Linguistic Variety	Application	Setting	Standardization Level	Purpose
Simplified English	Technical Documents	Aeronautic Industry	(Terminology) Core Vocabulary Syntax	Readability Consistency
Controlled English	Technical Documents	Other Industries ▪ Complex Manufacturing ▪ ICT	Terminology (Core Vocabulary) Syntax	Readability Consistency Translatability
Plain English	Administrative Documents	Government	Core Vocabulary Syntax	Readability

Standardized Terminology

Frieda Steurs - 2009 - TSS

What is Standardized Terminology?

- Principle One: 1 Concept = 1 Term

⇒ disallow synonyms, orthographic and morphological variants

Principle One : 1 Concept = 1 Term

Synonyms

GSM

handy

mobile

wireless phone

Morphological variants

indexes

indices

GSM

indexes

What is Standardized Terminology?

- Principle One: 1 Concept = 1 Term

⇒ disallow synonyms, orthographic and morphological variants

- Principle Two: 1 Term = 1 Concept

⇒ disallow homonyms

Principle Two : 1 Term = 1 Concept

Homonyms

airbag	airbag airbag system
noise	unwanted sound electronic interference
application	piece of software result, action of applying

What is Standardized Terminology?

- Principle One: 1 Concept = 1 Term

⇒ disallow synonyms, orthographic and morphological variants

- Principle Two: 1 Term = 1 Concept

⇒ disallow homonyms

- Principle Three: 1 Term = 1 Part of Speech

⇒ disallow homographs

Principle Three :
1 Term = 1 Part of Speech

Homographs

test	noun verb
abrasive	noun adjective
absent	adjective verb

What is Standardized Terminology?

- Principle One: 1 Concept = 1 Term

⇒ disallow synonyms, orthographic and morphological variants

- Principle Two: 1 Term = 1 Concept

⇒ disallow homonyms

- Principle Three: 1 Term = 1 Part of Speech

⇒ disallow homographs

- Principle Four: Acceptance of Term

⇒ prefer terms that are linguistically adequate, transparent in meaning and common in use

Why Standardize Terminology?

TERMS = BASIC KNOWLEDGE ELEMENTS

Terms help you to organize, distribute and standardize knowledge

- Comprehensibility: common understanding of documents
- Translatability: only one correct translation
- Maintainability: reduced set of terms and translations
- Consistency: always the same terms for the same concepts
- Retrievability: unambiguous set of basic knowledge elements

Restricted Core Vocabulary

Frieda Steurs - 2009 - TSS

What is a restricted core vocabulary?

DO NOT USE	USE
accomplish	do
utilie	use
endeavor	try
accelerate	speed up
encounter	find
detail	instruction
normally	usually, correctly

Restricted Syntax

Prohibited Features

- Long and complex sentences
- Many noun phrases per sentence
- Many nouns or adjectives in a row
- Many different sentence structures
- Complex verb forms and tenses
- Compressive structures such as attribute clauses, participle clauses, ellipses, contractions
- Inconsistent or wrong punctuation

Readability

Non-controlled:

Tighten the screw several turns, remove the screwdriver and check to see that the cover still is in place on the phone before connecting the phone wire.

Controlled:

1. Tighten the screw several turns.
2. Remove the screw driver
3. Check to see that the cover still is in place on the phone
4. Connect the phone wire

Comprehensibility

Non-controlled:

When fasteners are removed, always reinstall them at the same location from which they were removed.

Controlled:

Always reinstall fasteners in the same location.

Translatability

Non-controlled:

A 10 mega-ohm input impedance digital voltmeter

Controlled:

A digital voltmeter with 10 mega-ohm input impedance

Non-ambiguity

Non-controlled:

Unscrew the plug-in unit from the wall and paint it.

Controlled:

Unscrew the plug-in unit from the wall and paint the plug-in unit.

Case 1 : ILS
Integrated Language Services

- Printing business
- Language related issues :
  - Copywriting
  - Translation
  - Terminology management
  - Translation memory
  - Remodelling and finetuning of workflows

ILS : Main Tasks

- Coordination and streamlining of the translation process
- Creation, hosting and updating of translation memories and termbases (in-house terminology)
- Editing, copywriting and localisation issues
- Content creation for multilingual websites
- Source text control and management

ILS : Workflow

- Pre-study of existing documentation
  - Style briefing of translators
  - Reference material evaluation by the customer
  - Terminology study of source texts and mapping of inconsistencies
- Implementation of MAHT
- Internal and external evaluation
- Full project execution
- Follow-up and evaluation

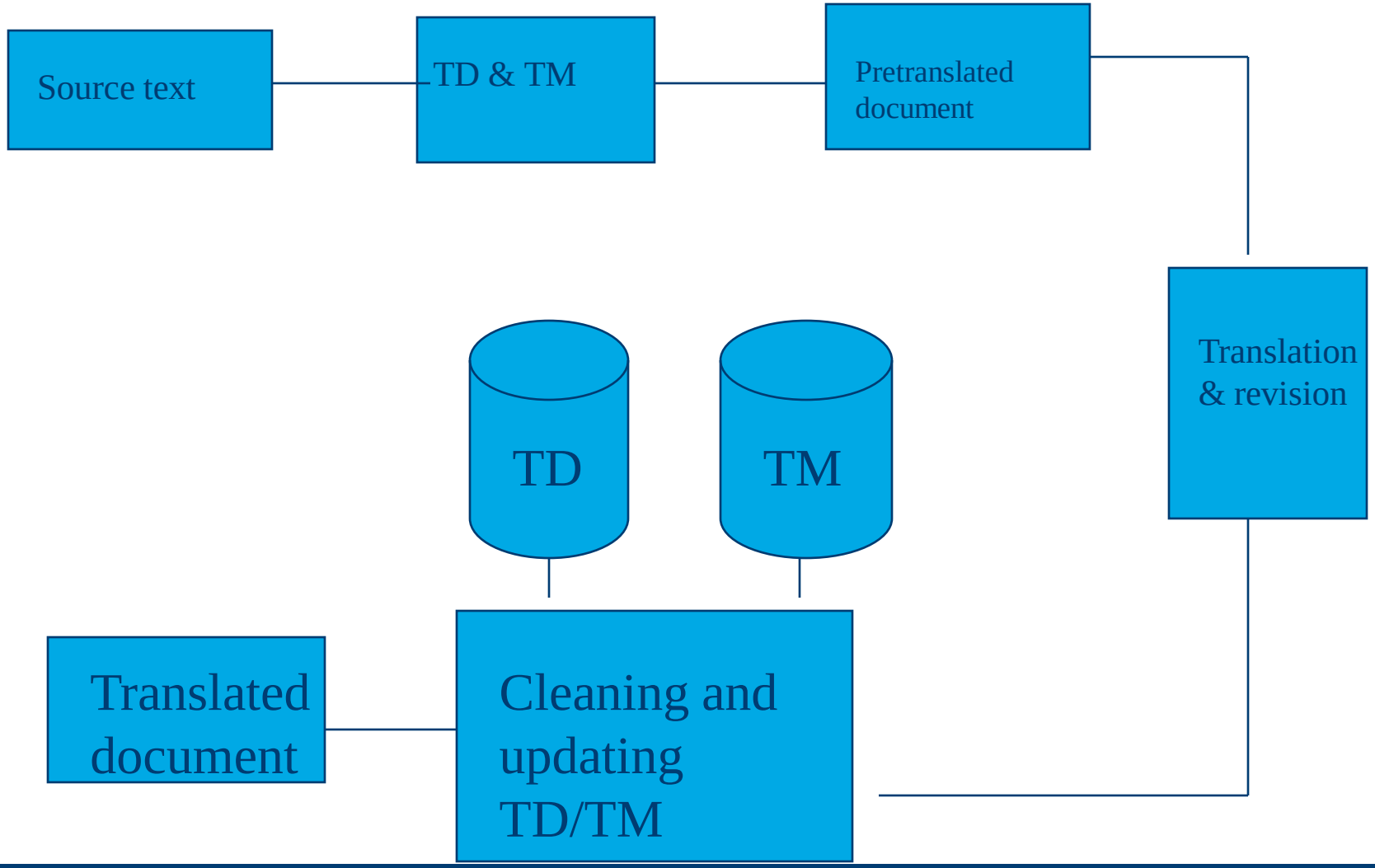
Intercultural Problems

- Asian customers and the European market
- Files are in unknown or incompatible file format
- Source text has been translated into 'Japenglese'
- Content is not adapted to European market
- Rewritten sourcetexts need local market guidelines for translators
- Text expansion

Knowledge management

- Data/ Information / Knowledge
- Information :
  - Textual
  - Graphical
  - Corporate Identity
- Input : source control
- Output : a variety of formats
  - CD-ROM, Website, media-neutral databases, printed matter , etc.

Workflow



Case 2 : Cummins Engine Company

- Leading manufacturer of gas and diesel engines (4500 distributors in 130 countries)
- 50 service manuals and 60 parts manuals every year

- http://www.cummins.com

Towards more efficient localisation procedures

- The publications had to be localised in a more efficient way
- Reduce publication time and translation cost
- Tools for :
  - Selection and output of source/target data for translation
  - Use of translation memory
  - Post-editing by translators
  - Controlled language (authoring process)
    - More accurate matching in the translation memory

Source text management

- Source language English
- Information chunks in manageable fragments
- Diffing algorithm : comparison of the last and current versions of the document to determine the editing changes
- Comparison in TM : unmatched items are sent for translation.
- Upgrade of TM with the new information in the translation and in the source document

Results

- 6 languages : Sp, Fr, G, I, Sw, Port
- Operator and Maintenance Manual (400 pages) : from 6-8 months to 10 weeks.
- 65-70% reduction in translation costs :
  - Use and reuse of data across publications
  - Integration with terminology and TM
  - Batch composition system for layout and graphics (saves 30% on production costs)

Case 3 : Tweddle Litho Company

- Technical publishing house for automotive industry
- http://www.tweddle.com
- Complete data management services including authoring, translation and other related services

Better, faster, cheaper

- Ford Motor Company : 30 languages, 60 countries (localisation!) supporting the vehicle release.
- Lay-out , graphics, had to be more accurate
- Culturally neutral global format and meet local requirements for engineering, regulatory, safety and environmental conditions.
- Reduce publication cycle time and costs

Problems with the information flow

- The use of flat files without data reuse across publications
- Inconsistent information from one publication to another
- Publications had to become portable to a variety of software applications and output to multiple media formats
- Translators had to start each manual from scratch, and this made the production cycle very long

Key to the solution

- Data model and DTD : SGML
- Information granularity / storage
- Media specific output : paper / electronic
- Effectivity control : target audience, type of vehicle involved
- Documents and CDs in over 30 languages. Increase in shared data up to 90%. Turnaround time for all markets from 6 months to 2 weeks.

Content management

Content management systems are systems that capture, archive, index, manage, author, combine, link, and distribute internal and external information to create a knowledge repository

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

The need for a centralised process

- Write it once
- Reuse it in many places
- Translate it once
- Reuse it in many places

Yamagata Europe

- Automated Quality Control in Technical Translation
- Sony, Ricoh, Hitachi, ..
- http://www.yamagata-europe.com

Yamagata Europe

- Translation work : 100% outsourced
- 30.000.000 words translated in 2005
- 30 different language pairs
- 13 different formats
- 700 translators
- Several specialisations
- → Vendor management: selection and evaluation : important for a language service providor

Problems (1)

- Control
  - Reliability
  - Size of the translatorsgroup
  - Transparancy
  - Lack of good and specialised translators
  - Capacity
  - Confidentiality

Problems(2)

- Lack of time
  - Size of the translatorsgroup
  - No inhouse-expertise for every language pair
  - Manual revision is both financially and practically not feasible

Problems (3)

- Subjectivity of the existing evaluationcriteria(e.g. SAE J2450, LISA QA model,…)
  - Based on manual revision
  - Arbitrary and subjective catagories
  - No consensus

  - http://www.sae.org/technicalcommittees/j2450p1.htm

Translation Quality Control

- →Need for quality control

- → Need for automated processes:
  - Formal mistakes in translation work are most of the time indicative for other problems in the text
    - "Everything that can be measured, can be traced"
    - Distinction between formal mistakes and stilistic (subjective) mistakes

QA Distiller

- Error categories:
  - Omissions
  - Inconsistencies
  - Formatting issues
  - Terminology mistakes

Vendor testing

A. Processrelated translation quality
   - Communication
   - Punctuality
   - Segmentation
   - Tagging

Vendor testing

- B. Productrelated translation quality
  - Omissions
  - Inconsistenties
  - Formatting
  - Terminology
  - Spelling

Case 5 : Atlas Copco

- Atlas Copco :
  - Industrial and professional tools compressed air equipment, construction and mining equipment etc.
  - (26.000 employees worldwide)
- Bowne Global Solutions :
  - Management of the technical documentation and localisation of the compressor product information
  - (now part of Lionbridge) http://www.lionbridge.com

Atlas Copco : the challenge

- Specialised manuals in more than 12 languages

  Conditional publishing
- Source once, Write many times

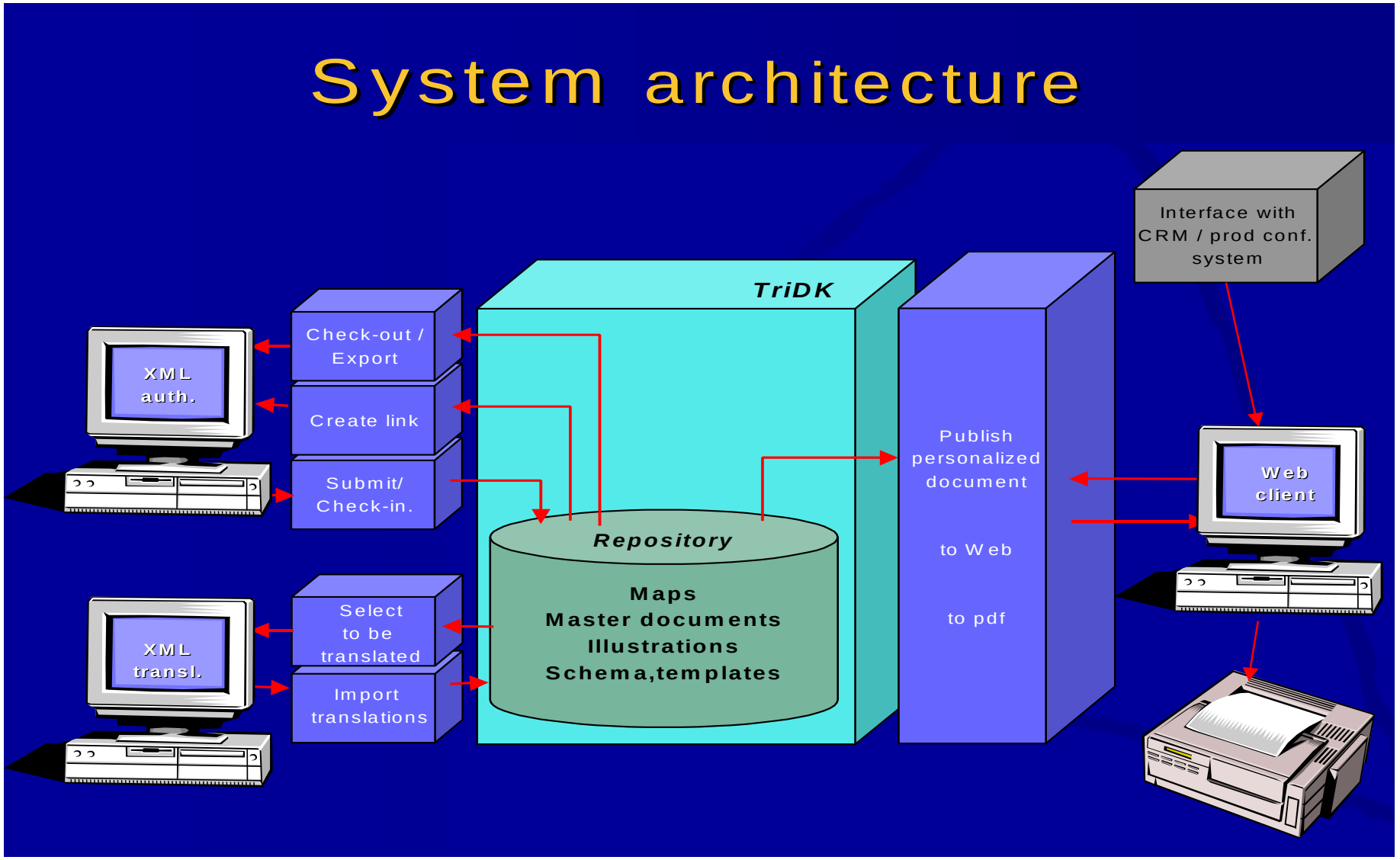
Atlas Copco manuals

- Publish instruction manuals on the Web in web standard XML format - requiring no specific software from user
- Personalisation of the instruction manual in 3 ways:
  - **conditional publishing** to the specific equipment (no more/no less info): all information maps in master documents in repository are provided with retrieval tags, corresponding to model and its features
  - in the **language of choice**
  - with **metric or British units** of measurement
- Manual content is defined by sales dept and customer
- A few mouse clicks automatically convert retrieved XML file to PDF and e-mail it to user

Atlas Copco manuals

- Each manual is dynamically generated on-the-fly, allowing instantaneous update of content
- Allowing decentralised and fragmented technical authoring and translation
- Reduce/eliminate time-to-market, pre-press, paper, printing, distribution cost

System architecture



Translations

- 12 EC languages
- Project allows increase to 18 languages
- Trados translation software:
  - builds up translation memories per language pair for utmost re-use of already translated sentences
  - enables Multiterm terminology management, **forcing external translators to use pre-determined terminology,** improving quality
  - allows **pre-calculation** of translation cost (once translation memories are built up)

Conclusion

- Optimal procedures in technical document creation and translation :
  - Source text control
  - Terminology management (both source text and translation oriented)
  - Translation management
  - Content management
  - Critical analysis of the needs of the user
  - Workflowmanagement