

Lost for Words

Maximizing Terminological Quality and Value at an LSP

T*rans***F***orm*

Gesellschaft für
Sprachen- und Mediendienste mbH



Why?

- Quality—we need the appropriate term for:
 - the language variant
 - the customer
 - the subject area
 - any applicable standard
- Economics
 - less research necessary
 - fewer complaints
 - increased customer loyalty

What's it about?

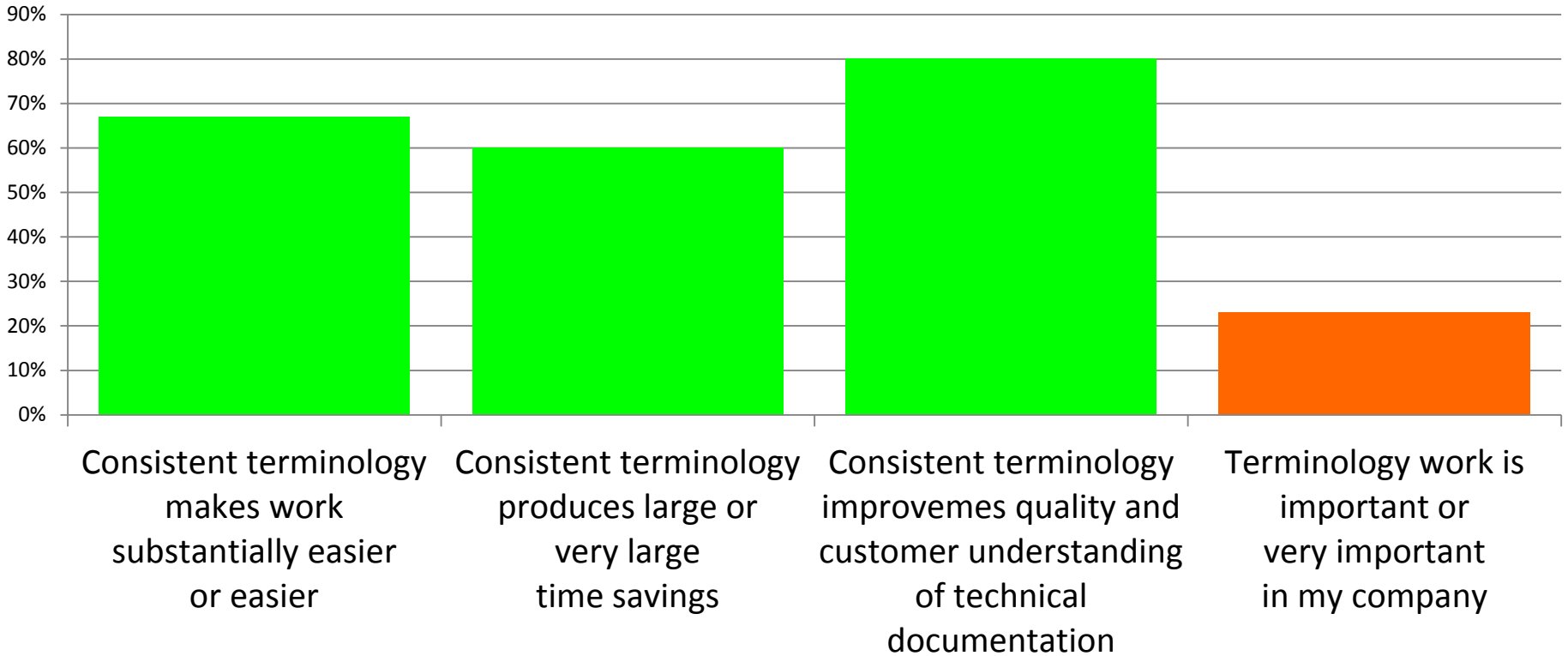
- **“The set of special words belonging to a science, an art, an author, or a social entity...”**

OR

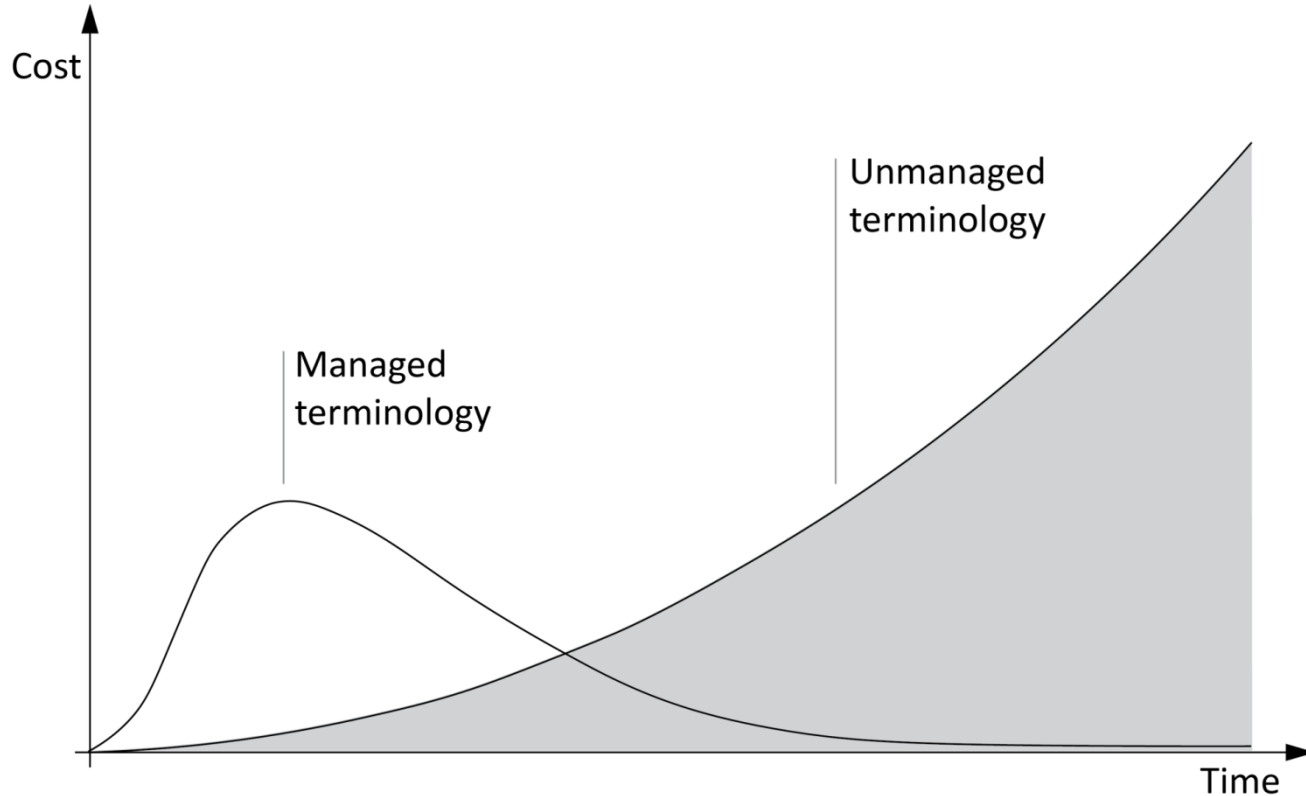
“The language discipline dedicated to the scientific study of the concepts and terms used in specialized languages”

(Pavel, Handbook of Terminology)

Whose job is it anyway?

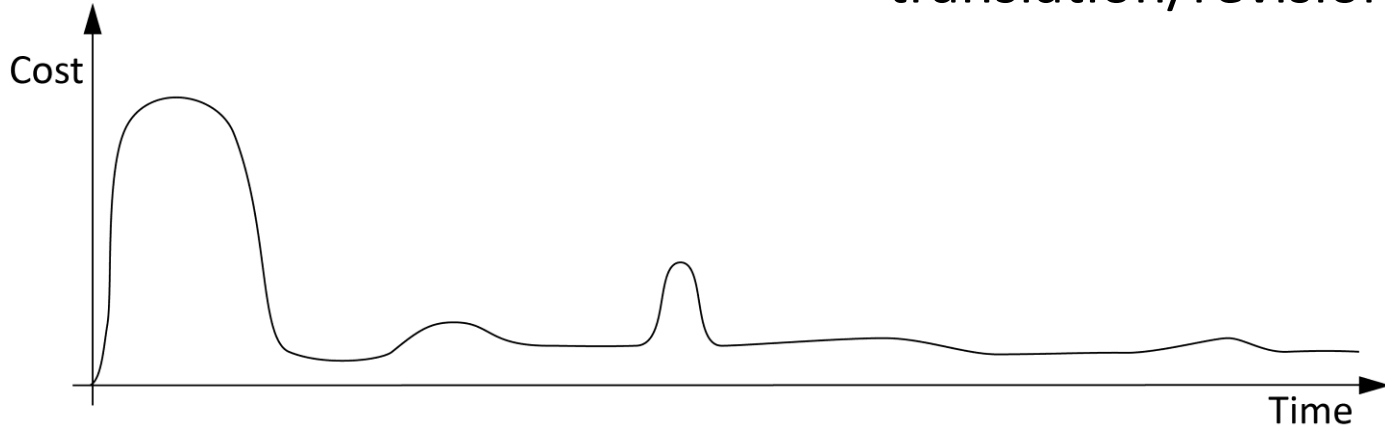


Is it because of the money?



Costs

- Directly attributable
 - e.g. preparation and maintenance for a specific account
- Difficult to isolate from general overhead
 - e.g. terminology work during the translation/revision process



Magazines

- Can cover a wide range of topics
- Can require highly specialist terminology
- Will have recurring headings, features, section titles which must be consistent between issues
- May feature company-specific nomenclature inaccessible from outside
- Can be long-term projects



Folio voucher copies, 2008–2011

Reports

Type	Standards	
Annual report	IAS/IFRS (International Accounting Standards/ International Financial Reporting Standards), GAAP (USA), HGB (Germany)	Different accounting standards, so terms are not conceptually equivalent
CSR	ISO 26000	
Sustainability	GRI (Global Reporting Initiative)	Currently running a call for tenders for the translation of the standards

Subject to regularly changing standards

External risks—what if it's wrong

- Red faces, red numbers or worse?
 - For print jobs, the later the error is discovered, the greater the financial consequences
 - Electronically published material is not subject to such rapid cost escalation as the correction costs are much lower.
 - Less-prestigious work such as technical documentation entails a higher level of risk

Investment risk/reward

Topic range	Regularity, frequency	Examples	Investment	Investment risk	Payoff
Limited	Irregular, frequent	Press releases, technical documentation	Initially high, declining to reasonably constant low value	Low	High (quality, savings)
Very limited	Regular, infrequent	Reports (Annual, CSR, Sustainability)	High, over short period	Low	High (quality, savings)
Wide range of topics	Regular, can be frequent	Magazines	Medium, over long period	Medium	Medium (quality)
Limited	One-off	Pitches	Medium	Very high	High

Interested parties

- Clients. Translations should:
 - conform to existing material
 - not cost much
 - be ready yesterday
- Suppliers. Want work that:
 - can be done economically
 - straightforward, leverage available, no extra work required
 - has appropriate support
 - has reasonable deadlines
- LSPs. Want work that:
 - Allows them to get paid for adding value

Best practices

- Clients
 - Create and use professionally prepared source language terminology
 - Target language terminology review by a native speaker of the target language who is familiar with the domain
 - Provide supporting material before translation start
- Suppliers
 - Conformity and consistency
 - Documentation and communication

LSPs

- Distinguish between different types of client
- Supply terminology in an appropriate form
- Facilitate and encourage supplier feedback
- Remove barriers to cooperation

ISO 17100 and terminology

- Annex A Workflow Section 4: Pre-production processes and activities specifies terminology as optional (4.6.3.2)
- Section 5, Production process requires compliance with client and domain terminology and terminological consistency (5.3.1 a)

Fit for purpose

- Quality is relative to required specification
 - 1) Customer acceptance
 - 2) Consistency
 - 3) Correctness
 - Customer acceptance of terminology trumps other concerns
- For an LSP, quality is a question of ROI

Ideal starting points

- Customer has well-defined source terminology
- Customer has matching target terms
- Customer terminology is available in useable form
 - TBX or other XML
 - In Excel but at least lemmatized and “clean”

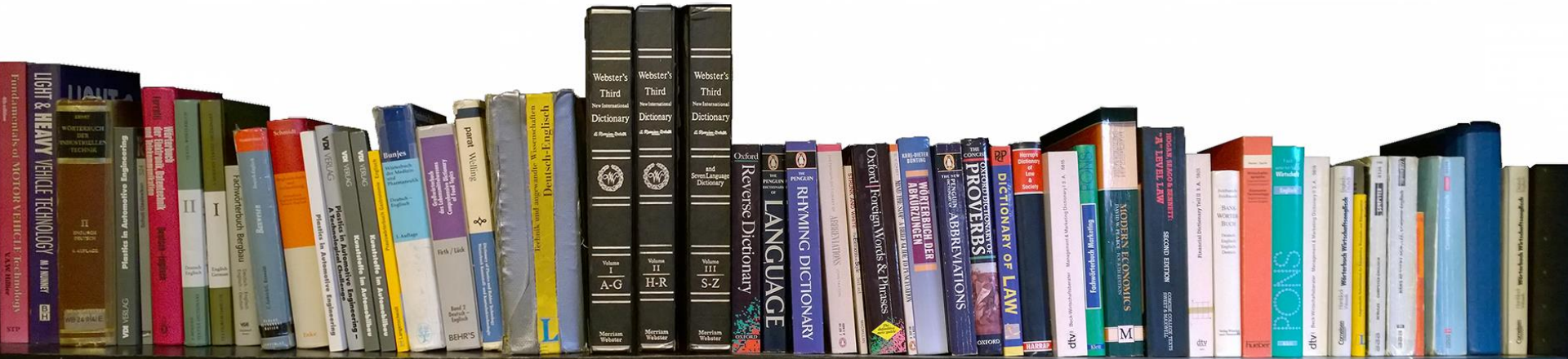
Back in the real world

- Information structured for customer's own use (or not well-structured)
- Inconsistent source prepared by inadequately supervised interns
- Target terms from non-native speakers
- In Excel
- Only available as a (protected) PDF
- “What do you mean, terminology?”

Or any combination of the above

Subject-specific challenges

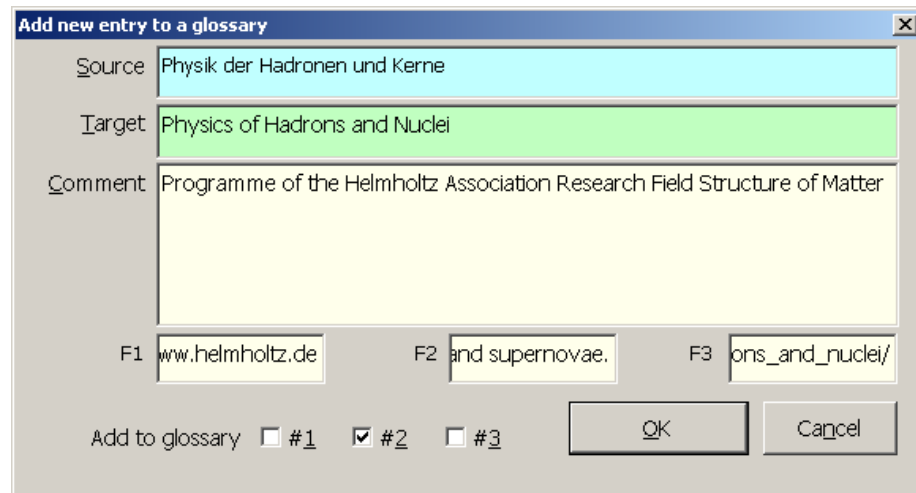
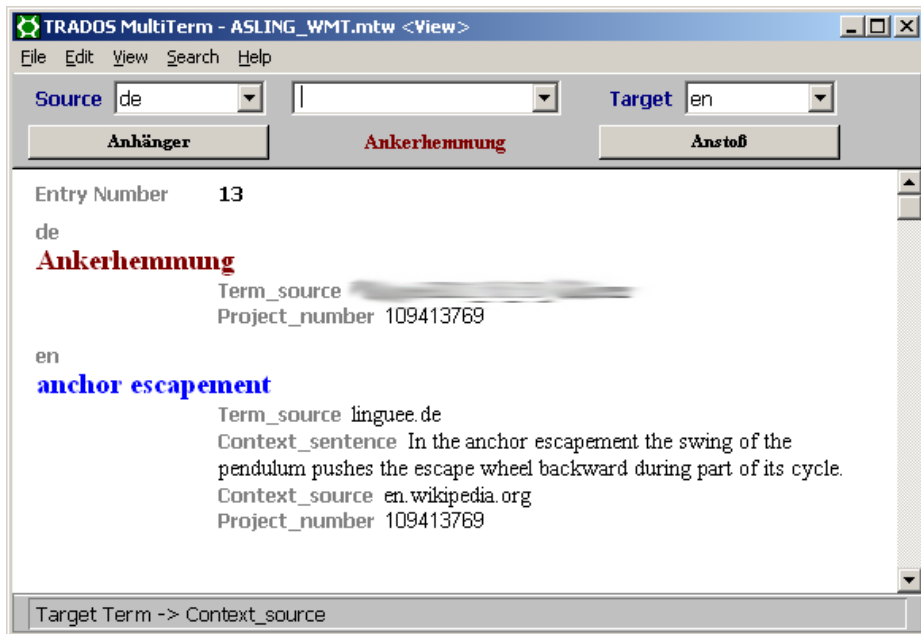
- One-to-many term relationships
- Country-specific topics use different concepts to describe equivalent subjects
- Not all subject areas are well-covered online



Ways and means

- Abolishing duplication of effort
- Reusing the results
- Terminology work never ends, so we need systems to capture, document and manage terminology
- In our case, making use of extensive legacy data

History—MultiTerm and Wordfast



Wordfast

Source Term

Target Term

Comment

F1

F2

F3

memoQ

memoQ

Entry-level data (Concept level)

Transactional

Modified at

Last modified by

Created at

Created by

Metadata

Domain

Subject

Client

Project

Note

ID

Language level

Language variant

Definition

Term level

Matching

Matching

Case sensitivity

Usage

Forbidden

Example

Grammar

Part of speech

Gender (grammer)

Number (grammer)

- No term source field
- No terminological context source field

The TransForm standard terminological record

Konto:	\2011_05_25_Folio_6_de_en	okay
Auftrag:	U4_Portrait_..._N4	zurück
Quellterminologie:		
Quellsprache:	Deutsch	
Context of Source Term:		
Source of context:		
Zielterminologie:		
Zielsprache:	English US	
Source of Term:		
Context of Target Term:		
Source of context:		
Subject Area:		
Kommentar:		
Benutzer:	djc	

Add new entry to a glossary

Source:	Physik der Hadronen und Kerne
Target:	Physics of Hadrons and Nuclei
Comment:	Programme of the Helmholtz Association Research Field Structure of Matter

F1: F2: F3:

Add to glossary #1 #2 #3

OK Cancel

TRADOS MultiTerm - ASLING_WMT.mtw <View>

File Edit View Search Help

Source: de Target: en

Anhänger Ankerhemmung Anstoß

Entry Number 13

de

Ankerhemmung

Term_source: Project_number 109413769

en





anchor escapement


Term_source: linguee.de
Context_sentence: In the anchor escapement the swing of the pendulum pushes the escape wheel backward during part of its cycle.
Context_source: en.wikipedia.org
Project_number 109413769

Target Term -> Context_source

Data from three legacy systems
Data from current system +
New, high-quality terminology





















Concept and language

▼ Entry 						
Operations	Name	Type	Built-in	Required	Default value	Description
	Last modified by	Text	✓			
	Last modified	Date	✓			
	Created by	Text	✓			
	Created on	Date	✓			
	Domain	Text	✓			
	Subject	Text	✓			
	Client	Text	✓			
	Project	Text	✓			
	Image caption	Text	✓			
	Image	Media	✓			
	Note	Text				
 	ID	Number				

▼ Language 						
Operations	Name	Type	Built-in	Required	Default value	Description
	Definition	Text	✓			

- Extra fields to maintain memoQ compatibility

Term

▼ Term						
Operations	Name	Type	Built-in	Required	Default value	Description
	Case sensitivity	Case sensitivity	✓			
	Matching	Matching	✓			
	Forbidden	True/false	✓			
 	Usage example	Text				
 	Part of speech	+{N, Adj, Adv, V, X}				
 	Number (grammar)	+{Sing, Pl}				
 	Gender (grammar)	+{Masc, Fem, N}				
 	Term source	Text				
 	Usage example source	Text				
 	Note	Text				
 	Validation status	1{Unvalidated, Validated}				
 	Validated by	Text				
 	Term type	+{Full_form, Acronym, Abbreviation, Short_form, Variant}				

- Extra fields to maintain memoQ compatibility

A measure of quality

- Term source
 - Proxy for quality rating
 - Customer's term—it must be right
 - Reliable source—good reason for use
 - “I made it up”—see terminological context
 - leo.org or similar—see terminological context
- Terminological context
 - Usage example also supports source
- Source of terminological context
 - Proxy for quality of context

QTerm input/output

- Terminology transfer formats

- Multiterm XML

- TBX, TBX-Basic

- The TBX and MTXML import functions overwrite any existing termbase structure and content

- Excel CSV, TXT

- Can be used to append records to an existing termbase

Excel

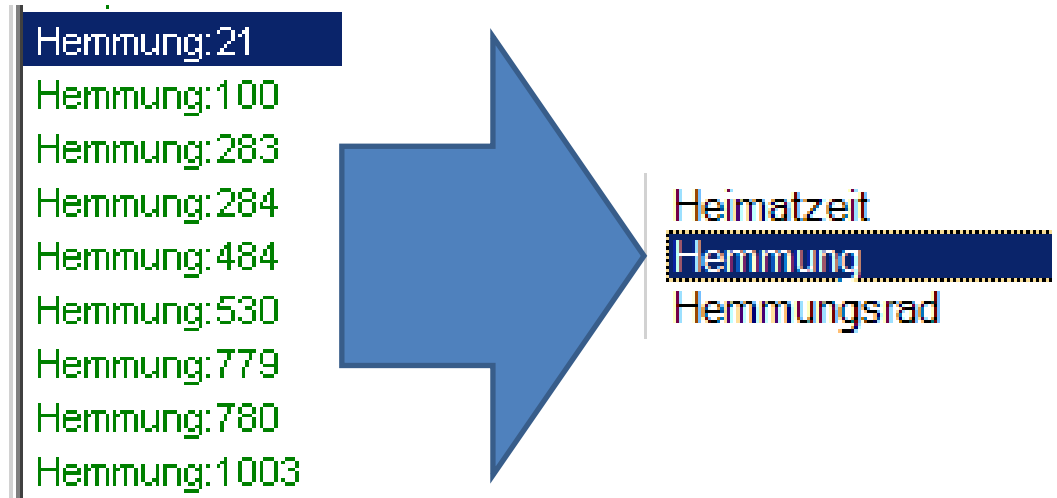
- Number formats and text delimiters change according to region settings
- Not Unicode-friendly (defaults to OS encoding)
- Vast potential for loss or corruption of data
 - " character in string is frequently interpreted by Excel as a text delimiter.
 - ' = feet, minutes (of arc and of time)
 - " = inches, seconds (of arc and of time)

Kit estetico esterno Elegance 1", cofano motore e tetto - argento"
Kit estetico esterno Sport", cofano motore e portellone - bianco"

14 inch Complete Steel Wheel with Winter Tire	14 komplette vinterhjul, stålhjul med vinterdæk"
15 inch Complete Alloy Wheel with Winter Tire	15 komplette vinterhjul, letmetalfælge med vinterdæk"

Reunification

- Define termbase with initial data
- Import legacy data
- Merge duplicates



Combined



English (United Kingdom)

escapement

CaseSense: Near

Forbidden: false

PartialThreshold: Half

Term source: ~~XXXXXX~~ Glossary Oct. 2016



English (United Kingdom)

escapement

CaseSense: Near

Forbidden: false

PartialThreshold: Any

Usage example: In mechanical watches and clocks, an escapement is a device that (1) transfers energy to the timekeeping element and (2) enables counting the number of oscillations of the timekeeping element.

Term source: djc

Usage example source: en.wikipedia.org

Validation status: Validated

Validated by: djc



English (United Kingdom)

escapement

CaseSense: Near

Forbidden: false

PartialThreshold: Half

Usage example: In mechanical watches and clocks, an escapement is a device that (1) transfers energy to the timekeeping element and (2) enables counting the number of oscillations of the timekeeping element.

Term source: ~~XXXXXX~~ Glossary Oct. 2016

Usage example source: en.wikipedia.org

Validation status: Validated

Validated by: djc

Matching

- Different tools, different degrees of control
- Automatic fuzzy matching
 - Case
 - Type of matching
- Manual fuzzy matching
 - Wildcards in terms

Wildcards

memoQ

Wordfast Professional

▼ erteilte* Auftrag*

Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes
Matching	<input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact

```
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilte Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilte Aufträge rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteiltem Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Aufträge rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Aufträgen rpr id=1 transform=close rpr id=0 XXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Auftrags rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Auftrags rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilter Aufträge rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 Erteilte Aufträge rpr id=1 transform=close rpr id=0 XXXXX
```

▼ erteilte* Auftr|g

Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes
Matching	<input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact

```
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilte Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilte Aufträge rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteiltem Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Aufträge rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Aufträgen rpr id=1 transform=close rpr id=0 XXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Auftrags rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilten Auftrags rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilter Auftrag rpr id=1 transform=close rpr id=0 XXXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 erteilter Aufträge rpr id=1 transform=close rpr id=0 XXXX
rpr id=0 X rpr id=0 transform=close rpr id=1 Erteilte Aufträge rpr id=1 transform=close rpr id=0 XXXXX
```

Index	Source	Target
1	erteilte* Auftr*	awarded contract

```
{ut1}X{ut2}{ut3}erteilte Auftrag{ut4}{ut5}XXXXXXXXXXXXXXXXX{ut6}X{ut7}
{ut8}erteilte Aufträge{ut9}{ut10}XXXXXXXXXXXXXXXXX{ut11}X{ut12}
{ut13}erteiltem Auftrag{ut14}{ut15}XXXXXXXXXXXXXXXXX{ut16}X{ut17}
{ut18}erteilten Auftrag{ut19}{ut20}XXXXXXXXXXXXXXXXX{ut21}X{ut22}
{ut23}erteilten Aufträge{ut24}{ut25}XXXXXXXXXXXXXXXXX{ut26}X{ut27}
{ut28}erteilten Aufträgen{ut29}{ut30}XXXXXXXXXXXXXXXXX{ut31}X{ut32}
{ut33}erteilten Auftrags{ut34}{ut35}XXXXXXXXXXXXXXXXX{ut36}X{ut37}
{ut38}erteilter Auftrag{ut39}{ut40}XXXXXXXXXXXXXXXXX{ut41}X{ut42}
{ut43}erteilter Aufträge{ut44}{ut45}XXXXXXXXXXXXXXXXX{ut46}X{ut47}
{ut48}Erteilte Aufträge{ut49}{ut50}XXXXXXXXXXXXXXXXX{ut51}
```


Results of using wildcards

Term as entered	Term as displayed								
<p data-bbox="125 252 958 292">▼ erteilte* Auftrag*</p> <table border="1" data-bbox="125 295 958 396"><tr><td data-bbox="125 295 432 347">Case sensitivity</td><td data-bbox="432 295 958 347"><input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes</td></tr><tr><td data-bbox="125 347 432 396">Matching</td><td data-bbox="432 347 958 396"><input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact</td></tr></table>	Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes	Matching	<input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact	<p data-bbox="1000 210 1792 303">ASLING term recognition test  Modified 27/10/2016 11:39 by admin </p> <table border="1" data-bbox="1000 303 1792 442"><tr><td data-bbox="1000 303 1174 380">ger</td><td data-bbox="1174 303 1792 380">erteilte Auftrag</td></tr><tr><td data-bbox="1000 380 1174 442">eng-US</td><td data-bbox="1174 380 1792 442">awarded contract</td></tr></table>	ger	erteilte Auftrag	eng-US	awarded contract
Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes								
Matching	<input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact								
ger	erteilte Auftrag								
eng-US	awarded contract								
<p data-bbox="125 533 958 574">▼ erteilte* Auftr g</p> <table border="1" data-bbox="125 577 958 678"><tr><td data-bbox="125 577 432 628">Case sensitivity</td><td data-bbox="432 577 958 628"><input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes</td></tr><tr><td data-bbox="125 628 432 678">Matching</td><td data-bbox="432 628 958 678"><input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact</td></tr></table>	Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes	Matching	<input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact	<p data-bbox="1000 503 1792 596">ASLING term recognition test  Modified 27/10/2016 11:43 by admin </p> <table border="1" data-bbox="1000 596 1792 734"><tr><td data-bbox="1000 596 1174 672">ger</td><td data-bbox="1174 596 1792 672">erteilte Auftrg</td></tr><tr><td data-bbox="1000 672 1174 734">eng-US</td><td data-bbox="1174 672 1792 734">awarded contract</td></tr></table>	ger	erteilte Auftrg	eng-US	awarded contract
Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes								
Matching	<input checked="" type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact								
ger	erteilte Auftrg								
eng-US	awarded contract								

- Source terms displayed without wildcard characters.
- Termbase no longer fully bidirectional

Discrimination

memoQ Web Search

Search

Project home

Source

1. B2C·Car·Sharing:

B2C·Car·Sharing:

▼ German (Germany)

▼ ca.

Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes
Matching	<input type="radio"/> Custom <input checked="" type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact
Forbidden	<input type="radio"/> Yes <input checked="" type="radio"/> No

Source

1. B2C·Car·Sharing:

2. Ca·100·Versuche

Approx·100·attempts

3. Es·gab·ca·100·Versuche

▼ German (Germany)

▼ ca.

Case sensitivity	<input type="radio"/> No <input checked="" type="radio"/> Permissive <input type="radio"/> Yes
Matching	<input type="radio"/> Custom <input type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input checked="" type="radio"/> Exact
Forbidden	<input type="radio"/> Yes <input checked="" type="radio"/> No

50% prefix vs. Fuzzy

16. Betriebswirte, Physiker, Chemiker und
which includes management experts, p

Changed admin 14/11/2016 18:18

View pane

▼ Betrieb	
Case sensitivity	<input type="radio"/> No <input type="radio"/> Permissive <input checked="" type="radio"/> Yes
Matching	<input type="radio"/> Custom <input checked="" type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact
Forbidden	<input type="radio"/> Yes <input checked="" type="radio"/> No

16. Betriebswirte, Physiker, Chemiker und
which includes management experts, p

Changed admin 14/11/2016 18:18

View pane

▼ Betrieb	
Case sensitivity	<input type="radio"/> No <input type="radio"/> Permissive <input checked="" type="radio"/> Yes
Matching	<input type="radio"/> Custom <input type="radio"/> 50% prefix <input checked="" type="radio"/> Fuzzy <input type="radio"/> Exact
Forbidden	<input type="radio"/> Yes <input checked="" type="radio"/> No

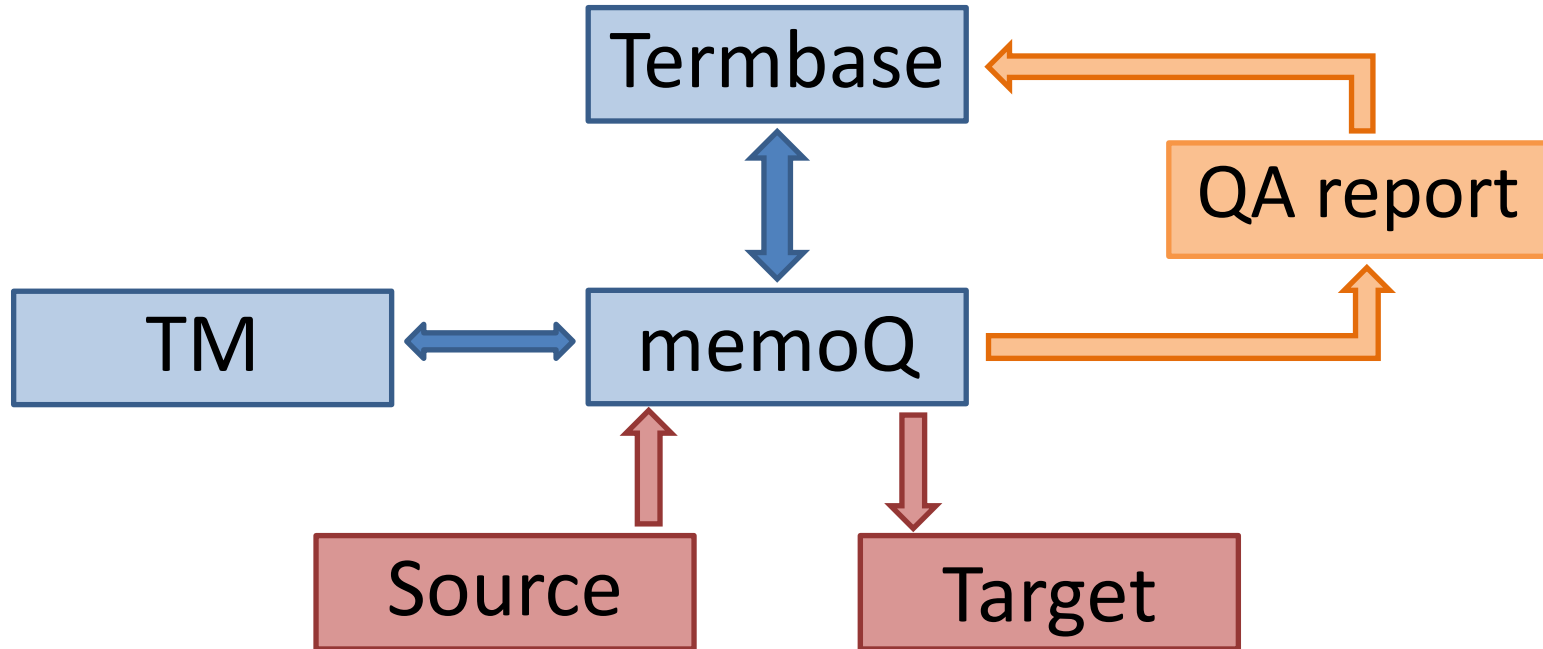
Composite terms

<input type="checkbox"/>	★	Betrieb	facility
<input type="checkbox"/>	☆	Betriebswirt	management expert

15.	anzutertigen, die vom Vorsitzenden od geleitet hat, unterzeichnet werden.
	Betriebswirte, Physiker, Chemiker und
16.	which includes management experts, p
Changed admin 14/11/2016 18:18	

▼ Betriebswirt	
Case sensitivity	<input type="radio"/> No <input type="radio"/> Permissive <input checked="" type="radio"/> Yes
Matching	<input type="radio"/> Custom <input checked="" type="radio"/> 50% prefix <input type="radio"/> Fuzzy <input type="radio"/> Exact
Forbidden	<input type="radio"/> Yes <input checked="" type="radio"/> No

Iterating with QA reports



Lost for Words

Maximizing Terminological Quality and Value at an LSP

References

- 3 <https://en.wikipedia.org/wiki/Terminology>
- 3 <http://itia.ir/farsi/documents/ha.pdf>
- 3 Warburton, K. after Rondeau, G. *Tekom Proceedings tcworld 2013, CHAT 1.*
- 5 R. Herwartz: *Wann macht sich Terminologie bezahlt? Erstellung einer Kosten-Nutzen-Analyse*, in: *technische Kommunikation 5/2011*, pub. tekomp
- 5 A. Zerfaß: *Experiences in Practical Terminology Management*, in conference proceedings, tekomp-Jahrestagung und tcworld conference 2013
- 5 Matthias Kurrus: *Erfolgsfaktor Terminologiemanagement: Nutzen, Kosten, Rentabilitätscheck*, in conference proceedings, tekomp-Jahrestagung und tcworld conference 2014
- 6 TSS2009_FS_EconomicIssues, Prof Dr Frieda Steurs, Lessius/KULeuven
- 33, 34 https://www.wordfast.net/wiki/Fuzzy_Terminology_Recognition
- 33, 34 WFC see manual p. 33 onward.

T*rans***F***orm*

Gesellschaft für
Sprachen- und Mediendienste mbH