

## The Power of Words... Why Smart Companies Create Their Own Dictionaries

Presentation to the OMG Plenary

Kara Warburton, Consultant at Termologic and  
Chair of ISO TC37  
[kara@termologic.com](mailto:kara@termologic.com)

Dec 10, 2014

# Ter·mo·lo·gic

---

## Agenda

- What is terminology?
- Why are companies managing terminology?
- What are some of the useful applications?

# Ter·mo·lo·gic

---

Enterprises and institutions are developing electronic dictionaries that contain their proprietary “terminology”

IBM, Microsoft, SAP, SAS, Hewlett-Packard, Symantec, Agilent, Rockwell Automation, MedTronic, Scania, Bosch, Philips, Novell, Fuji Xerox, ISO, IMF, Avaya, Sony, Yamaha, Dell, Toyota, Cisco, Intel, Google, GE, Adobe, Oracle, Boeing, FMGlobal, Huawei, NATO, UN, EU, etc...

and others....

Why?

# Ter·mo·lo·gic

---

Terms --  
*linguistic expressions of  
significant concepts --*  
are the building blocks  
of communication that  
drive the global economy

# Ter·mo·lo·gic

---

The World Bank offers support to developing countries through policy advice, research and analysis, and technical assistance. Our analytical work often underpins World Bank financing and helps inform developing countries' own investments. In addition, we support capacity development in the countries we serve. We also sponsor, host, or participate in many conferences and forums on issues of development, often in collaboration with partners.

# Ter·mo·lo·gic

---

Our industry-leading technology searches the lowest fares from hundreds of airlines, and gives you access to Web Fares. Choose your plane seat with Seat Maps, and know when fares to your favorite destinations drop with Fare Watcher.

Print out a custom mini-guide before you go using Frommer's Guides. Browse our deals pages for popular destinations, including insider tips, great deals, and travel basics.

## What are we talking about?

- a *term* is an expression (words, picture) that has a distinct meaning (expressable via a definition) in a given *context*
- a *context* is the real use of language by people, such as any kind of document, an email, a web site, a dialog, a video, a contract, marketing collateral, etc.
- *terminology management* includes collecting *terms* that people use, describing what they mean (writing definitions), identifying synonyms and homonyms, establishing and monitoring correct usage (norms), establishing relations between terms (taxonomies, ontologies), determining equivalents in other languages (translation, localization), and structuring and storing this work so that it can serve as a useful resource

Again, why do those companies manage their terminology?



# Ter·mo·lo·gic

Companies “manage” terminology to improve their bottom line

- increase employee productivity
- reduce production costs
- improve information content
- improve information retrievability
- compete in global markets
- leverage NLP applications
- avoid risk and loss due to ambiguity and inconsistency



## Terminology data drives quality in informational content and consistency in business processes

- Informational content is increasingly subject to manipulation by various Natural Language Processing (NLP) applications for knowledge mining, reuse, content management, etc.
- Terminology provides structured lexical constructs and domain knowledge that NLP applications need in order to process unstructured information
- Lexical resources (next-generation “dictionaries”) can be leveraged for diverse applications: translation and localization, search, indexing, controlled authoring, glossaries, classification systems, thesauri, word lists for various text analysis applications, etc.
- Managed terminology facilitates regulatory compliance

# Ter·mo·lo·gic

---

Some examples...

## Search query synonym suggestions

Search for:

Did you mean: [\*fatal error\*](#) , [\*irrecoverable error\*](#) , [\*unrecoverable error\*](#) ?

## Search query contraction

### Narrow Your Search

- › ethernet switches
- › **limit switches**
- › interlock switches
- › disconnect switches
- › safety switches
- › stratix switch
- › systems ethernet switches
- › catalog safety switches
- › control load switches
- › ethernet switches specifications

### Narrow Your Search

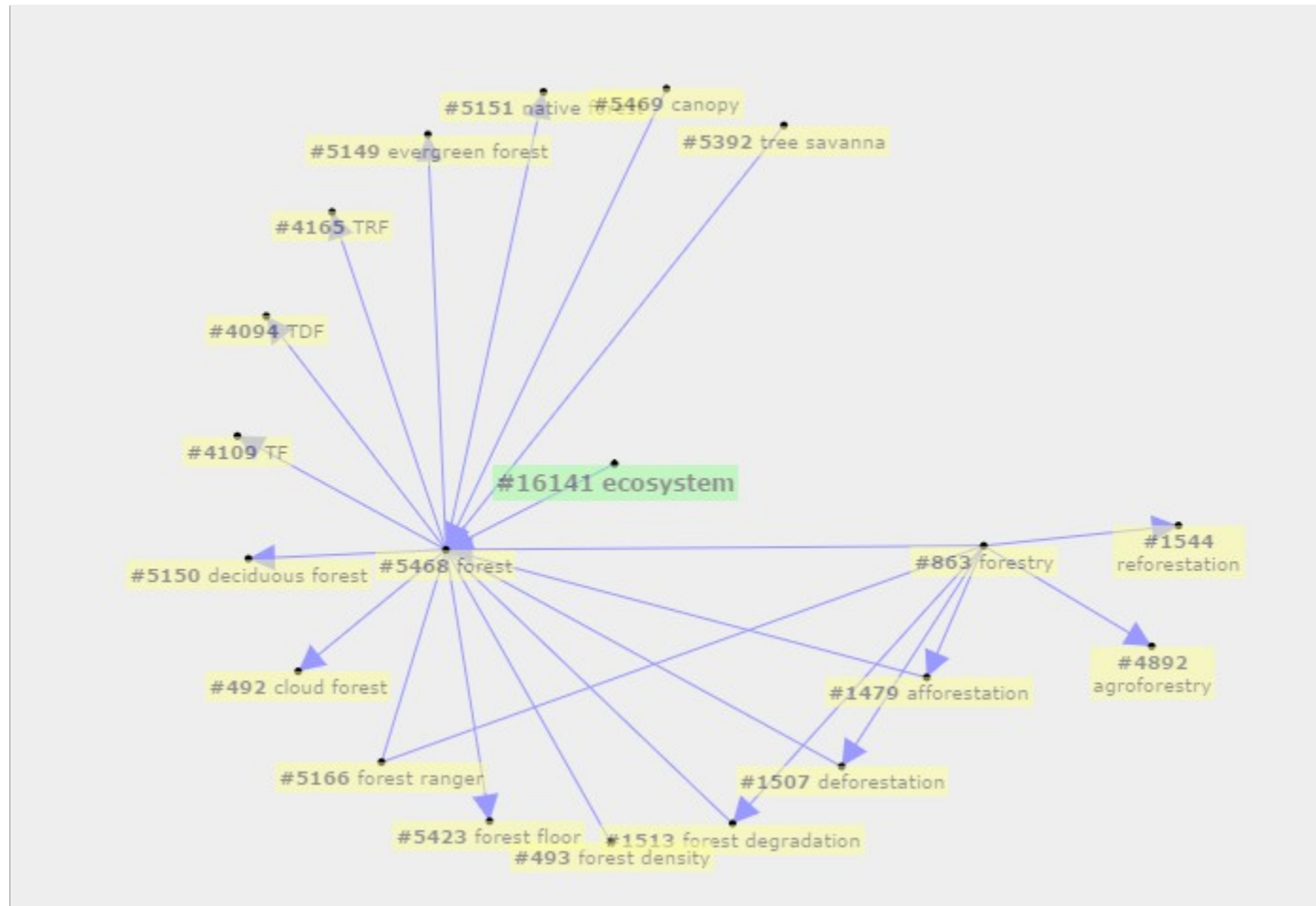
- › **safety limit switches**
- › switches technical data limit
- › technical data limit td001
- › limit switch installation
- › limit switch publications
- › devices switching elements electromechanical
- › limit switch family offers
- › limit switch publications literature
- › switch family offers full
- › switch publications literature library

### Narrow Your Search

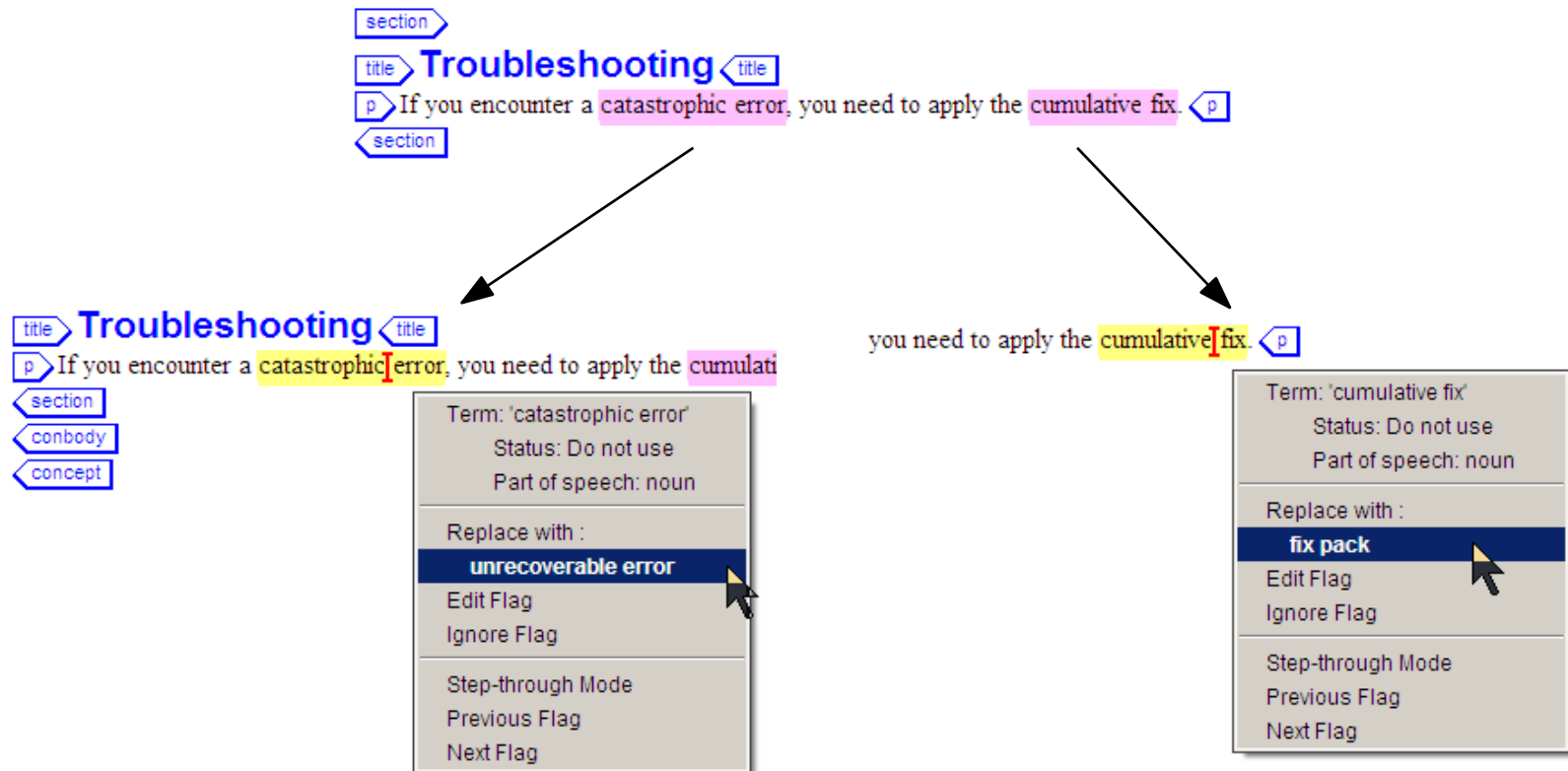
- › limit switches
- › safety limit
- › safety products
- › limit switch family offers full
- › switch family offers full range
- › safety standard sensing applications
- › 802t plug safety limit switches
- › limit switches include
- › safety switches
- › software updates product safety advisories

# Ter·mo·lo·gic

termweb



## Controlled authoring



# Ter·mo·lo·gic

---

## English

### unrecoverable error

Part of speech Noun

Usage status Preferred

Created by	Kara Warburton (kara)	Creation date	2014-10-10 07:46:46
Change date	2014-10-10 07:46:46	Term ID	20796-1

## English

### catastrophic error

Part of speech Noun

Usage status Not recommended

## English

### cumulative fix

Part of speech Noun

Usage status Not recommended

Created by	Kara Warburton (kara)	Creation date	2014-10-10 07:48:36
Change date	2014-10-10 07:48:36	Term ID	20797-2

## English

### fix pack

Part of speech Noun

Usage status Preferred



## Terminology – a structured language *resource* supporting effective communication

Concept ID 1517

 Add Comment



**Dictionary** World Bank Terminology  
**Section** MC TermBase  
**Domain** Environment and Natural Resources > Pollution Management



**Source of entry** WDR10-TermBase export  
Kyoto Glossary  
WDR 2010 Glossary

Term ID 1517-1

### English

#### greenhouse gas

**Part of speech** Noun

**Term type** Full form

**Definition** Any of the atmospheric gases that cause climate change by trapping heat from the sun in Earth's atmosphere—producing the greenhouse effect.

**Source of definition** Kyoto Glossary, WDR 2010 Glossary

**Context sentence** The most common greenhouse gases are carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), ozone (O<sub>3</sub>), and water vapor (H<sub>2</sub>O).

**Source of context** Kyoto Glossary, WDR 2010 Glossary

Term ID 1517-13

### English

#### GHG

**Source of term** Kyoto Glossary, WDR 2010 Glossary, Acronyms in World Bank Documents

**Part of speech** Noun

Term ID 1517-11

### Spanish

#### gas de efecto invernadero

**Source of term** Spanish-English Glossary

**Part of speech** Noun

**Term type** Full form

## SBVR - Semantics of Business Vocabulary and Business Rules

- An OMG standard that defines methods for creating unified business dictionaries and rules
- Based on terminology management principles from ISO TC37, yet provides significant additional semantic rigour
- Following SBVR brings conventional terminology management practices and data structures to a higher level, required for rendering terminology resources multipurpose

## Semantic rigour in concept description

- Coherence, cohesion, logic, transparency, systematicity, connectivity
- Is essential for:
  - Creating usable and effective definitions
  - Establishing relations between concepts
  - Raising the quality and repurposability of terminology resources
- Is often low in existing terminology databases... but can be raised with SBVR

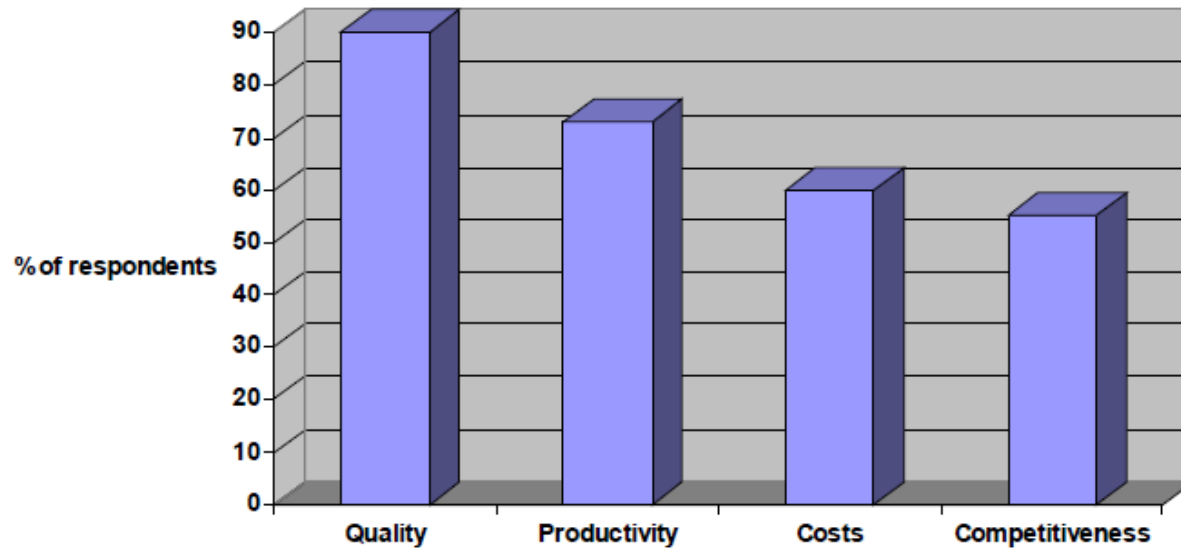
## The scope of potential needs...

*Terminology work is necessary  
for between 4% and 6%  
of all words in a text*

This means... every year... work is being done to figure out or fix...

- 21 million words for IBM
- 2.2 million words for the WTO
- 1.2 million words for the World Bank
- 400,000 words for the average mid-sized global company

## Benefits of managing terminology (LISA survey)



<b>When you evaluate the services of an expert in terminology, would you say that those services ...</b>	<b><u>LARGE FIRMS</u></b>
Improve the quality of the work	86% A lot*: 77%
Add value to the work of the company	84% A lot: 63%
Improve the efficiency of client services	83% A lot: 63%
Improve the efficiency of writing	83% A lot: 68%
Reduce the risks of errors in terms of client services	82% A lot: 77%
Improve your company's customer satisfaction	78% A lot: 64%
Improve the efficiency of translation	76% A lot: 74%
Improve your company's customer loyalty	64% A lot: 50%
Reduce the costs related to writing/editing or translation	60% A lot: 52%

## The scope of errors/inconsistencies... what surveyed professionals say

- 84 % report that inconsistent terms are used very frequently in various departments/sections (Tekom)
- 70 % report that inconsistent terms are used very frequently in various documents (Tekom)
- 95% of translators regularly notice inconsistent terminology in the source language (Hurst, SDL)
- 87% of translators say their productivity would improve with access to terminology resources (Hurst)

Language is constantly changing. People use language in different ways. Without establishing clarity at the institutional level, everything is open to interpretation.

## Relations

Generic: ↓ 11067 - biological diversity

Show graph »



**Source of entry** GEF Glossary

**Definition** One of the program priorities established by the Conference of the Parties to the Convention on Biological Diversity. Also an operational program.

<b>Created by</b> Administrator (admin)	<b>Creation date</b> 2013-02-09 08:09:41	<b>Changed by</b> Administrator (admin)
<b>Change date</b> 2013-02-09 08:09:41	<b>Term ID</b> 11039-1	

## English

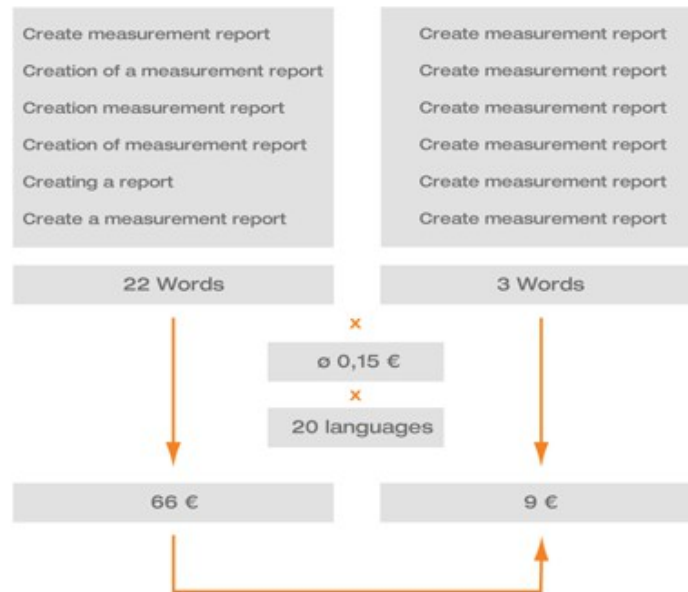
### agricultural biodiversity

**Part of speech** Noun

- English** agrobiodiversity
- Spanish** agrobiodiversidad
- Spanish** biodiversidad agrícola
- French** biodiversité du paysage agricole
- French** valorisation agricole de la biodiversité
- French** diversité des espèces élevées
- French** diversité des espèces cultivées
- French** biodiversité agricole
- French** biodiversité des cultures
- French** agrobiodiversité

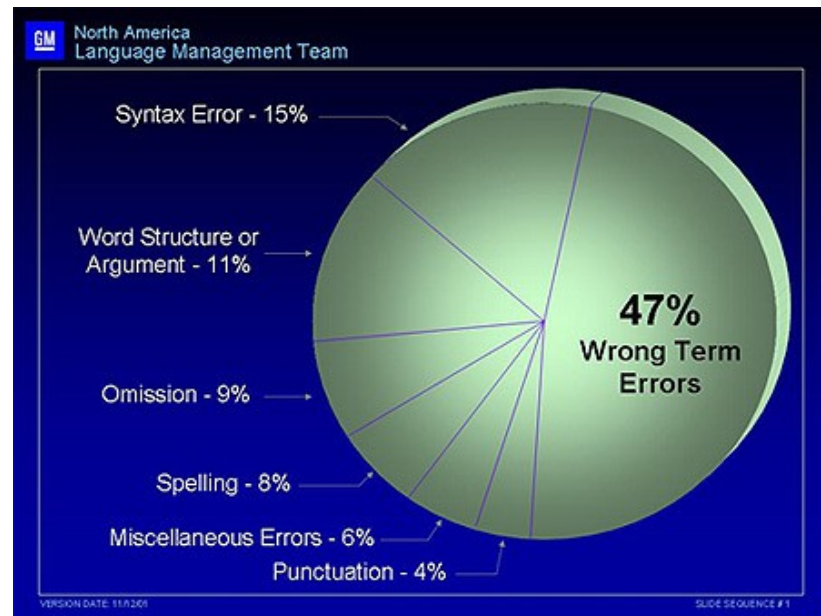


## The cost of inconsistency



# Ter·mo·lo·gic

---



Source: Rick Woyde

## Immeasurable benefits



Risk reduction

Avoidance of lawsuits

Regulatory compliance

ISO 9000 certification

Lower production costs

Customer satisfaction

Brand reinforcement

Better search results

Heightened content reusability

Interoperability with suppliers and partners

Easier transitioning of new employees

Vendor mobility

Increased market opportunity

Preparedness for NLP

Responsiveness to mergers and acquisitions

## Final word

Through SBVR and other standards and implementations built on it, the OMG gains a reputation as an standards body that not only develops modelling standards, but also attempts to address the challenges of implementing those standards in business environments by infusing semantic rigour in commercial information content.

On behalf of ISO TC37, I would like to express gratitude to the OMG for the significant work put into this important standard. Not only does it serve the business community but it brings great value to ISO TC37 standards as well.