
Terminology standardization, terminology management and best practices

Or... what's going on with terminology?

Kara Warburton
kara@termologic.com

Agenda

- Introspection

*"Knowing yourself is the
beginning of all wisdom."*

~Aristotle



EPIC Inspirational Quotes

What is terminology?

- *terminology* is the science studying the structure, formation, development, usage and *management of terms*
- *terminology management* includes collecting *terms*, defining their meaning, indicating their correct usage, translating terms, and storing this work so that it can be useful to others

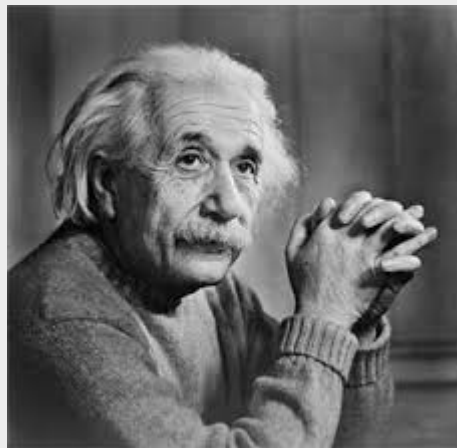
But what is a *term*?



According to ISO TC37

A term is a designation of a concept belonging to a language for specific purposes (LSP)

- an *LSP* is a language used in a *subject field*
 - a *subject field* is a field of *special knowledge*
 - *special knowledge* is ????

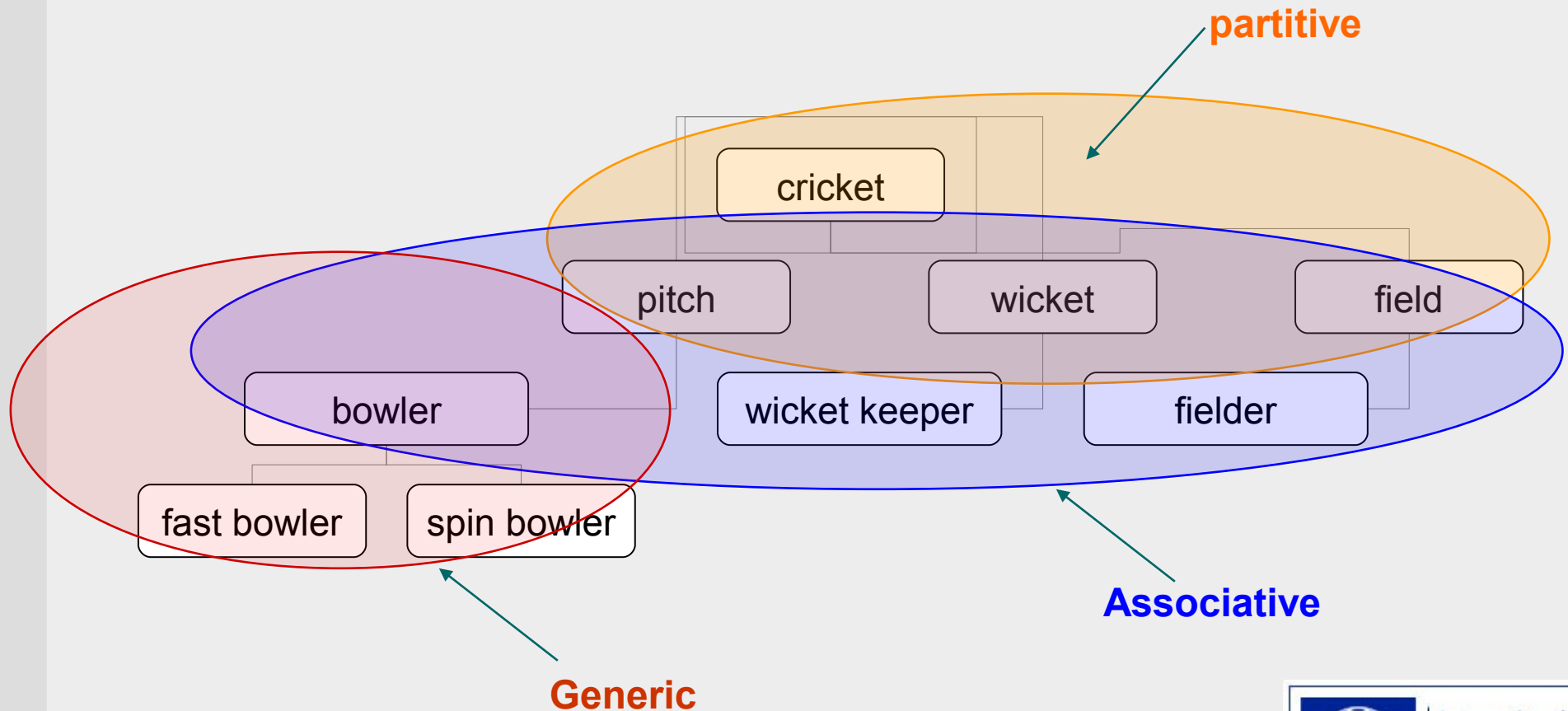


The truth is,
we don't really know
what a term is.



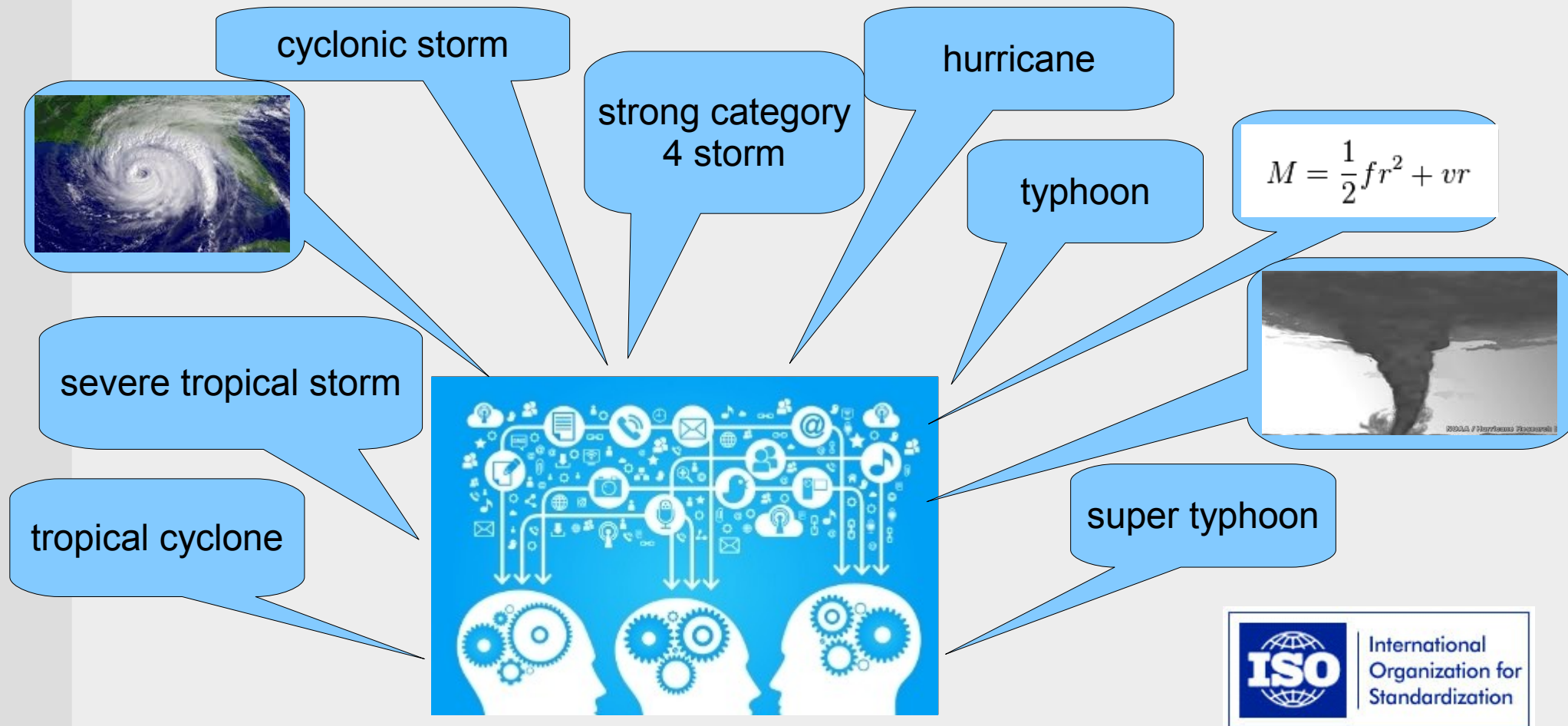
General theory of terminology (GTT)

A term is the designation of an **object**, the conceptualization of which can be classified into a **system of concepts**



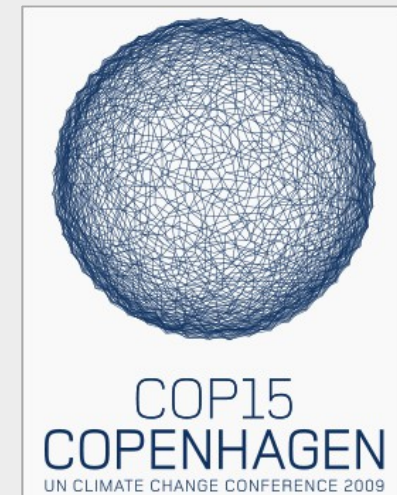
Socio-cognitive theory of terminology

A term is a **natural language** representation of a **unit of understanding**, considered **relevant to given purposes, applications, or groups of users**



Lexico-semantic theory of terminology

A term is a meaningful **text unit** that is defined according to the position it occupies in a **specialized corpus** considered **useful for an application**



Communicative theory of terminology

Terms are **lexical units** that, through **pragmatic** conditions and **communication** situations, can assume a terminological value.



Textual theory of terminology

A term is a **construct** that takes shape through an analysis which gives consideration to **corpus evidence**, validation by **subject-matter experts**, and the **purpose of the terminographical product**

shield cables 3-phase, brake, thermal switch MF-16MF-17 MF-S Low-profile
differential input+ S1 11 Motor thermal switch (normally closed) TS 4 Cosine
accessories thermal switch thermal switch The thermal switch and brake
Only passive shunts with a thermal switch are wired to the TS connector
wires, brake wires, and thermal switch wires. To improve the EMC
(+5V) EPWR_5V 12 Motor thermal switch- TS- 6 Common These
Connector Resistor External Thermal Switch(TS) Connector ThermalSwitch
input+ BM+ / COS+ 11 Motor thermal switch (normally closed) TS 4 BM- /
input- COS- 11 Motor thermal switch+ TS+ 5 Encoder power (+5V)
Connector External Thermal Switch(TS) Connector Refer to the
shunt does not have a thermal switch, leave the jumper (default

TS

Term type: Acronym

termweb

Section Rockwell
Domain Semiconductor and electronics

Relations

Generic: ↑ 1 - switching device Partitive: ↓ 29 - thermal switch jumper

Show graph »



Explanation

English

thermal switch

Part of speech Noun

Context The 1326AB (resolver) power cables (catalog number 1326-CPx1T-L-xxx) contain the three-phase wires, brake wires, and thermal switch wires.

English

temperature switch

Part of speech Noun

French

interrupteur thermique

Part of speech Noun

Gender Masculine

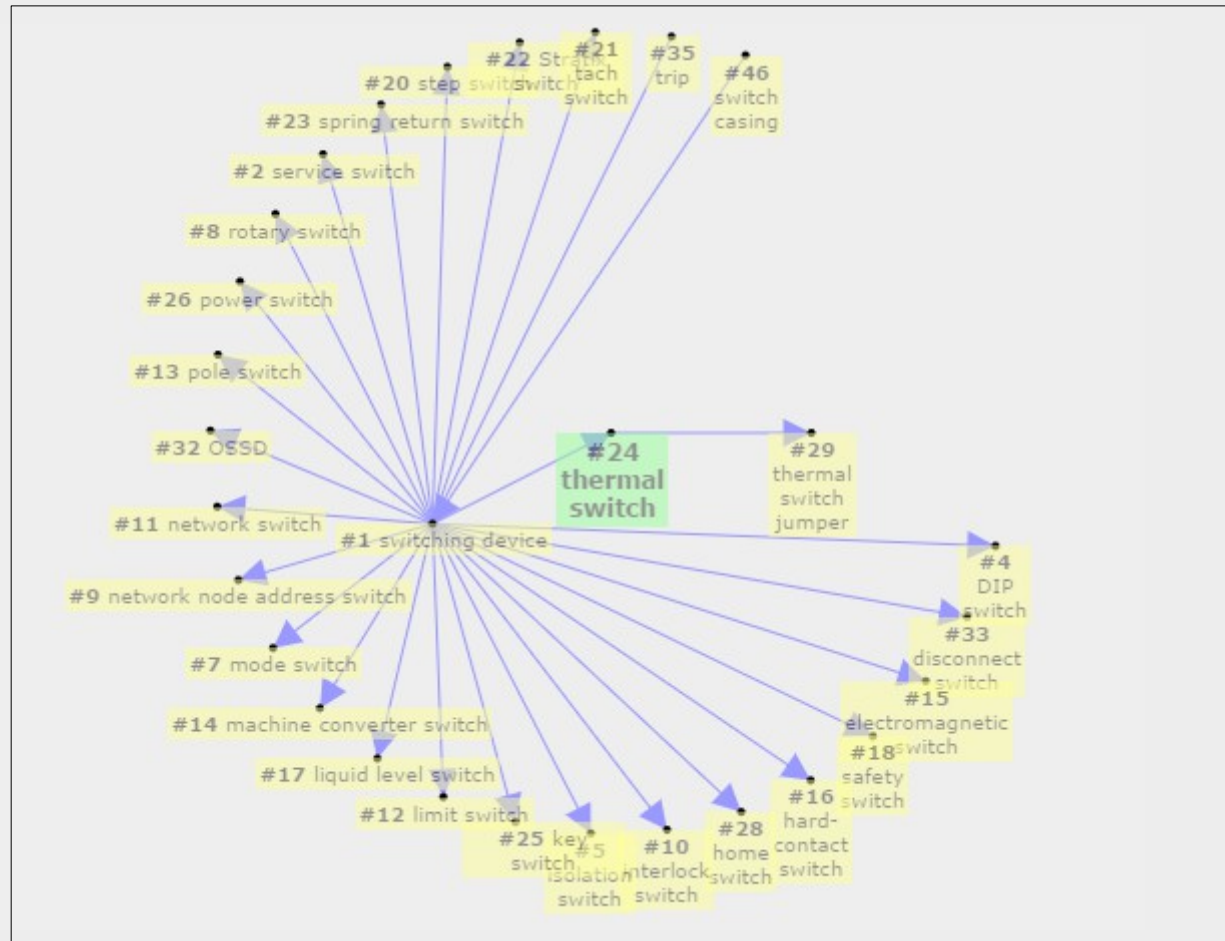
Spanish

interruptor térmico

Part of speech Noun



International
Organization for
Standardization



SDL Trados Studio - Project 1

File Edit View Translation Format Project Tools Help

Open Document... New Project... Open Package... Open Server Project...

Display: All segments Containing: In Source

Editor

Thyristorarten.docx.sdlxliff

Project 1_DE-UK_Memory_2012_02,Project 1_DE-UK-Memory,120309-Delta_120309-Delta - Translation Results

Project Settings...

Technische-Daten: Asymmetrischer Thyristor, Ausschaltthyristor und sonstige Thyristorarten.

No matches found.

Term Recognition

asymmetrischer Thyristor
asymmetrical thyristor
Ausschaltthyristor
turn-off thyristor
Thyristor
thyristor

Project 1_DE-UK_Memory_2012_02,Project 1_DE... DE-UK_Memory_2012_02,Project 1_DE-UK_Mem... Comments Messages

Thyristorarten.docx.sdlxliff [Translation]*

Thyristorarten.docx	Thyristorarten.docx	
1 Technische-Daten	100% Technical data	P
2 Technische-Daten: Asymmetrischer Thyristor, Ausschaltthyristor und sonstige Thyristorarten.	Technical data: a	P
3 Modell	Model	
4 Produkt-Bezeichnung		
5 Umgebungstemperatur-während-Betrieb		
6 Umgebungstemperatur-während-Lagerung		
7 3TH6330-1FR06		
8 Asymmetrischer Thyristor		
9 -25...+60-°C		
10 -55...+80-°C		
11 3TH6330-1FR13		
12 Ausschaltthyristor		

Home

Projects

Files

Reports

Editor

Translation Memories

Termbase Search

Preview

View: Real-time Preview

Source Side-by-side Target

Click here to generate initial preview

INS 83.82% 11.76% 4.41% Chars: 17



Terms are the fundamental building blocks of communication that drives the global economy

Take advantage of SQL Server 2008 R2 features such as partitioning, snapshot isolation, and 64-bit support, which enable you to build and deploy the most demanding applications. Leverage enhanced parallel query processing to improve query performance in large databases. Increase query performance further with star schema query optimizations.

(Text from Microsoft SQL Server 2008)

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.

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 is the purpose of managing terminology?

- Support global communications?
- Promote multilinguality, multiculturalism?
- Reduce economic inequality?
- Are these goals restricted to “special knowledge”?

NLP applications need terminology

- Terminology is domain knowledge that many NLP or text analytics applications need in order to process unstructured information
- A properly-designed termbase can be leveraged for diverse applications: MT, search, indexing, CAT, controlled authoring, glossaries, classification systems, thesauri, word lists for various text analysis applications, etc.


Types of data required

- Controlled authoring
 - synonyms, with preferred term marker
- Search engines
 - synsets – synonyms, abbreviations, variants
 - broader and narrower terms – expand or narrow the search
 - subsets, for faceted search – product names, brands, etc.
- Machine translation
 - grammatical information: part of speech, gender, transitivity
 - headword of compounds
 - subject fields
- Text analytics
 - inflected forms

Search query expansion

Search for: **within** ☒ ☒

Language: ☒ in English ☐ in


Did you know?  Who has been recognized as a Search Champion of the month? [Here they are.](#)

1 - 10 of about 1,834 for catastrophic error sorted by relevance

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

Search for: **within** ☒ IBM w3 intranet ☐ www.ibm.com

Language: ☒ in English ☐ in

 Have you read the latest news..."w3 Search: 2009 Update" [Here's where you can find it.](#)

1 - 10 of about 25,571 for Storage Manager sorted by relevance

Did you mean: [Tivoli Storage Manager S/390 Edition](#) , [Tivoli Storage Manager](#) , [Tivoli Storage Manager Enterprise Edition](#) ?

Controlled authoring



Diagram illustrating the first step of the controlled authoring process:

- Section: Troubleshooting
- Paragraph: If you encounter a catastrophic error, you need to apply the cumulative fix.
- Context menu for 'catastrophic error':
 - Term: 'catastrophic error'
 - Status: Do not use
 - Part of speech: noun
 - Replace with : **unrecoverable error**
 - Edit Flag
 - Ignore Flag
 - Step-through Mode
 - Previous Flag
 - Next Flag

Diagram illustrating the second step of the controlled authoring process:

- Paragraph: you need to apply the cumulative fix.
- Context menu for 'cumulative fix':
 - Term: 'cumulative fix'
 - Status: Do not use
 - Part of speech: noun
 - Replace with : **fix pack**
 - Edit Flag
 - Ignore Flag
 - Step-through Mode
 - Previous Flag
 - Next Flag

Managing terminology is really about managing lexical units of any kind to support the needs of an application or a stakeholder in a given situation.

Managing “lexical units” would be more appropriately named “lexicology”.

But that's not possible.

Lexicology already exists as a field with its own theory and methodology

Lexicology

- Study of words and their meanings
- Preparation of glossaries and dictionaries
- Headword-based structure
- Descriptive

Terminology

- Study of concepts and their designations
- Preparation of terminological entries
- Synset structure (concept orientation)
- Often prescriptive

Concept orientation

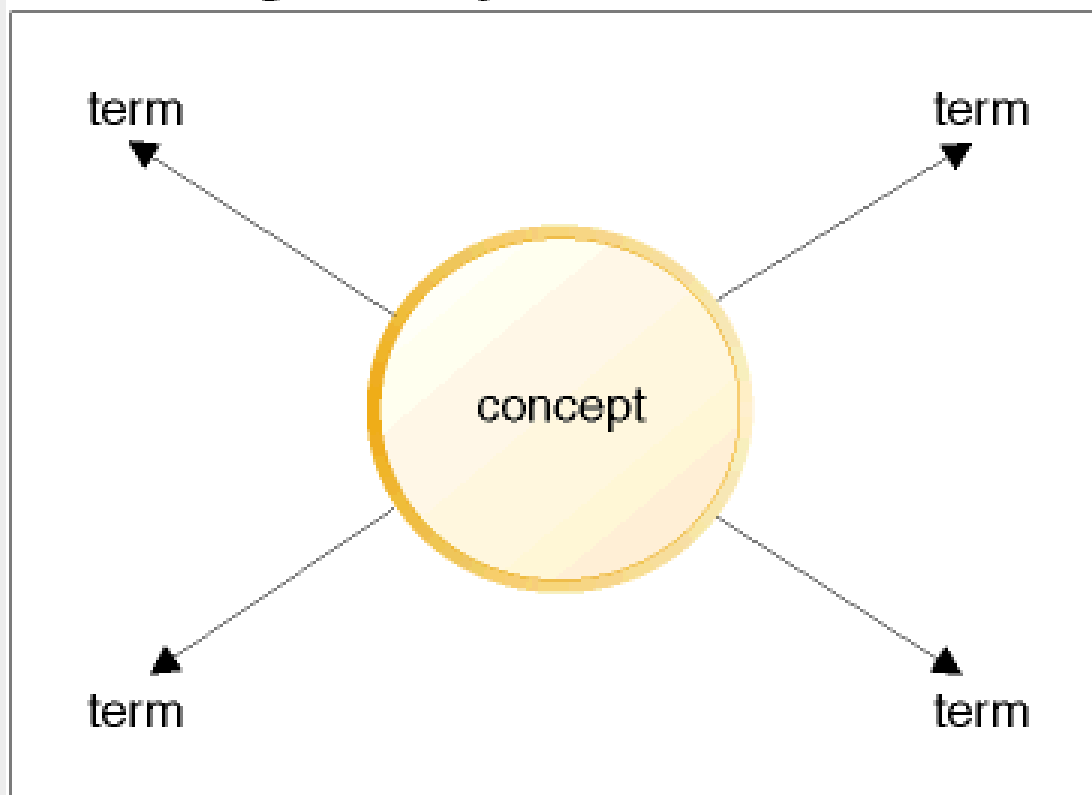
A terminological entry describes only one concept, and includes, *for that one concept only*:

- All terms in the source language (synonyms, acronyms, abbreviations, spelling variants, etc.)
- All terms in the target language (translations)
- Descriptive information (definitions, explanations, usage notes, etc.)

Often includes:

- Relations to other concepts
- Pictures
- Semantic data categories such as subject field

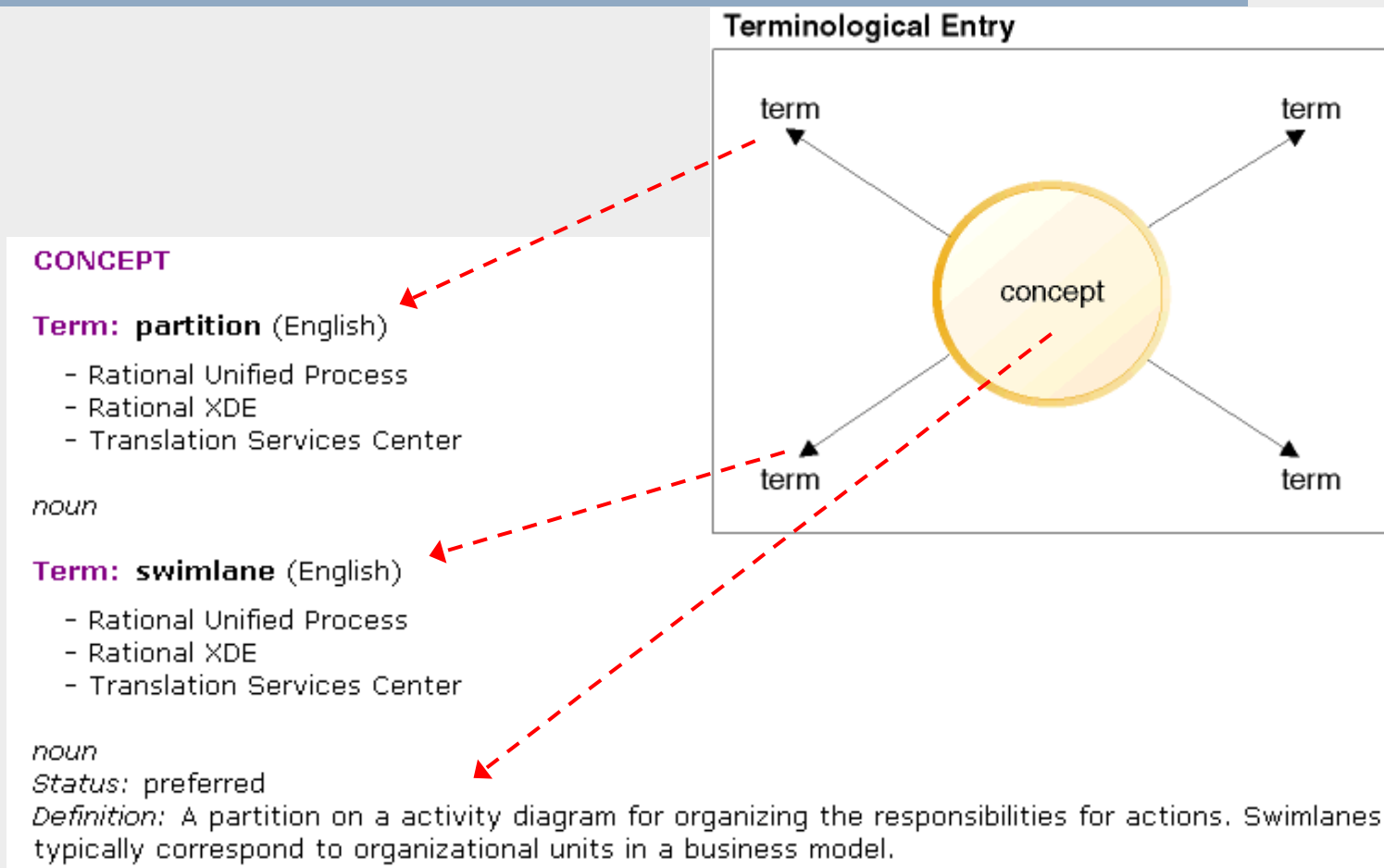
Terminological Entry



Concept orientation

English	French	Spanish
1. Subject Field(s) <ul style="list-style-type: none"> - Petroleum Deposits - Petroleum Asphalts - Crude Oil and Petroleum Products (General) 	Domaine(s) <ul style="list-style-type: none"> - Gisements pétroliers - Bitumes - Pétroles bruts et dérivés (Généralités) 	Campo(s) temático(s) <ul style="list-style-type: none"> - Yacimientos petrolíferos - Asfalto de petróleo - Petróleo bruto y derivados (Generalidades)
bituminous sands Source CORRECT, SEE OBS, PLUR, STANDARDIZED	sables bitumineux Source CORRECT, SEE OBS, MASC, PLUR, STANDARDIZED	arenas bituminosas Source CORRECT, FEM, PLUR
tar sands Source CORRECT, SEE OBS, PLUR, STANDARDIZED	sables bitumeux Source CORRECT, SEE OBS, PLUR	
oil sands Source CORRECT, SEE OBS, PLUR, CANADA, STANDARDIZED	sables pétroliers Source CORRECT, SEE OBS, MASC, PLUR	
asphaltic sands Source CORRECT, SEE OBS, PLUR, STANDARDIZED	sables asphaltiques Source CORRECT, SEE OBS, MASC, PLUR	
bituminous oil sands Source CORRECT, SEE OBS, PLUR		
petroleum tar sands Source CORRECT, SEE OBS, PLUR		

Terminological entry



Headword orientation

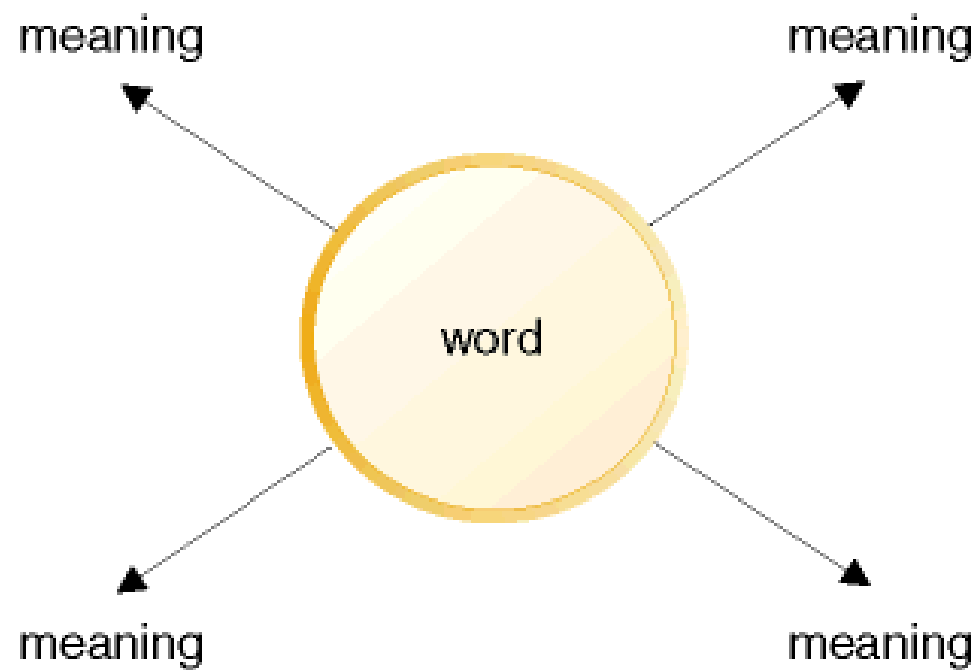
A lexicological entry describes only one lexical unit, and includes:

- All the meanings of that lexical unit

Because of the focus on the lexical unit, often includes:

- Etymology
- Pronunciation
- Collocations

Lexicological Entry



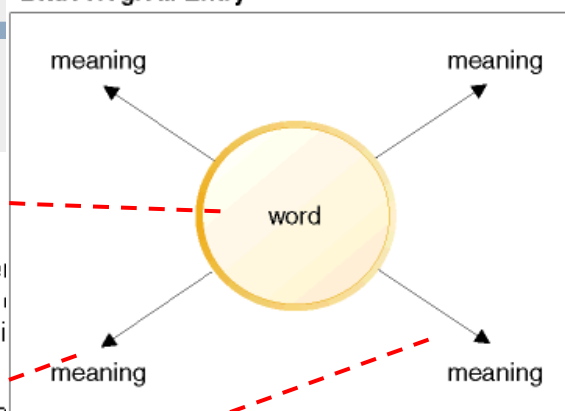
Lexicological entry

partition

noun

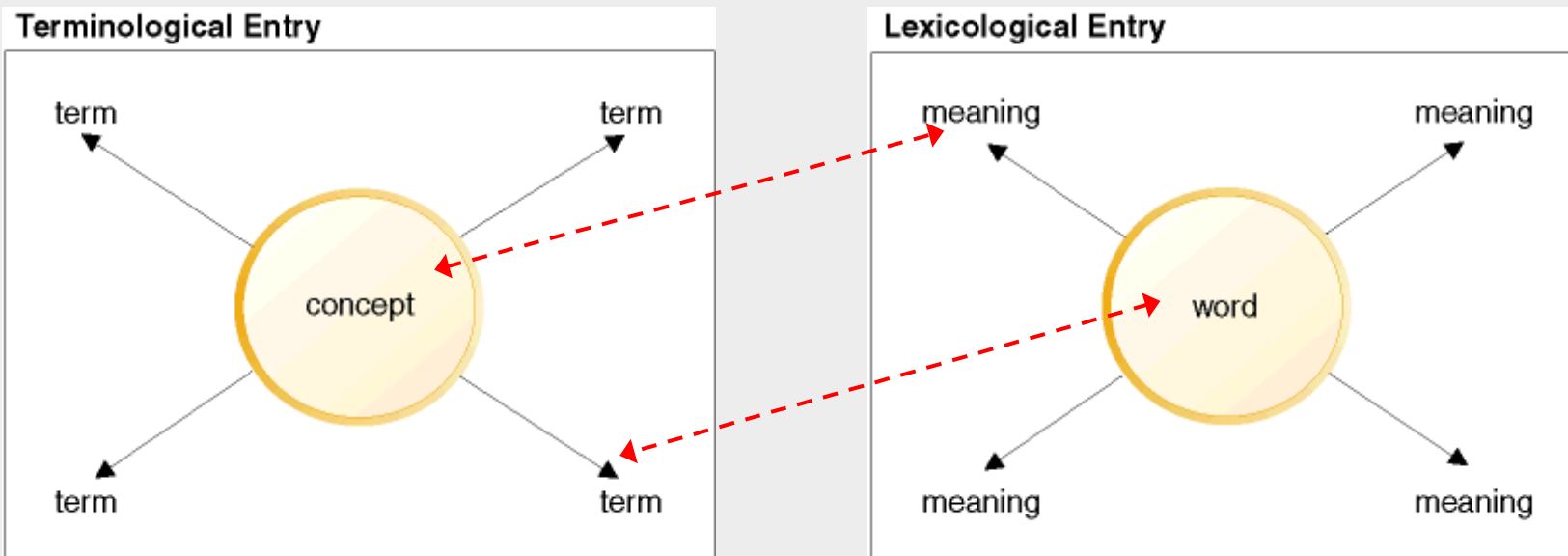
1. Definition: A portion of a page set. Each partition corresponds to a single, independent partitioned page set. All partitions of a given page set have the same maximum size.
Source: *DB2 Glossary, DB2 UDB for z/OS, Translation Services Center*
2. Definition: On a personal computer hard disk, one of four possible storage areas of a hard disk that can be accessed by DOS and each of the others may be assigned to another operating system.
Source: *Dictionary of Printing, iSeries, Translation Services Center*
3. Definition: A subset of the active cluster nodes that result from a network failure. Members of a partition maintain connectivity with each other.
Source: *iSeries, Translation Services Center*
4. Definition: A logical division of storage on a fixed disk.
Source: *AIX Glossary, Dictionary of Printing, iSeries, Translation Services Center, Tivoli Product Terminology, System p5, eServer p5 and i5 and OpenPower servers, x430 and NUMA-Q Glossary*
Comment: Partitions make it easier to organize information. Each partition can be formatted for a different file system. A partition must be completely contained on one physical disk, and the partition table in the Master Boot Record for a physical disk can contain up to four entries.
5. Definition: In VSE, a division of the virtual address area that is available for program execution.
Source: *CICS Glossary, Dictionary of Printing, Translation Services Center*
Comment: CICS/VSE runs in a VSE/ESA region, usually referred to as the CICS region.
6. Subject Area: Hardware \ Printers
Definition: In FD:OCA, a conceptual subdivision of a string of data fields. A partition can be further divided into subpartitions.
Source: *Dictionary of Printing, Translation Services Center*
Related Terms: [partitioning](#)

Lexicological Entry



International
Organization for
Standardization

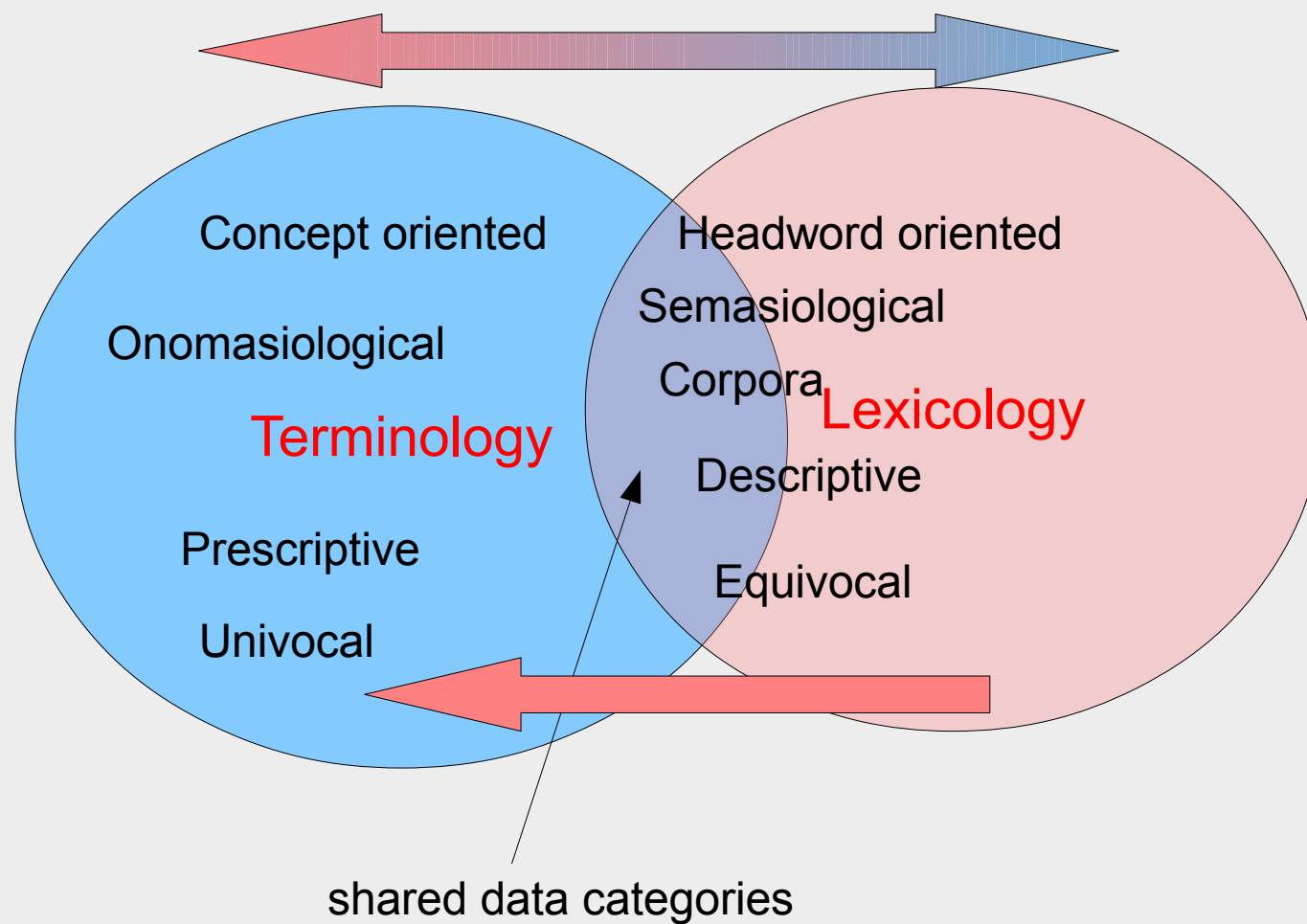
Both lexicological and terminological structures are needed for different purposes



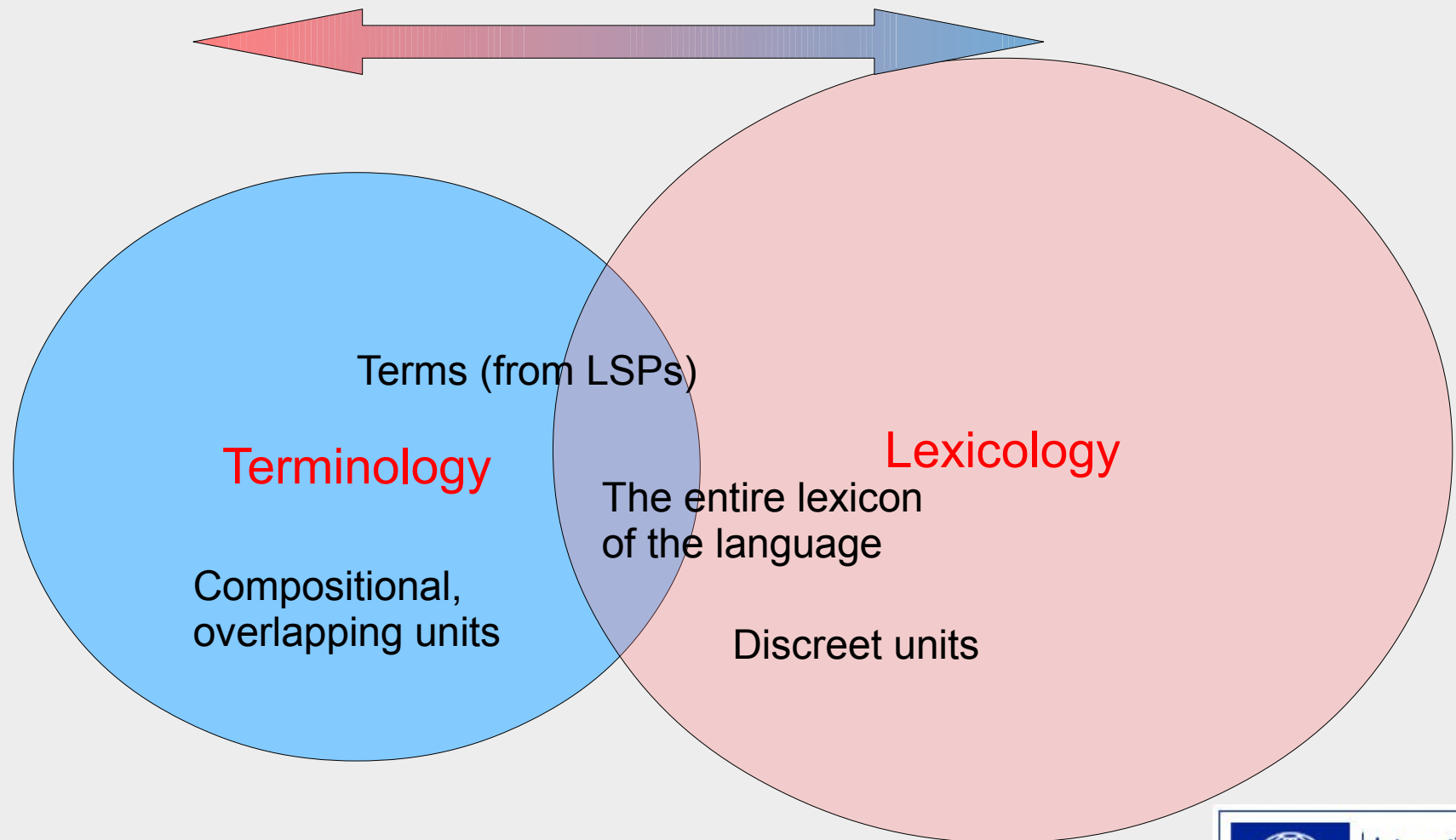
Principles and methods of Terminology, extended to lexical units of interest for various purposes

- Computer-assisted translation
- Controlled authoring
- NLP
- Sentiment analysis and opinion mining
- Search optimization

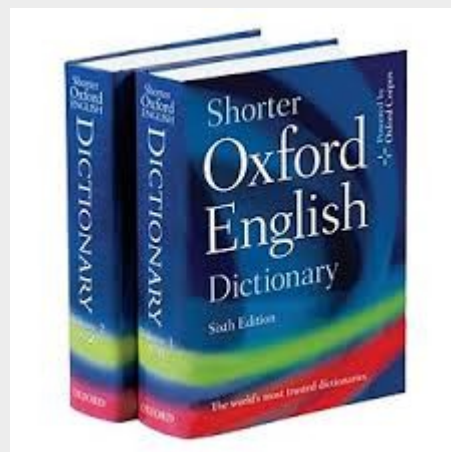
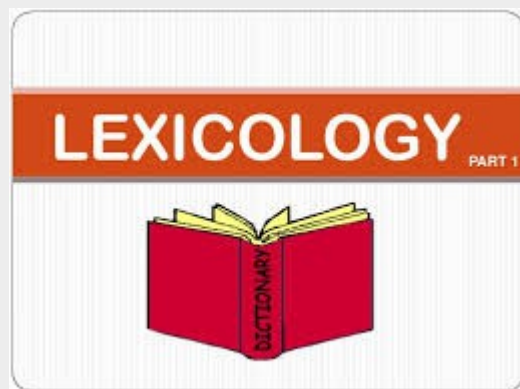
Theories and methods



Lexical items of interest



Lexicology – oriented to the publishing medium



Set (set) *verb* (sets, setting, past and past participle) put, lay, or stand (something) in a specified place or position. Delaney set the mug of tea down. Catherine set a chair by the bed. (be set) be situated or fixed in a specified place. the village was set among olive groves on a hill. represent (a story, play, film, or scene) as happening at a specified time or in a specified place. a private-eye novel set in Berlin. mount a precious stone in (something, typically a piece of jewellery): a bracelet set with emeralds. mount (a precious stone) in something. Printing arrange (type) as required. Printing arrange the type for (a piece of text): article headings will be set in Times fourteen point. prepare (a table) for a meal by placing cutlery, crockery, etc. on it in their proper places. (set something to) provide (music) so that a written work can be produced in a musical form: a form of poetry which can be set to music. Bell-ringing move (a bell) so that it rests in an inverted position ready for ringing. cause (a hen) to sit on eggs. put (a seed) in the ground to grow. Sailing put (a sail) in the wind.

Terminology – a granularly structured language *resource*



Concept ID 1517

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 (CO₂), methane (CH₄), nitrous oxide (N₂O), ozone (O₃), and water vapor (H₂O).

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

ISO/TC 37 Terminology and other language and content resources



Approaches to terminology management

- Originally
 - Systematic – work on a set of terms that are semantically related
- Increasingly
 - Ad-hoc – one term/entry at a time, usually to respond to an immediate need
 - Corpus-based

Systematic terminology work

A number of terms are worked on as a collection. Within this set, there are “families” of terms that are related:

- Morphologically, i.e. have similar parts:
 - forestation, deforestation, afforestation, reforestation
 - server, Web server, application server, domain name server, email server, file server, SQL server...
- Semantically, i.e. have a related meaning
 - forest, forestry, forest ranger, canopy, tree, lumber, biodiversity, forest fire, woodlot
 - server, client, replication, server farm, network, server pool

Ad-hoc terminology work

- Very common
- Pick up a term for a project, client, etc.
- Clarify it, add descriptors
- Translate
- Get feedback

Approaches to terminology management

Originally

- normative - “standardizes” terminology

Increasingly:

- descriptive – describes terms, their meaning, their usage, without making recommendations
- prescriptive - makes recommendations, chooses “preferred” terms. Requires a review/approval process.
- hybrid:
 - Some entries contain recommendations/decisions
 - A few entries might be “standardized”
 - Most entries are purely descriptive

Terminology standards vs standardized terminology

Terminology standards

- Standard ways to manage terminology data (ISO TC37 standards)
- Example: ISO 16642 – Terminology Markup Framework

Standardized terminology

- Terms that are “approved” over other terms
- Quite rare
- Example: Clause 3 of ISO standards, “Terms and definitions”
 - E.g. terminology of offshore drilling rigs

INTERNATIONAL STANDARD

**ISO
10256**

Second edition
2003-09-15

Head and face protection for use in ice hockey

*Protections de tête et de visage destinées à être utilisées en hockey sur
glace*



International
Organization for
Standardization

3.3 Face protection

3.3.1

chip

readily visible macroscopic particle missing from the protector

3.3.2

collimated light source (source of illumination)

quartz halogen lamp (17 lx or 1,68 foot candles) producing a 100 mm beam at 6 m distance which is centred on the pupils of the eyes of the headform or on the midpoint between the pupils of the eyes of the headform, this centring being maintained at all times during the test

3.3.3

combination

combined unit of a face protector or visor placed on a hockey helmet with which it is designed to be used

3.3.4

computer interface

linkage between the computer, the goniometer and the sensors, enabling a fully automated measurement process via a menu-driven operation

3.3.5

diopetre

measure of the power of a lens or a prism equal to the reciprocal of its focal length expressed in metres

3.3.6

face protector

device intended to reduce the risk of injury to the eyes and face of ice hockey participants

3.3.7

field of vision

projection outward of all retinal points (the nervous layer of the eye) at which visual sensations can be initiated



International
Organization for
Standardization

Approaches to terminology management

Originally: Onomasiological

- Identify and clarify a concept before naming it
- Develop a “concept system” with definitions based on logic
- Adherence to the “univocity principle”

Increasingly: Semasiological

- Identify a term then define it
- Recognition of polysemy and homonymy

To sum up: terminology “in practice” is changing

- Increased text focus and use of corpora
- Recognition of semasiological approach
- Need for ad-hoc approach
- Rarely normative
- Rarely univocal
- Increasingly pragmatic: role of cognition, communication, application
- Expansion of addressed lexicon
- Increased applications for NLP and commercial settings
- Adoption of methods from lexicography where suitable

But...ISO terminology standards have not kept up

- TC37 standards espouse the GTT and have not recognized any of the other theories and methodologies
- Many of the pragmatic approaches are dismissed as invalid
- There are no standards or best practices that address:
 - Use of corpora
 - Developing terminology resources for various applications

As a result, Terminology has an identity crisis

- Its not lexicology
- Its not the same “terminology” as before
- Yet it has great potential as a language resource
- We need visionaries to help develop this field



Thank You!