

# Terminology 101

## ILI Symposium

### Native Language Terminology Development

Kara Warburton  
October, 2010

Who am I?  
Who am I not?  
What this presentation will be about.

# What is a term ?

- One or more words that denote a concept in a “special” domain: technical, scientific, legal, political, cultural, etc:
  - Terms specific to native concepts: tribal council, treaty, First Nation, healing circle, trust responsibility, medicine wheel, Gathering of Nations
  - plants, animals, natural phenomena
  - traditions, customs, ceremonies, social titles
  - religious and spiritual concepts
  - principles: truth, respect, responsibility, reverence of elders
  - social issues: youth intervention, education, health, addiction

# Uses of structured terminology data

- Language reinforcement
- Support for communication
- Terminology checker (assisted authoring)
- Spell checker
- Support for educational materials
- CAT tools
- Soon (already?) terminology will be disseminated through mobile phones.

# What is terminology management ?

- Taking measures to ensure effective communication within the community, with a focus on:
  - identifying concepts
  - defining concepts
  - denoting concepts (with words, terms, expressions)
  - distributing the information, so that other people learn to use the words/terms.

# What are the tasks involved ?

- Collecting the terms in a domain
- Describing terms with metadata, such as definitions
- Creating terms when needed
- Determining target language equivalents
- Storing terminology data in a database, for re-purposing
- Distributing information about terms as a reference

# Goals of terminology management in *commercial* environments

- Goal: control language
- Prescriptive approach
  - One concept = one term
  - Minimize variations (synonymy)
  - Minimize sense overloading (polysemy)
  - The objective is consistency and standardization

# Goals of terminology management in *language planning* environments

- Goal: enrich language
- Approach is first descriptive, then (maybe) prescriptive
- Polysemy and synonymy have a purpose
  - Synonyms – language enrichment
  - Polysemes – a means of term creation (adding a new sense to an existing term)
- Prescriptive approach is utilized to “expel contaminants”, such as anglicisms

# Language is constantly changing

- Market changes: globalization, “buy local”, e-commerce, e-government, fair trade, trade deficits, etc.
- Shifting social focus: environment, social responsibility, aging population, global community, human rights, language preservation, social networking (Facebook, Twitter, etc.) cultural, linguistic, and religious diversity, etc.
- New innovations: agro-fuels, renewable energy, multi-media, wireless technologies, medical breakthroughs, etc.

# Terminology databases

- Tool for collecting, documenting, managing, and distributing terminology
- Increases productivity of translators and writers
- Valuable resource for community workers “in the field” to deliver programs in the native language
- Typically a prerequisite for language planning activities
- Example: Avataq Inuttitut Data Base

# Synonyms

## CONCEPT

**Term:** **partition** (English)

- Rational Unified Process
- Rational XDE
- Translation Services Center

*noun*

**Term:** **swimlane** (English)

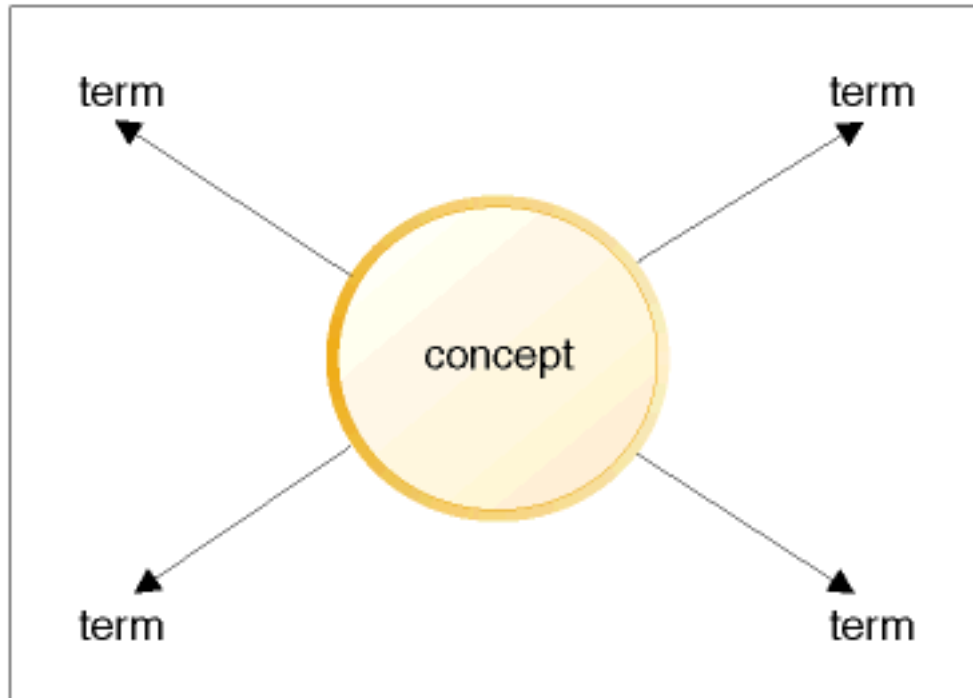
- Rational Unified Process
- Rational XDE
- Translation Services Center

*noun*

*Status:* preferred

*Definition:* A partition on a activity diagram for organizing the responsibilities for actions. Swimlanes typically correspond to organizational units in a business model.

## Terminological Entry



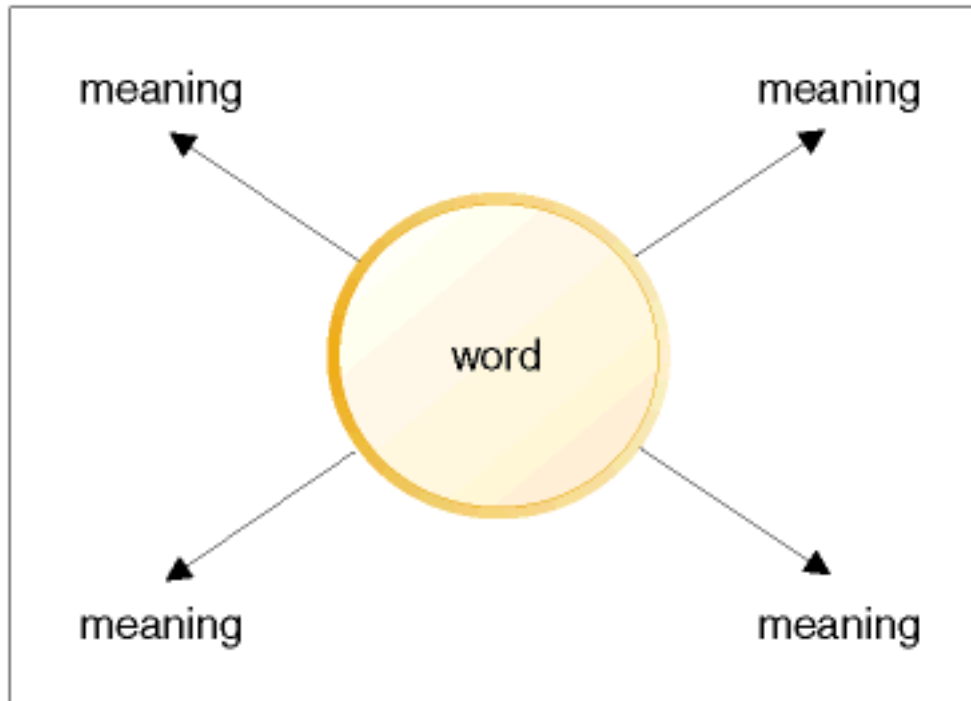
# Homonym

## partition

noun

1. Definition: A portion of a page set. Each partition corresponds to a single, independently extendable data set. Partitions can be extended to a maximum size of 1, 2, or 4 gigabytes, depending on the number of partitions in the 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 variable size; one may 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



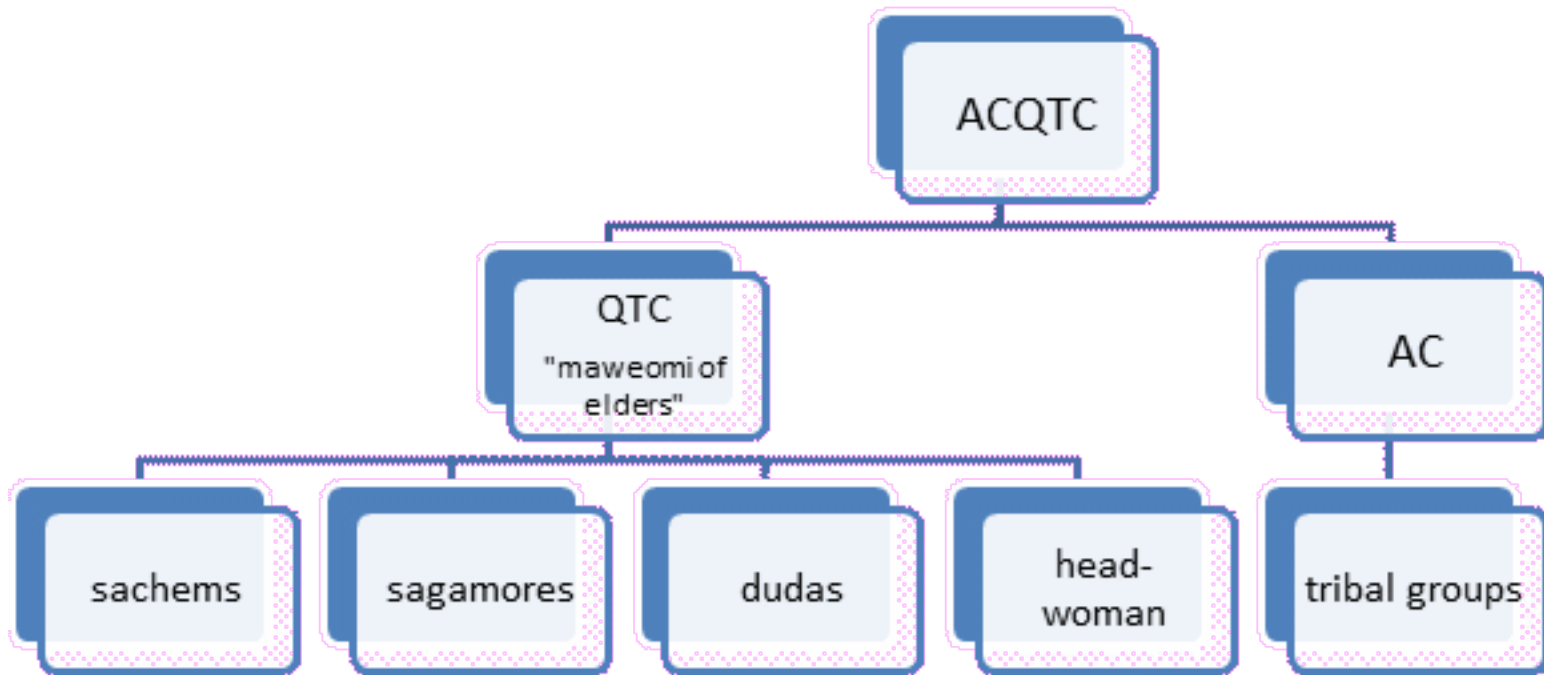
# Gathering terminology in a domain

- Domain: traditional dances
- Method: corpus analysis, interviews, existing dictionaries/glossaries
- Terms found: sun dance, thirst dance, rain dance, medicine dance, stomp dance, gourd dance..., sacred fire, firekeeper, flametender, medicine man, sweat lodge, piercing, drums, sacred pipe, fasting, tobacco, praying...

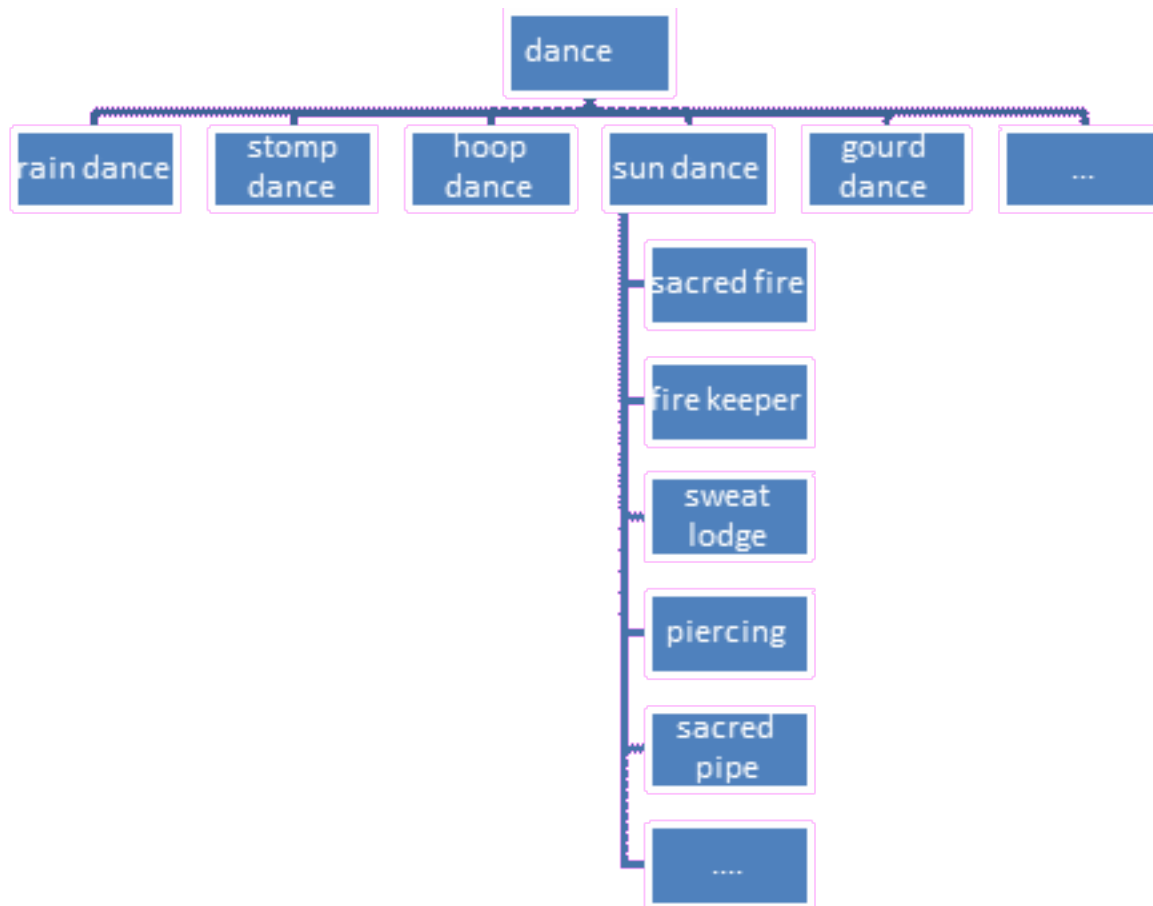
## Initial research indicates...

- *thirst dance* (Plains Cree), *rain dance* (Ojibwa), and *medicine dance* (Blackfoot) are synonyms of *sun dance*, used in different tribes
- However, *rain dance* may also have a different meaning, therefore, *rain dance* may be an example of both a synonym and a polyseme
- *firekeeper* and *flametender* are synonyms

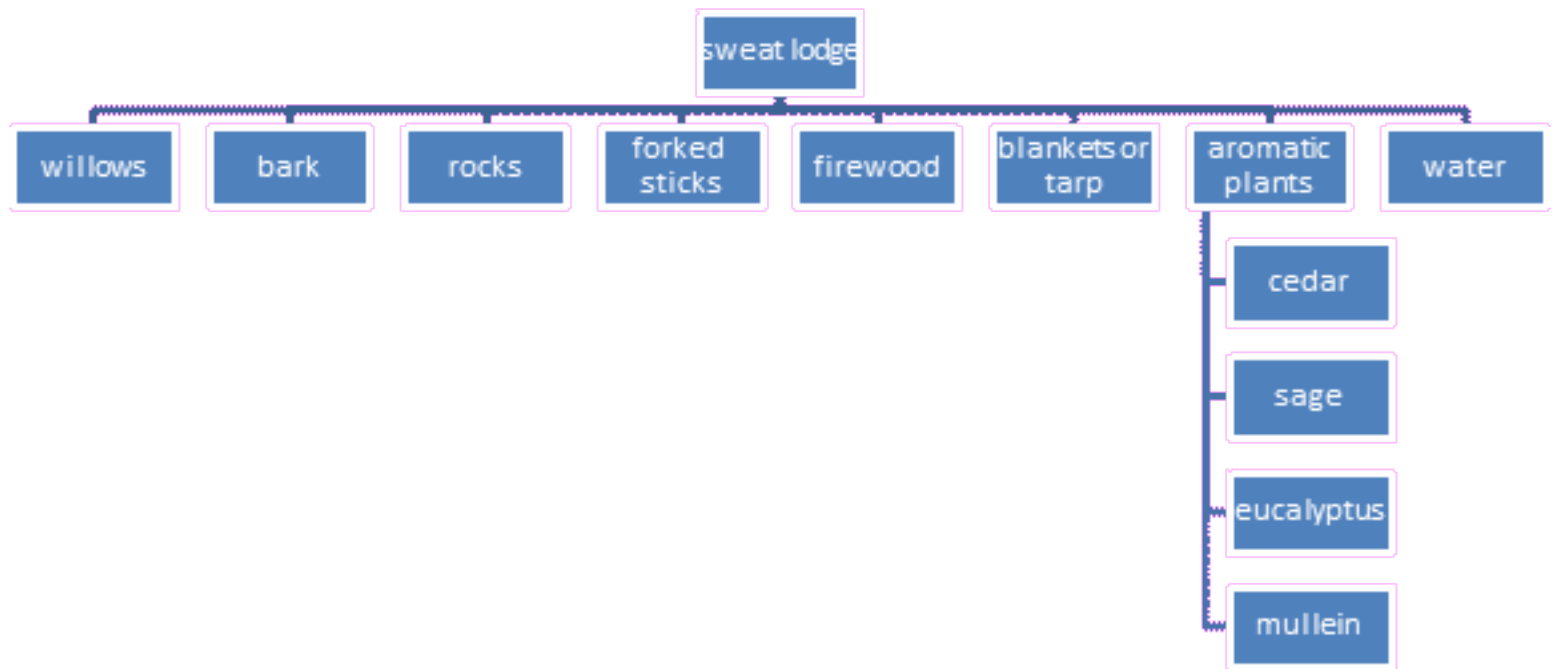
# Typical organizational chart – a model for concept organization



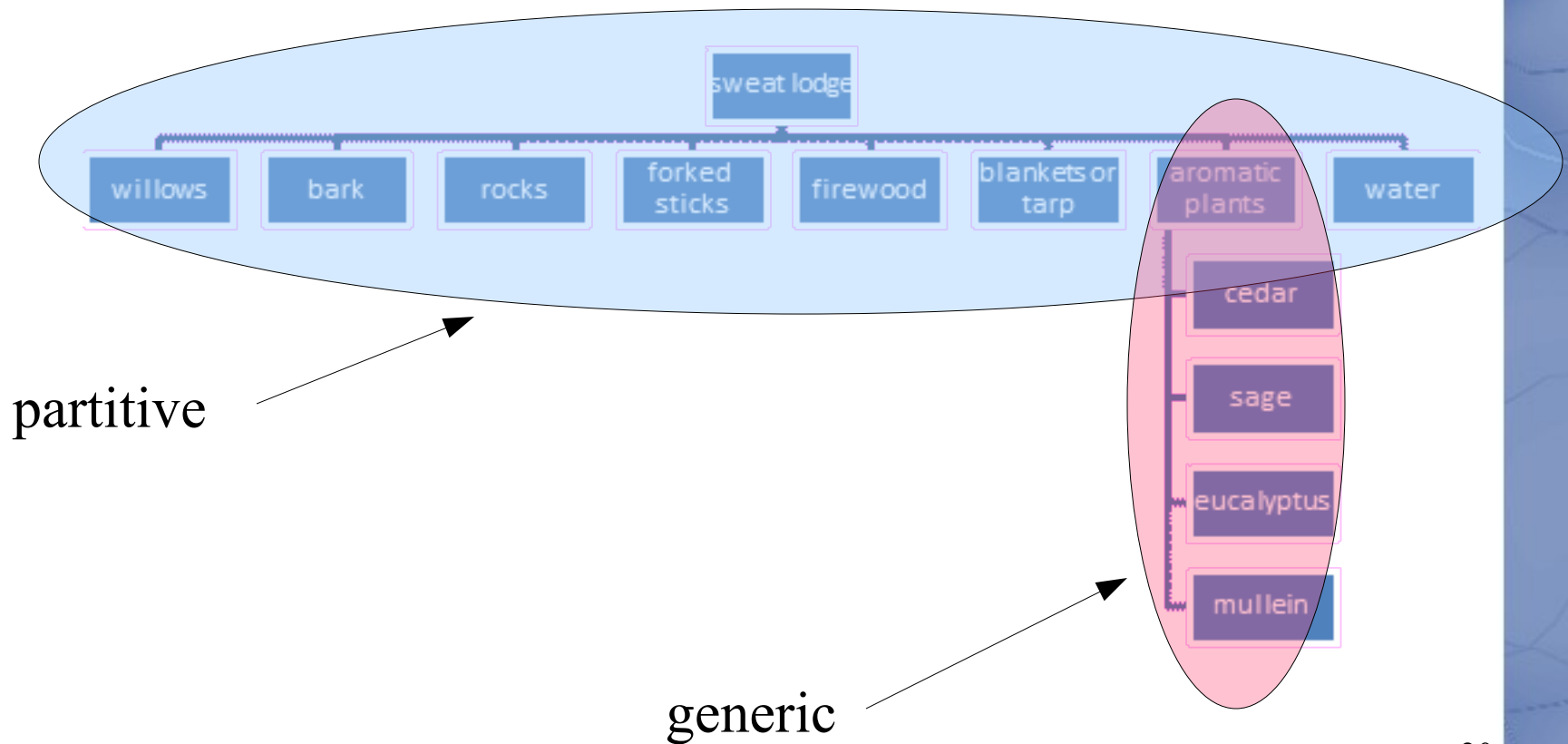
# Concept relations



# Concept relations



# Concept relations



# Generic definition

***term***

***definition***

[An] **electrical storm** [is] **a meteorological disturbance in which the air becomes highly charged with electricity.**

***genus***

**electrical storm**

a **meteorological disturbance** in which the **air becomes highly charged with electricity**

***differentia***

# Verify definitions by substitution

## wicket

set of three stumps and two bails at either end of the pitch

The **wicket** is broken if either or both of the bails fall from the stumps.

The **set of three stumps and two bails at either end of the pitch** is broken if either or both of the bails fall from the stumps.

# Contexts as a substitute for definitions

## diagnosis

we make a diagnosis when we identify certain symptoms as characteristic of specific conditions

- Sentence that sheds light on the meaning of the term
- Contains the term
- Can be automatically extracted from a corpus

# Creating new terms

- New terms are called a “neologisms”
- There are various methods:
  - coin a new term through morphological construction (preferred method)
  - semantic transfer – give an existing term another meaning
  - borrow a term from another language, possibly with modifications

# Example

- “let” is a diminutive affix in English, for example, piglet, booklet. Using this model:
  - doclet, portlet, applet, viewlet
- software – logiciel
  - “logi” is a latin-based affix meaning logical, based on computation.
  - “iel” is a French affix used for forming nouns, such as
  - From this model: pourriel, courriel, matériel, tutoriel, ludiciel, didacticiel, progiciel.

# The ideal term

- Well motivated - transparent (self-evident): torque wrench vs monkey wrench; square head vs Robinson head; recycling box vs blue box;
- No connotations (e.g. “cheat sheet”)
- Relevant to the target audience’s frame of reference: fire fighter vs pompier; fire truck vs pumper
- Respect rules of language: morphology, syntax, semantics

# Google and Microsoft

Google has launched Amharic, Oromo, Tigrinya (Ethiopia), Somali (Somalia), Kinyarwanda (Rwanda), Kirundi (Burundi), Lingala (Democratic Republic of Congo), Hausa, Yoruba (Nigeria), Zulu, Xhosa, Sesotho, Afrikaans (South Africa), Shona (Zimbabwe), Malagasy (Madagascar), and Akan (Ghana and Ivory Coast).

In its quest to reach more people, Microsoft is developing products in more than 60 African languages under the Local Language Program.

The Microsoft Language Portal uses a “crowd-sourcing” approach for collecting terminology.... you can participate directly in MS development of your terminology.

# Facebook

Also uses a crowd-sourcing approach for developing language-specific interfaces.

Could your youth be motivated to contribute?