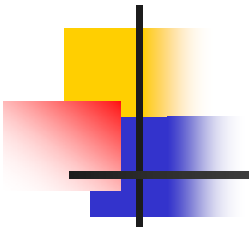
Excerpting terms in text



- Relates to term harvesting
- Departure phase – foundation for follow-up procedures
- Essential step for databases and technical dictionaries
- Compells the terminologist to make mental notes of what should be documented

What does term excerption mean

- Means that terms have to be recognised as specialised vocabulary where they cohere with the general language in technical communication
- Originates from the Latin word "***excerptum***", (something **picked out**); from "*excerpere*", (to **select** something); to **pick terms out** from where they are embedded in text

Excerption and information science

- Information science works with parts of text from text
- Terminologists concern themselves with even smaller entities of text (terms & phrases)
- Extract data to be added as contextual information
- Abstracts produced by information scientists are also valuable sources for term extraction.

Terminology and terminography

- The discipline of terminology (specialised field of lexicology) is particularly concerned with terms and their meaning
- terminography (specialised field of lexicography) is concerned with that which is captured in electronic term bases and the presentation thereof in technical dictionaries (hard-copy format)

Terms:
A few general characteristics

- A term is unique to a subject field, particular domain, science, art, profession, trade.
- and has a clearly defined meaning in a precise application of usage,
- can be a single word, word group or phrase, characterising a technical register
- terms are used to designate specific concepts

Value of terminology

- term bases and term lists, technical dictionaries play an important role:
- in transferring knowledge,
- serving as enabling tools for specialised communication (standardising technical communication),
- and also plays an overall important role of empowering people in terms of elaborating their linguistic capacity in general.

Simple term (simplex)

simple term (**simplex**): these are terms consisting of only one stem without affixes, eg

- **sediment** {*geology}*
- **mine** {*mining}* and
- **state** {*political sciences}*
- **twin** n. **{***crystallography}* [Afr tweeling]
- ***twin*** v. *{crystallography}* [Afr vertweeling]

Compounds

Terms that are formed by combining two or more free morphemes, eg

- **mountain chain** {*geomorphology*}
- **soil genesis** {*soil sciences*}
- **velvet bean** {*horticulture*}
- **solution collapse** {*spelaeology*}
- The first part (determinant) delimits or qualifies the meaning of the second part (determinatum).

Compounds (continued)

Some typical constructions:

- adjective plus verb, eg **nuclear spin**
- verb plus noun, eg **shear crack** {*tectonism}*
- verb plus preposition , eg **flash-over**
- noun plus noun, eg **shingle beach** {*sedimentology}*, **shield volcano** {*volcanology}*
- adjective plus noun, eg **diagonal fault** (*structural geology*), **scalloped upland** (*geomorphology*)

complex (multiword terms)

complex (multiword terms): terms consisting of a combination of compounding components, eg

- **phototypesetting**
- **siliceous cementing matrix** and
- **deep pressure sensibility**

Derivatives

derivatives: terms that are formed by adding one or more affixes to an existing stem or root, eg

- **breakage**, **orphanage**.
- In English the stem is usually a free morpheme, eg
- **walker**, **walking** (from walk)
- **conductive**, **conduction**, **conductivity**, **conductor** (from conduct)

Derivatives (continued)

The stems of many terms in technical languages are borrowed from Greek and Latin.

Many derivatives are formed from dependent stems, eg **dissip**- (does not have meaning on its own) in derivatives such

- as **dissipate**, **dissipation** and **dissipative**.

Collocations

- the constituent elements are **semantically fully transparent**
- are **syntactically predictable** in the sense that if one occurs, we expect the others to co-occur
- Abound in the technical language of certain fields such as finance, political science, and special domains such as meeting procedure, weather predictions, etc

Collocations (continued)

Examples

- **disposable profits**
- **scattered showers**
- **vote by ballot**
- **voters' roll**
- **convene a meeting / adjourn a meeting**

synonyms

When terminologists are collecting data for a specific terminological corpus, they also remain on the look out for possible synonyms.

If these don’t appear in the text itself synonyms could be collected from other existing sources, e.g. dictionaries

synonyms (continued)

A synonym is a :

- **member of a word pair or group of terms**
- and is **characterised by the fact that its central meaning overlaps** to a large degree with that of another member of the synonym pair/- group
- may be **fully exchangeable in some contexts and registers**, or they may belong to different registers (formal; informal; highly technical; slang, etc)

homonym

- Term with the same spelling and pronunciation as another but with a different meaning, eg
  - **grade** n. {*position or degree in a scale, rank,size such as grain size}*
  - **grade** n. {*incline*}
  - **market** n. **{***place where goods are bought or sold***}**
  - **market n. {***all potential buyers}*

Homonym (more examples)

- match n. {*organised game or competition*}
- match n. {*stick for lighting fires*}
- match v. {*to have the same qualities or look the same*}

and

- human n. {*a human being*}
- human adj. {*of or concerning man*}

abbreviations

- Shortened form of terms, symbolising the full forms, eg
  - **NGO** *for* nongovernmental organisation
  - **DTA** *for* differential thermal analysis
  - **Oe** *for* oersted **{***unit of magnetic-field intensity*}
  - **rms** *for* route mean square
  - **cm** *for* centimetre
  - **AAM** for Anti-Apartheid Movement

Abbreviations (continued)

Few more examples

- **DTP** *for* desk-top publishing
- **LRA** *for* Labour Relations Act
- **LSP** *for* Language for Special Purposes
- **TM** *for* translation memory
- **ID** *for* identification document

acronyms

Words (terms) created by the **combination of the initial letter(s) of each of the consecutive or most important parts of a complex term**.

Type of abbreviation that is **pronounced as a word**, and with time becomes so generally adopted by the speakers of the language that they do not always know what the abbreviation originally stood for.

Acronyms (A few examples)

- **JOIDES** *for* Joint Oceanographic Institutes for Deep Earth
- **quango** *for* quasi-autonomous government organisation
- **ABET** *for* adult basic education and training
- **Aids** *for* Acquired Immune Deficiency Syndrome
- **Eskom** *for* Electricity Supply Commission
- **NEPAD** for New Partnership for Africa's Development

Loanwords

Terms borrowed from other languages.

Loanwords become generally adopted in their foreign form in the borrowing language, eg

- **solfatara** term from **Italian** for a vent from which gases and vapours are emitted, characteristic of a late stage of volcanic activity

Loanwords (more examples)

- **tsunami** term from **Japanese** for a giant wave of water associated with earthquakes
- **kamikaze** term from **Japanese**, referring to suicidal military operations
- **imperium** term from **Latin** which refers to Roman political power and authority
- **impimpi** term from **isiZulu** which means "informer"

Questions for term collection

Terminologists and their collaborators work in response to the following questions:

- **who are the target users** and
- **what are the specific needs of the target user.**

The needs will influence decisions regarding the what is to be collected, the style guide, and the typology of the end product.

- **what sources should be used for term extraction**

Sources

- should be **sufficiently representative** of the scope of scientific and technical publications available in the subject field
- should be **current works** of reference
- terminologists determine the sources for excerption in **consultation** with **subject-field specialists**

Terminologist's continued task

- Fields of study are dynamic and **continue to develop**, which **generates new terms**
- Have to keep pace with such progress. It is therefore vital that terminologists read **recently** published **books**, **handbooks**, **manuals**, **journals**, **periodicals**, **news papers**, **lecturing material**, **research reports**, **abstracts**, documents pertaining to **legislation** and **draft legislation** (such as acts, bills, white papers, green papers and regulations), etc.

Continued search for new terms

- **Existing dictionaries** and **encyclopedias** are also indispensable sources of terminological information
- The **Internet** is an indispensable source for term harvesting

Aim of excerption

Not an action in isolation. It is done in relation to:

- **basic aims of terminography and lexicography** (creating containers of knowledge)
- with the **particular needs of the end user** in mind,
- **considering** throughout the **subject field** and its **subdomains**

Aim of excerption (continued)

- **short-term aims** are the **term collection** (excerption) and **demarcation** of the broader subject field, including the subfields**.**

- The **long-term aim** is the **documentation** of terminological data and the gradual **expansion of the terminological corpus** as well as the **publication** of **term lists and dictionaries**.

The process of excerption

- Orientation and demarcation
- Setting up conceptual frameworks
- Collecting terms for concepts
- Considering throughout important aspects regarding the users, language conditions, access & availability of similar/related sources

Orientation & demarcation

- **Reading** is essential - the more extensive the excerptor's knowledge of the relevant field of study, the better the quality of the excerption
- **texts that are representative of the particular field** (textbooks, periodicals, lectures, conference material, etc)

Orientation & demarcation (continued)

Reading helps to demarcate the knowledge base of a subject field or domain

- Terminologist becomes familiar with

  concepts, terms & term patterns, concept relations, terminological conventions in the field
- Human mind processes text in terms of the lexical items that will be excerpted as part of the technical vocabulary of the field.

Orientation & demarcation (continued)

- **Intrinsic** (colour, shape, size, composition) & **extrinsic** features (purpose, function, manufacture, origin)

Setting up conceptual frameworks

- The progressive discovery of terms renders the terminologist capable of setting up conceptual frameworks, and determining the relations between terms.
- Analysing the partly coinciding meanings of related terms, and considering the aspects in which related terms differ in meaning, terminologists are quite successful in distinguishing terms from ordinary words in text.

Aspects to be considered in term selection

- Target users
- Their specific terminology needs
- Types of lexical items for inclusion
- Main & peripheral fields
- Exact purpose for which terminological corpus is created, level of difficulty, register
- Literacy level of target users

Aspects to be considered in term selection (continued)

- Current multilingual language dispensation – not only the most prominent subject or domain-specific terms, but also lexical items interfacing or overlapping with the general language “grey-area terms”
- Availability or not, and access to dictionaries (general & special)

Aspects to be considered in term selection (continued)

- In various practical situations where critical needs exist for empowering people through their language proficiency, **particular factors and special circumstances** (as in the case mentioned above), **will impact on the scope of terminological data collection**

Aspects to be considered in term selection (continued)

- lexical items for inclusion is also influenced by the intrinsic nature of the language of the subject field
- As a rule, terms in texts dealing with topics in the natural sciences are more technical in nature, and are therefore more obviously identifiable than in texts dealing with topics in the humanities

Aspects to be considered in term selection (continued)

- Owing to the inherent nature of the subject content in the humanities terms are not always that obviously identifiable and many lexical items fall within the “grey areas” of the language where certain levels of interfacing and overlapping prevail

Contextual information

- Providing contextual information in a terminological data corpus is of great importance to **elucidate the meaning of terms**
- Context indicators are **words** or **phrases**, **indication** of **subject field**, **usage**, or **typical context** in which both source language terms and target language equivalents occur

Contextual information (examples)

- **crack array** {*structural geology*}
- **tie bar** {*eg a sand bar connecting an island with the mainland*}
- **ultimate landform** {*close to end of cycle of erosion*}

and

- **impeachment** {*charge against incumbent of a specified office*}
- **beehive industries** {*eg Small Business Development Corporation*}