Beyond facets: Semantic roots and modifiers as elements of a conceptual morphology

Dagobert Soergel
College of Library and Information Services
University of Maryland
College Park, MD 20742
301-405-2037

Abstract

This paper presents initial ideas on a conceptual morphology in which concepts such as Fermentation, Fermented, and Fermentable are represented as combinations of a semantic root, in the example *Ferment*, with a modifier, in the example **process**, **state/condition**, and **susceptible to process**, respectively. This makes it possible to generate a large number of concepts from a much smaller list of semantic roots and modifiers. It also allows for great flexibility in indexing and searching. The paper gives a preliminary scheme of modifiers and invites ideas from classification researchers, logicians, and linguists.

1. Introduction

Facet analysis identifies the semantic components of a concept and arranges the resulting elements into facets. For example

Hepatography = Liver & Imaging & X-rays

Hepatograph = Liver & Camera & X-rays

Hepatogram = Liver & Image & X-rays

The facets are **body part** (liver), **process** (imaging), **apparatus** (camera in the broad sense of any imaging device), **thing** (image, X-rays). This analysis results in great conceptual economy, since a given set of elemental concepts can be used to express a much larger number of combinations. Facet analysis also facilitates searching for broad concepts, such as *Camera (imaging device)* regardless of type, while still allowing for specific searches - e.g. for *X-ray camera* - by combination.

Three of the concepts used in the example, Imaging, Camera, and Image, while belonging to different facets, share the core meaning of *image*, which is broader than any of the three concepts. Some queries require this broad meaning, possibly combined with an imaging principle, such as X-rays. We can take facet analysis and conceptual economy a step further to support this type of broad search. We can express a concept such as Imaging by a combination of its semantic root - *Image* - with a general modifier, in the example **process**. Here are the results of such an analysis for the three sample concepts (semantic roots in italics):

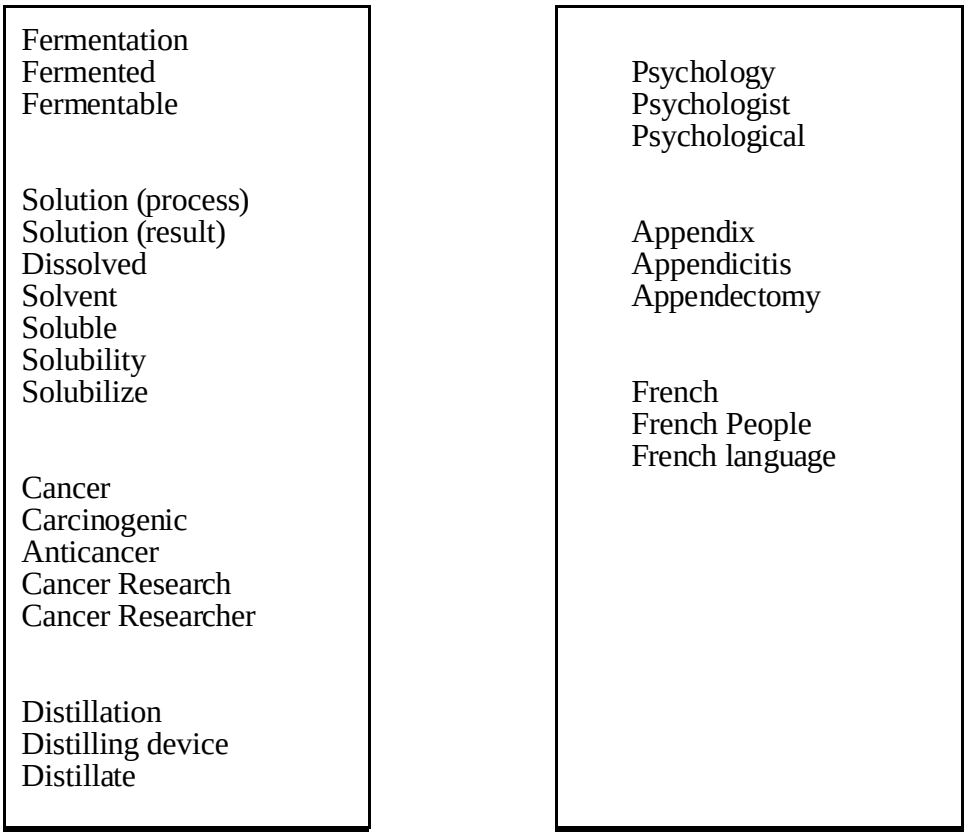
Imaging = *Image* - **process**

Camera = *Image* - **apparatus**

Image = *Image* - ***resulting thing***

Again, this analysis allows a broad search for the semantic root *Image* (combined with X-ray or ultrasound or whatever, as required), but also a specific search for Camera or, even more specifically, X-ray camera.

Figure 1 lists some other groups of concepts that share a semantic core.



a. Mostly clear-cut examples

b. More difficult examples

Figure 1. Groups of concepts that share a semantic root

The examples in Figure 1 suggest the modifiers already mentioned, **process**, **apparatus/device**, and **resulting thing**, as well as others, such as **susceptible to process** or **causing**. Other examples are modifiers that specify **comparison to a standard** (under, sub-; normal (default); over, super-) or **relative size** (micro-; normal size (default); macro-).

The principle of forming concepts by combining semantic roots with modifiers is, of course, borrowed from natural language, which uses roots and modifiers, such as **solu-ble**, to construct many words out of a much smaller number of elements. But natural language has many exceptions and irregularities, and some perfectly reasonable combinations (such as image-able) would sound quite unusual. I propose to apply this familiar principle in a kind of **conceptual morphology**. I envision a classification consisting of a well-structured skeleton of core concepts or semantic roots and a carefully worked-out system of modifiers. Many more concepts can be formed through combining these elements.

The task, then, is to develop a **logical system of modifiers that is expressive yet simple enough to be learned and used**. The paper provides the beginnings of such a system but is primarily a call for ideas from classification researchers, logicians, and linguists.

2. Developing a scheme of modifiers

One might assume that linguists would have worked out such a scheme in an analysis of modifiers used in natural languages. But consultation with a linguist produced no promising leads to work that would be applicable directly. So it seemed useful to make an independent start and than look at other work that is related and from which many useful suggestions can be drawn once the idea of a conceptual morphology has crystallized a little more. The task of developing a list of modifiers can be approached as a bottom-up thesaurus development task: Start with a sufficiently large number of concepts and perhaps arrange them into groups around core meanings. Then express each concept as a combination of semantic root and modifiers, making up new modifiers as needed. Often linguistic form will help in this process. Finally arrange the modifiers into a coherent schema and consolidate those that are very similar.

Analyzing a limited set of terms collected in conjunction with the Alcohol and Other Drugs Thesaurus resulted in the preliminary scheme presented in the appendix. The following examples illustrate the approach.

The first set of examples revolves around the semantic core *Ferment*.

Fermentation = *Ferment* - **process**
Fermenting

Fermented = *Ferment* - **state/condition**

Fermentable = *Ferment* - **susceptible to process - state/condition**

Thinking about fermentation, one can easily come up with other combination for which there is no word in English:

Ferment - **apparatus/device**

Ferment - **agent**

Ferment - **susceptible to process - process**
(the process of making something fermentable)

The next set of examples revolves around the semantic root *solution*. Actually, this is a polysemous term or a homonym, depending on what commonality one perceives between problem solution and chemical solution. Most combinations are valid for either meaning, but their may be two English words or the English word may refer only to the problem-solving meaning or only to the chemical meaning.

Solving = *Solution* - **process**
Dissolving

Solution = Solution - **resulting thing**

Solved = *Solution* - **state/condition**
Dissolved

Solvent = *Solution* - **agent**

Solvable Soluble	=	*Solution* - **susceptible to - state/condition**
Solvability Solubility	=	*Solution* - **susceptible to - state scale**
(solvabilize) solubilize	=	*Solution* - **susceptible to - process**

Some of these examples illustrate the need for having two modifiers in a row. There may even be longer modifier chains. The concept to which the last modifier is applied is called the **base concept**. A base concept may be a semantic root or a combination involving one or more modifiers. It may even be necessary to admit base concepts which are constructed by combining several semantic roots with or without modifiers.

The modifiers are used to define concepts, not to make statements about reality. For example, *Cancer* - **agent causing** defines Carcinogen, which is a valid concept whether or not substances or other agents causing cancer actually exits. The statement

Tar isa *Cancer* - **agent causing**,

on the other hand, uses the concept in making an assertion about reality.

The appendix gives a first sketch of modifiers to be used in conceptual morphology.

3. Applications in retrieval systems and language processing

As was mentioned in the introduction, the motivation for the development of conceptual morphology is two-fold: to allow for great flexibility in indexing and searching and to limit the number of elements needed in the index language. The first is achieved by being able to search broadly for semantic roots or specifically for semantic root - modifier combinations. The second can be achieved by limiting the index language to the semantic roots and modifiers as the elemental building blocks from which the indexers and searchers can construct indexing concepts.

In some systems the specificity made possible by modifiers may not be needed. In that case, the index language can be limited to semantic roots. If some specific combinations are needed, they can be introduced as precombined descriptors.

An index language restricted to semantic roots and modifiers may put a burden on the indexers and searchers, who must understand the system and construct the appropriate combinations. This problem can be addressed by including many lead-in terms in the thesaurus or by introducing precombined descriptors. Both approaches are limited by what can be expressed in natural language. One might argue that a concept that cannot be expressed by a word or established phrase in the language is not important, but that is debatable.

Precombined descriptors not only facilitate indexing and searching but they also serve as examples on which indexers and searchers can pattern other combinations. On the other hand, precombined descriptors introduce a degree of inconsistency in that some combinations are already there whereas

others need to be created. Internally, precombined descriptors should always be resolved into their components for two reasons: (1) Entities indexed by a semantic root - modifier combination should always be retrievable by the semantic root alone, whether the combination is in the index language as a precombined descriptor or constructed by the indexer. (2) Some indexers and searchers may find it easier to construct a combination than to use a precombined descriptor. This should make no difference to the system. This internal resolution into components also assures that all precombined descriptors are formed in the same systematic way, and that semantic relationships are thus made explicit.

The specificity and precision achievable by the approach described here should be particularly useful for the representation of "substantive" data (as opposed to bibliographic or other directional data). It may also be useful in dictionaries for natural language understanding.

4. Outlook

The thinking about this system is still in an embryonic stage. The patterns of combinations and the modifiers listed in the appendix are those that are most obviously suggest themselves from an analysis of a limited list of terms. The less obvious cases must be dealt with, and the somewhat ad hoc listing given below must be shaped into a coherent system that balances the need for expressiveness against the need for simplicity. As was stated in the introduction, this paper is primarily a call to classification researchers, logicians, and linguists for ideas and references to related work which would serve as a basis for developing the schema further and apply it in some pilot systems.

Appendix

First sketch of modifiers to be used in conceptual morphology

In the following table, the left column gives a type of base concept and the right column gives modifiers that might be applied to that base concept. As noted in the text, the base concept can be a semantic root or a semantic root plus modifier. Usually it is clear to what base concept type a combination with a modifier belongs. For example,

Acid - **addition**

can become the process Acidification by further combining with the modifier **process** or the state/condition Acidified by further combining with the modifier **state/condition**.

Base concept	**Modifier**
Any concept	pertaining to that concept (adjective derived from the concept)
Any concept or topic	realm/discipline pertaining to the concept or topic (e.g., Heart - Cardiology)
Any concept, particularly Process State/condition Being/thing/object Substance	agent causing or furthering (-genic) agent causing agent furthering agent killing or inhibiting (anti-) agent killing agent inhibiting Note: Special modifiers could be created to define any of the following agent types: Process Thing/object Apparatus, device, machine Substance Being Organism Person One might also want to look at the distinction between animate and inanimate agent and the related distinction between agent in a more narrow sense and instrument. susceptible to functionality good function (default) dysfunction (mal-, mis-) Note: Dysfunction could be subdivided into many specific modifiers for the type of dysfunction, particularly type of disease, but these then loose the intuitive character of a modifier.
State/condition Being/thing/object	process of development (-genesis, -poiesis) (Combinations can be further modified, see process/ state/condition) process of reduction or dying reduction dying Note: Killing would be the modifier Agent killing with process as the agent type.

Base concept	**Modifier**
Process State/condition	process (If the base concept suggests a condition, this is the process that leads to the condition.) being/thing or state/condition (often the result of a process). being/thing state/condition state/condition as such (e.g., freedom) state scale state/condition as property of some being/ thing (e.g., free) being/thing having that property degree of severity (related to Status of affirmation) not at all semi- fully
Being/thing/object seen in relation to a larger system Substance	addition removal, extraction replacement dysfunction/disease due to lack dysfunction/disease due to overexposure
Any quantity	measurement of that quantity (may be further modified as Process/state/condition)
Realm, area of endeavor, field, discipline, occupation	area/field/discipline/occupation as such, as a social organization, role of the area/field/discipline discipline profession, occupation person working in the area/field/ discipline/occupation pertaining to the area/field/ discipline (special case of pertaining to)

Base concept	**Modifier**
Space/time/figurative dimension (Concepts that can be used in two or three ways, such as before or under)	space use time use figurative use
Geographical area/ethnic group/language (There could be one list of semantic roots; each root may designate one or more of these, but see note.)	geographical area (e.g. France) ethnic group (e.g. French people) language (e.g., French language) Note: This one may be stretching the point. While the economy of having only one list replacing three, the relationship between France, French people, and French language is more empirical than semantic. The hierarchy of languages in particular follows geography only loosely. So perhaps it would be better to maintain three different lists but add a fourth which would have concepts such as France, French people, French language to facilitate searching for all these aspects at once.

Base concept	**Modifier**
Any concept Very general modifiers	status of affirmation negated (non-, un-, in-) affirmed (default) presence vs. absence presence absence, without, -less comparison to standard under, sub- normal (default) over, super- relative size micro- normal size (default) macro- number of elements one, mono-, uni- two, dual, dyadic, bi- three, triadic, ternary, tri- four, quaternary, tetra- five, penta- six, hexa- etc. many, multiple, multi-, poly- temporal aspects before, pre- during (default), present after, post-