Understanding Systematic Conceptual Structures in Polysemous Medical Terms

Aldo Gangemi, Domenico M. Pisanelli, Geri Steve
Istituto di Tecnologie Biomediche - CNR Roma, Italy

Polysemy is a bottleneck for the demanding needs of semantic data management. We suggest the importance of a well-founded conceptual analysis for understanding some systematic structures underlying polysemy in the medical lexicon. We present some cases studies, which exploit the methods (ontological integration and general theories) and tools (description logics and ontology libraries) of the ONIONS methodology defined elsewhere by the authors. This paper addresses an aspect of the project we are involved in, which investigates the feasibility of building a large-scale ontology library of medicine that integrates the most important medical terminology banks.

INTRODUCTION

Polysemy – the phenomenon by which a term may have multiple meanings – greatly affects data management. In fact the current demand is for an unambiguous sharing or integration of the semantic content of data, either between different databases or for intelligent information access, natural language processing, terminological standards definition, etc.

In the face of such demand, domains and applications often deal with a lack of conceptual foundation. For example, within the domain of molecular biology, Schulze-Kremer[1] reports an interesting case of the relevance of semantic mismatches. Even an - apparently - unambiguous term like *gene* may be found conceptualized in different ways in different genome data banks. According to one (GDB), *gene* is a DNA fragment that can be transcribed and translated into a protein, whereas for others (Genbank and GSDB), it is a "DNA region of biological interest with a name and that carries a genetic trait or phenotype".

Since many years, our research has focused on the use of ontology libraries that are formally defined according to rigorous general theories, in order to analyze, integrate, and formalize medical terminologies. The resulting libraries are one contribution to the set of tools necessary to cope with semantic mismatches in biomedicine[2].

In this paper, we present some case studies of conceptual analysis applied to the formal modelling of polysemous terms, and we suggest that the definition of solid, formal conceptual structures is the only way to cope with the pervasive presence of polysemy in the lexicon, even in scientific domains like medicine.

The case studies presented here have been investigated within a project that exploits the ONIONS methodology: for a comprehensive presentation of our methodology for conceptual analysis and ontology integration, see [2].

In the first section, we introduce some basic notions. In the second we present some case studies. In this paper we assume some basic knowledge of the notation of description logics and their semantics (subsumption, relation composition, etc., cf [3]).

CONTRASTIVE VS. LOGICAL POLYSEMY

Polysemy is widespread in verbal communication. The main reason is that linguistic items are used 'economically': humans try to use already known terms, instead of creating complex sentences for each intended meaning thay want to express. For example, by *window* one could mean either a glass-filled frame, or an opening in a wall. Terminological economy is diffused in specialized languages as well and medicine makes no exception. For example, by *inflammation* one could mean a physiological function, a condition, or the area of an organ that bears an inflammation process.

Many standard vocabularies try to organize medical terminology in order to enumerate the main meaning shifts of terms. In particular, the big effort devolved in the construction of the merging repository called UMLS[4], succeeded in controlling the more superficial kind of ambiguities, due to historical relatedness, accidents of orthographic blending, or syntactic alternation (eg, noun vs. verb). In linguistics, this kind of polysemy is named *contrastive polysemy*[5].

On the other hand, there are more subtle cases of polysemy in the lexicon, like the ones mentioned in the *window* and *inflammation* examples. In linguistics, such cases are named *logical polysemy* and their main feature is that the multiple meanings of a polysemous term are interrelated. For example, a window as glass-filled frame is located at a window as opening, and it carries out the functions of controlling air exchange between interior and exterior, letting light reach the interior, etc. An inflammation process is embodied in an area of an organ, and both are contextualized in a condition.

Actually, logical polysemy creates a network of multiple related senses for a given word. What is interesting for ontological engineering is that multiple senses are

systematically related, in that the *categories* of the related senses form a restricted set.
For example, the categories of the two related senses of *window* are *object* and *place*, while two categories for *inflammation* are *process* and *outcome*. The logical polysemy that creates a category alternation between object and place is frequent, and acts for example in the intended meaning of *body region*, that may mean either an anatomical space, or the anatomical structures located at that space (see examples below).
In linguistics, category alternation in logical polysemy is named *metonymy*. Recent developments in cognitive linguistics and lexical semantics[5, 6] show that metonymy is a general feature of conceptual structures used to communicate and to organize knowledge, and it goes beyond the polysemy of terms.
In this paper, we present some conceptual analyses of metonymies in the medical lexicon.

FORMAL ANALYSIS OF METONYMIES

A dictionary like Webster's is quite direct defining *metonymy*: "the use of the name of one thing for that of another of which it is an attribute or with which it is associated". Although intuitive, the definition does not make justice of the conceptual complexity of metonymies.
Our formal definition of metonymy is: the *systematic* activation of two or more *related* concepts within the *same intended model* by using one name (proper metonymy), or two or more near-synonyms (metonymical network).
An intended model is a model that approximates the conceptualization (the intended meaning) of a vocabulary (cf. [7]). "Systematic activation" means that there should exist a conceptual structure (usually quite general) that allows us to use or interpret a name metonymically.

As a matter of fact, to understand the relatedness of any pair of metonymically related concepts, one must resort to some abstraction. For example, in the following pair of sentences (Fig. 1 shows a graph of the related model):

A. the humerus is in the arm
B. the arm is broken,

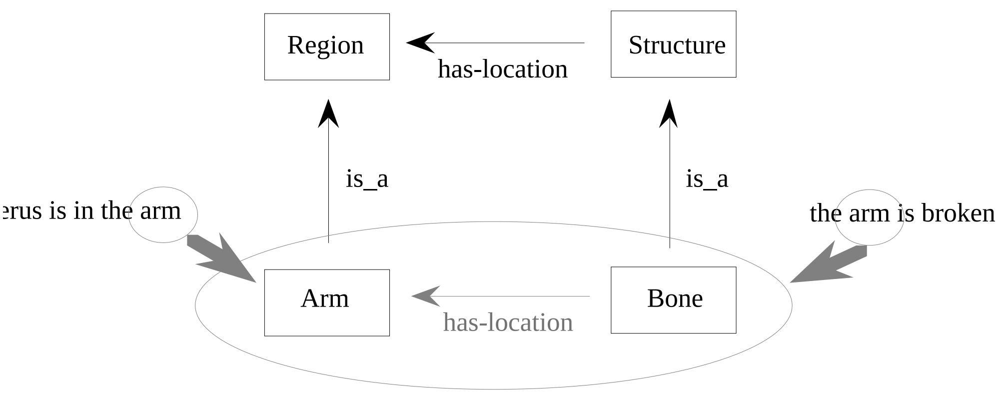

Figure 1. The polysemy of *arm* arising in two sentences

arm activates two different yet related parts of a model that contains at least the concept of *arm* as a place, the concept of *bone* as an anatomical structure, and the concept of *location*. Arm in (1) activates the concept of arm as a place, while arm in (2) activates the concept of bone *as far as* it is located at the arm. This metonymy is an instance of the general case of metonymy allowing us to use the same term both for a region and for the objects located at that region when the region is more cognitively salient than the objects located at it[6].
Another case is shown in the following pair of sentences (Fig. 2 shows a graph of the related model):

C. the neoplasm is organoid
D. the neoplasm is worsening,

where *neoplasm* activates the concept of neoplam as an abnormal structure at (3), and the concept of neoplastic process (that necessarily produces an abnormal structure) at (4). Such metonymy is an instance of the general case of metonymy allowing us to use the same term both for a process and for the product of that process when the product is more cognitively salient than the process itself.

For example, the same metonymy creates a semantic mismatch between ICD10[8] cand Snomed-III[9] in the form of a metonymical network. In ICD10 the terms for *inflammation* are classified as "inflammatory diseases" (a kind of process), while Snomed-III has *inflammation* under a separate taxonomy (branching from "morphology") containing properties or structures *produced by* an inflammatory disease.

Conceptual analysis puts into evidence other issues concerning morphologies. The most important is the dependence between a morphological condition, a function, and the related organ.
For example, an "ulcer" (as a condition) of a stomach implies that the stomach *embodies* an *ulceration function* (an ulcer as a function).
Another example is the mereological import of morphologies: some are featured by an organ, some only by a part of an organ. For instance, an "ectopic heart" is wholly ectopic, but an "ulcerated stomach" is only partly ulcerated. This example shows also how pervasive is polysemy in the realistic use of the lexicon.

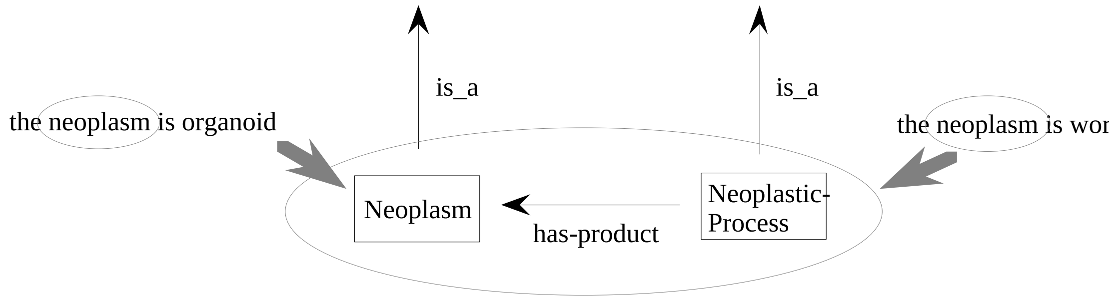

Figure 2. The polysemy of *neoplasm* arising in two sentences

METONYMY, RULES, AND AUTOMATIC CLASSIFICATION

In medical ontological engineering, it is sometimes mentioned the oddity deriving from the application of the sensible rule (assuming that "part" is transitive):

```
(implies (:composition embodied-in part)

         embodied-in)
```

From such rule, an "injury embodied in a part of an organ" is an "injury embodied in the organ". For example, "fracture of the phalanx of the thumb" would be "fracture of the arm", since a phalanx is a part of the thumb, which is a part of the arm.
Here the problem actually derives from the assumption of transitivity made on "part". If we use a non-transitive mereological relation, such as "component":

```
(implies (:composition embodied-in component)
         embodied-in),
```

the inference allowed by the rule will be such that a "fracture of the phalanx of the thumb" is a "fracture of the thumb", but not a "fracture of the arm". Indeed, this one-step inference through mereological relations is commonly accepted and used in everyday language, for example a "fist against the door knob" would be commonly accepted as a "fist against the door", but not as a "fist against the house".
[10] reminds us that transitive inference is used in medical classifications to talk of a "fracture of the phalanx of the thumb" as a kind of "fracture occurring in the upper limb".
Generic ontologies come into our aid to clarify the matter. "Fracture occurring in the upper limb", as well as "fracture of the arm", are metonymical terms that actually mean "fracture of a bone located at upper limb" (or arm), since "upper limb" and "arm" are body regions, and not body parts. A contribution of a good anatomical ontology is to define relations and rules that can support such metonymy, for instance:

```
(implies (:composition component located)
         located)
(implies (:composition embodied-in located)
         located),
```

Such rules allow to infer that:

- if a thumb phalanx is a component of a thumb that is located at an arm, that thumb phalanx is located at that arm as well;
- if a fracture is embodied in a thumb phalanx, it is also located at the arm where the thumb phalanx is located; then
- if "fracture occurring in the upper limb" is defined as a fracture *located* at the upper limb, a "fracture of the phalanx of the thumb" would be classifiable under it; and
- if "fracture occurring in the upper limb" is defined as a fracture of a *bone* located at the upper limb, a "fracture of the phalanx of the thumb" would be classifiable under it as well.

USING RELATION COMPOSITION TO DISAMBIGUATE VERB METONYMY

Metonymy creates polysemous relations that can be analyzed only by postulating 'chains' of relations. For example, physicians "treat" patients, patient groups, and conditions; therapies "treat" pathologies, abnormalities, and patients; devices "treat" abnormalities, etc.
Treat is not ambiguous in the experts' knowledge, but it is metonymically polysemous. Ontological theories should support the definition of relations that refer to the basic meaning of notions like "treat", but also they should reveal the relations chained in the metonymies. Formally, this is ideally accomplished by relation composition.
For example, after fixing the basic meaning of "treats" as ranging over healthcare operators and health conditions, we defined "treatment-action" for activities performed by operators during treatment:

```
(defrelation treatment-action
  :annotations ((DOCUMENTATION "The relation
for 'treats' when procedures used for
treatment are the domain."))
  :is (and clinical-actor
          (:composition performed-by treats)
          (:domain activity)))
```

Similarly, we defined "treatment-method", "treatment-device", "treatment-resource", etc.
An extreme example of this method has been defined in the theory for clinical guidelines, which are special plans describing the method of a medical procedure[11]. Guidelines usually focus on a "population group". But the intended meaning of *focus* is not simply stated. In our model, we chose to define a *target-population* as the composition of five relations already defined in the library.

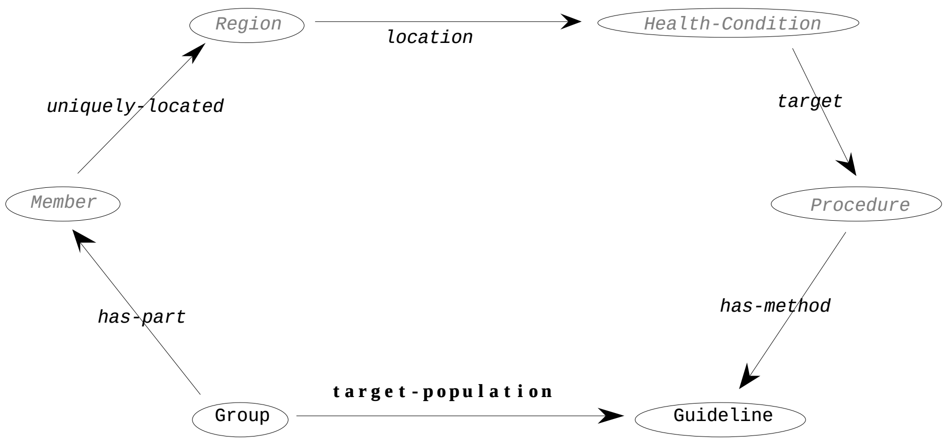

Figure 3. The definition of the relation "target-population" analyzed as a composition of five relations.

The metonymy here is very complex (Fig. 3), since a "group" is the "target population" of a guideline because it has "members" as parts that are "uniquely located" at some region, which is the location of some "health condition", which is the real target of the procedure that has the guideline as a method.

CONCLUSIONS

Polysemy is not only a matter of economy and context: it partly depends on some systematic conceptual structures that can be represented within ontology libraries with the aid of general theories of parts, connexity, time, localization, actors, etc. We have tried to apply this principle in the medical lexicon we are investigating since many years.
Conceptual analysis based on general ontologies helped us understanding systematic metonymies, the reason for multiclassification in the UMLS Metathesaurus[12], and greatly enhances the ongoing project of building an Integrated Medical Ontology Library cf [2].
We are aware that analyses like the ones presented here – although rigorous - are mostly not likely to become automatic and will remain a matter of craft in some measure. Nevertheless, we are also confident that a lot of groundwork like this would constitute a robust basis for the semantic management of data in the future.

References

1. Schulze-Kremer S, "Adding Semantics to Genome Databases: Towards an Ontology for Molecular Biology", 5th Int. Conf. on Intelligent Systems for Molecular Biology, Halkidiki, Greece, AIII Press, Menlo Park, 1997.
2. Gangemi A, Pisanelli DM, Steve G, "An Overview of the ONIONS project: Applying Ontologies to the Integration of Medical Terminologies", in *Data and Knowledge Engineering*, 1999, vol.31, pp. 183-220, 1999.
3. Borgida A, "Description Logic in Data Management", *IEEE Transactions on Knowledge and Data Engineering* 7(5): 671-682, 1995.
4. National Library of Medicine, *UMLS Knowledge Sources,* 1999 edition, available from the NLM, Bethesda, Maryland.
5. Pustejovsky J, "The Generative Lexicon", Cambridge, MA, MIT Press, 1995.
6. Langacker R, "Concept, Image, and Symbol", Berlin, Mouton De Gruyter, 1991.
7. Guarino N, "Formal Ontology and Information Systems" in N Guarino, *Formal Ontology in Information Systems,* Amsterdam, IOS Press, 1998.
8. WHO, *International Classification of Diseases* (10th revision), Geneva, WHO, 1994.
9. Coté RA, Rothwell DJ, Brochu L, eds. *SNOMED International* (3rd ed.), Northfield, Ill, College of American Pathologists, 1994.
10. Ceusters W, Buekens F, De Moor G, Waagmeester A, "The Distinction between Linguistic and Conceptual Semantics in Medical Terminology and its Implications for NLP-Based Knowledge Acquisition", in C Chute (ed): *Proceedings of IMIA WG6 Conference on Natural Language and*

Medical Concept Representation (IMIA WG6, Jacksonville, 1997) 71-80.

11. Pisanelli DM, Gangemi A, Steve G, "Towards a Standard for Guideline Representation: an Ontological Approach", *Journal of American Medical Informatics Association*, vol. S4, pp. 906-910, 1999.
12. Pisanelli DM, Gangemi A, Steve G, "An Ontological Analysis of the UMLS Metathesaurus", *Journal of American Medical Informatics Association*, vol. 5 (symposium supplement), 1998.