MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies

Domenico M. Pisanelli[a], Aldo Gangemi[a], Massimo Battaglia[b], Carola Catenacci[a]

[a] *National Research Council, Institute of Cognitive Science and Technology, Rome, Italy*
[b] *National Research Council, Institute of Neurobiology and Molecular Medicine, Rome, Italy*

Abstract

Polysemy is the linguistic phenomenon by which a term has more than one meaning. It is not a negligible issue in information management, since an effective and unambiguous sharing of the semantic content of data among different databases or knowledge repositories is needed. The paper illustrates a case of polysemy (concerning inflammation), and puts it within the framework provided by the DOLCE+ foundational ontology. This solution enables us to formally represent several senses of inflammation, and their interrelationships. This we take as a demonstration of how ontologies play an essential role in providing precisely the conceptual foundations that are needed in order to make the intended meaning of natural language expressions available to all the (artificial or human) agents that could be involved in the semantic web

Keywords
Ontology, Semantic Web, Inflammation

Introduction

Polysemy is the linguistic phenomenon by which a term has more than one meaning. It is widespread both in oral and written communication, due to a human tendency to make a thrifty use of linguistic items. Actually, the adoption of an already defined term to convey a different meaning is a far more common practice than creating complex sentences for each of the meanings to be expressed. Such procedure is at work in specialized languages as well, and medicine makes no exception [1].

Polysemy is definitely not a negligible issue in information management, since the current demand is for an effective sharing of the semantic content of data among different databases or knowledge repositories. It may seriously hamper crucial services such as intelligent information access, natural language processing or terminological standards definition and the widespread diffusion of the so-called "semantic web" [2].

Computer-based applications are often faced with a lack of conceptual foundations for terminology management, which makes them unable to meet the requirements of such added-value services.

Since polysemy is a natural linguistic phenomenon, it cannot be ignored in a truly semantic web, which has to rely on solid conceptual foundations for an effective and unambiguous terminology management.

Our position, presented in this paper, is that ontologies play an essential role in providing precisely the conceptual foundations which are needed in order to make the intended meaning of natural language expressions available to all the (artificial or human) agents that could be involved in a semantic service.

Is my inflammation your inflammation?

"Inflammation" is a typical polysemous word. At least five different meanings of inflammation can be recognized in different contexts in which the term may appear, i.e.:

- Inflammation *segregates* external agents
- The inflammation has a *diameter* of 5 cm
- The inflammation has changed its *shape*
- The inflammation evolved *during* three weeks
- The inflammation is *severe*

In addition to linguistic evidence, useful hints towards the development of a formal model can be found in medical dictionaries, where "inflammation" is defined in the following ways:

> "A local response to cellular injury that is marked by capillary dilatation, leukocytic infiltration, redness, heat, pain, swelling, and often loss of function and that serves as a mechanism initiating the elimination of noxious agents and of damaged tissue." [3]
>
> "A fundamental pathologic process consisting of a dynamic complex of cytologic and chemical reactions that occur in the affected blood vessels and adjacent tissues in response to an injury or abnormal stimulation caused by a physical, chemical, or biologic agent including: 1) the local reaction and resulting morphologic changes, 2) the destruction or removal of the injurious material, 3) the responses that lead to repair and healing." [4]

The first definition basically reflects the classic conceptualization of inflammation given by Celsus in the first century A.D., where its necessary attributes are listed: "rubor" (redness), "calor" (heat), "tumor" (swelling), "dolor" (pain). In the XIX century, Virchow – the founder of modern pathology – added the peculiar feature "functio laesa" (inhibited function) which, incidentally, is not "homogeneous" with the classical old four attributes. The second dictionary, on the other hand, stresses the nature of inflammation as a process.

We can start modeling such a process by defining it as: "response to a lesion that caused cellular death or suffering" and by listing its three main parts: 1) "elimination of causal agent", 2) "removal of cellular debris", 3) "repair of damaged tissue".

A complete inflammation model should also account for:

- *location* of inflammation, i.e. the area of an organ which bears an inflammation process;
- inflammation *agents* which may be exogenous (infection, trauma, temperature changing, toxins) or endogenous (ischemia, neoplasm, autoimmune pathology);
- *participants* in inflammatory response, e.g. vascular tissue (endothelial cells of vessels, blood cells, granulocytes, monocytes, lymphocytes, plasma proteins) or connective tissue (mast cells, fibroblasts, proteins of extracellular matrix).

Such a model should also consider various attributes:

- Acute vs. Chronic
- Type of vascular response
- Amount of infiltrated cellular substance (physical region)
- Type of lesion agent
- Duration (temporal region)
- Signs/symptoms (possible constituents of an inflammation condition, with different epistemological status, e.g. *pain* vs. *WBC count*)

The detail of an explanation of the intended meaning of inflammation, as it is available to experts, varies from gross distinctions to fine accounts of processes.

It should be clear now that inflammation is not an elementary concept to be modeled. We have collected useful material aimed at its formal conceptualization. In the next paragraph we will show how to build an ontology of inflammation and why, in our opinion, a particular kind of ontologies, i.e. the so-called “foundational ontologies”, are crucial in this process.

An ontology of inflammation

It is some years now that 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 [6]. The resulting libraries are one contribution to the set of tools necessary to cope with semantic mismatches in biomedicine.

The main principles of our formal ontology methodologies can be summarized as follows:

Logical consistency. Ontologies should be expressed in a logical language with an explicit formal semantics.

Semantic coverage. An ontology should assume to cover all the entities from its domain (*extensional* coverage), and it also should try to cover all the entity *types* of its domain (*intensional* coverage). For example, the assumption that all enzymes involved in inflammation pathways are in our domain of interpretation concerns extensional coverage, while a commitment to all the enzyme types that (according to the state of our knowledge) are conceivable in our domain of interest is an example of intensional coverage.

Modelling precision. An ontology should try to represent all and only the *intended models* for its domain of interest and for the tasks the ontology should accomplish. For example, an ontology of biochemical inflammation pathways would be *overcommitted* if it tried to represent relationships between enzymes and administrative healthcare concepts; it would be *undercommitted* if it failed to represent relationships between enzymes and the substances that play an enzymatic role.

Strong modularity. Ontologies should modularize the domain’s conceptual space as much as possible, by organizing the domain theories either into different ontologies or into different *descriptions* or *situations*. For example, “inflammation” as a healthcare concept will belong to a healthcare ontology module, while as a biochemical concept it will belong to a biochemical ontology module. Moreover, inflammation as a diagnostic concept (a description) will be different from inflammation as a condition (situation). In the terms of the Description and Situation theory, or D&S (see below), inflammation as a condition *satisfies* inflammation as a diagnosis (see [7] for an account of D&S theory and an example of its application)

Scalability. Expressive languages should be used to represent detailed accounts of intended meanings, according to the domain and tasks to be accomplished. Expressive ontologies can be used either as *reference* ontologies or to support sophisticated services, such as multiple databases querying. As reference ontologies, they can be scaled in order to support computationally hard services, e.g. information extraction from large repositories or retrieval across the semantic web. The scaled versions are “lightweight” versions of reference ontologies. Reference versions are maintained independently, and can produce updates of the lightweight versions according to knowledge management schedules.

Let us consider again from the five basic meanings of inflammation, as shown in the linguistic evidence:

1. Inflammation *segregates* external agents
2. The inflammation has a *diameter* of 5 cm
3. The inflammation has changed its *shape*
4. The inflammation evolved *during* three weeks
5. The inflammation is *severe*
6. The inflammation is *severe*

We can map these meanings into existing biomedical ontologies. In order to proof the concept, we will use our ON-9 top-level medical ontology [8], which is a merging of some authoritative medical ontologies. This can be done in the following way.

1. Inflammation as a *physiological function* performing segregation of external agents.
2. Inflammation as a characteristic portion of a *body part*, which embodies that physiological function.
3. Inflammation as a specific abnormal *morphology* of that portion (morphology is the appropriate medical category for shapes).
4. Inflammation as clinical *condition* encompassing all those entities (conditions *evolve*, hence they concern situations rather than single events or objects).

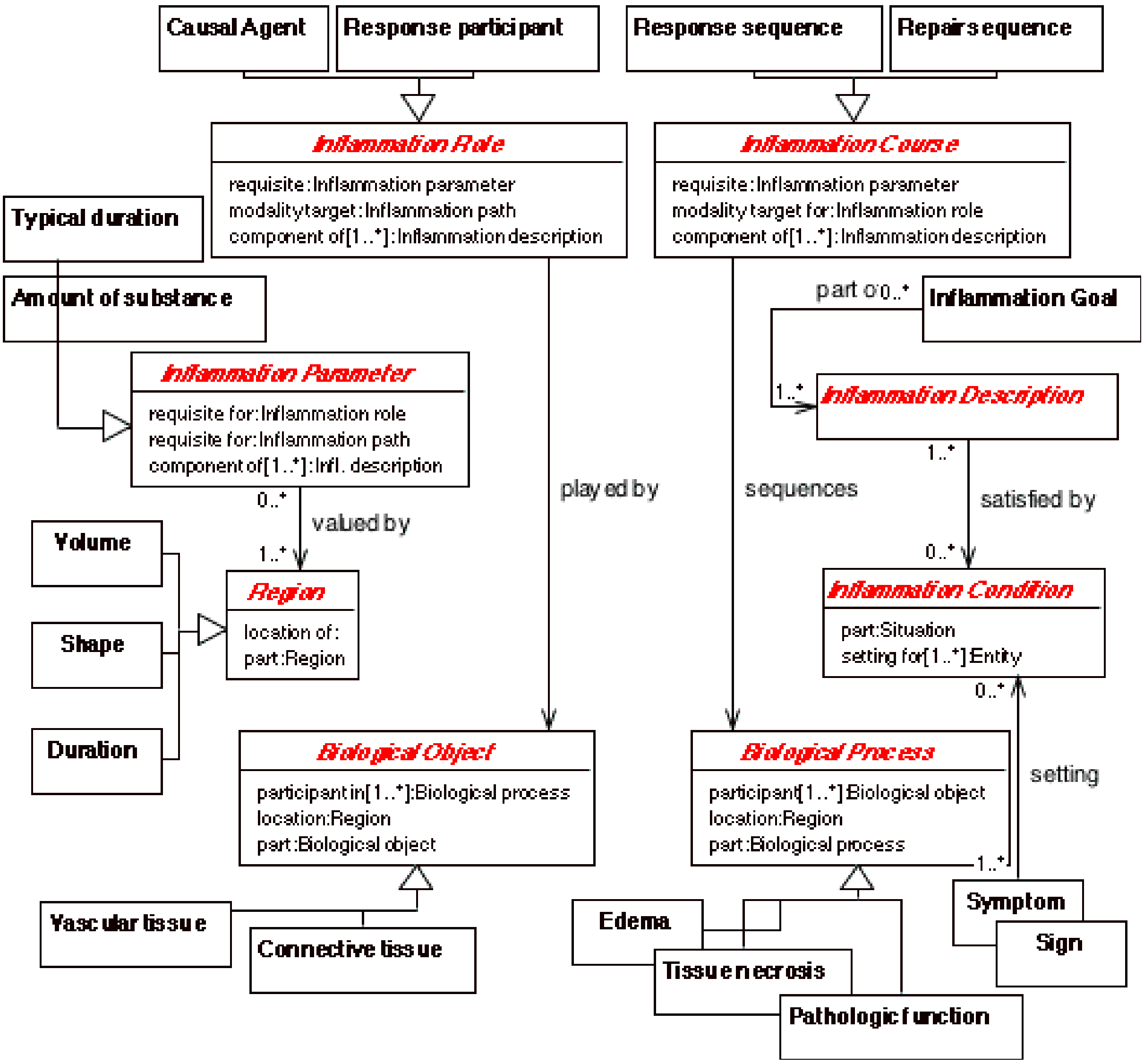

Figure 1 - Appying the DOLCE+ foundational ontology to the polysemy network that is instantiated by inflammation. Class with red names are those that specialize the D&S pattern for inflammation descriptions. «Inflammation as a diagnosis» is called in our ontology «inflammation description». A diagnostic description is adequate to some clinical condition («inflammation condition») only if the constraints posed by the diagnosis are satisfied by the condition's elements. Description constraints are called here: 1)»inflammation course», which can be «sequenced» by «inflammation as a process» (a kind of biological process); 2) «inflammation role», which can be «played» by «inflammation as inflamed object» (a kind of biological object); and 3) «inflammation parameter», which can be «valued by» by «inflammation as morphology» (a kind of region).

5. Inflammation as *diagnosis* applicable to that condition (being *severity* an appropriate attribute for an *assessment* rather than for an event or an objective condition).

These mappings to ON-9 leave unclear if e.g. diagnoses are something different from conditions. This "opaqueness" prevents an effective reconstruction of the interrelationships between the two concepts.

In order to obtain such reconstruction, we will use the DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) foundational ontology [5]. A foundational ontology is a very general (i.e. domain-independent) expressive ontology which defines a conceptual architecture that is shared by several communities of linguistic agents, thus being able of playing the role of a translator for intended meaning. Such kind of ontology contains an explicit, axiomatic description of the basic kinds of entities and relationships that are assumed to exist in a large variety of domains. DOLCE, in particular, has four basic top categories: *endurant* (including object- and substance-like entities, either physical or not), *perdurant* (process-, event- and state-like entities), *quality* (individual attributes), and *abstract* (including mainly conceptual *regions* for structuring attributes); and several primitive relations, such as *part*, *connection*, *constituency*, *inherence* of qualities in entities, *participation* of endurants in perdurants, etc.

The Descriptions and Situations ontology (D&S) is an attempt to support a first-order manipulation of *descriptive objects* (such as plans, diagnoses, norms, institutions, and so on – i.e., *theories*) and *situations* (such as cases, facts, settings, etc. – i.e, *models*). D&S commits to a distinction between an *unstructured world* or *context*, and an *intentionality* (description) that recognizes a *structure* (situation) in that world or context.

D&S is plugged into DOLCE (hence giving raise to "DOLCE+") as follows. A description is a non-physical endurant. A situation is a (new) DOLCE+ top category. A description satisfied by a situation is an s-description.

An example of D&S application is the following: a clinical condition (situation) has an associated diagnosis (s-description) made by some agent (endurant) participating in the diagnostic procedure (situation), in which the agent plays a functional role (physician), according to some diagnostic guideline (s-description) that has a set of steps (course), and certain thresholds (parameters) as components, etc.

Other examples of descriptions and situations include:

- A case in point (situation) is constrained by a certain norm (s-description)
- A murder (situation) has been reported by a witness (functional role) in a testimony (s-description)
- Information science as a topic (s-description) references the manipulation of data structures (situation), both as a pure or applied science (parent s-descriptions).

The linguistic evidence for the ambiguity of "inflammation" can be now revisited (Fig. 1):

Inflammation as a *physiological function* (*inflammation#1*) is a kind of <u>biological process</u> (a kind of perdurant)

Inflammation as a *characteristic portion of a body part* (*inflammation#2*) is a <u>feature</u> (a relevant part of an object) that embodies *inflammation#1*

Inflammation as a *specific abnormal morphology* (*inflammation#3*) is a <u>quality</u> (a <u>region</u> in a space of values) inherent in *inflammation#2*

Inflammation as a *clinical condition* (*inflammation#4)* is a <u>situation</u>, i.e. the setting for the entities (*inflammation#1, #2, #3*, etc.) whose relations have been evidenced through a diagnostic procedure or some other assessment.

Inflammation as a *diagnosis* (*inflammation#5*) is an <u>s-description</u> satisfied by *inflammation#4* iff *inflammation#4* constituents (including also *#1, #2, #3)* are compatible with its components (courses, functional roles, and parameters).

The ontology of inflammation so designed provides an analogue of the cognitive context available to an expert. For example, this ontology, together with other suitable tools for information extraction, might enable a service to contextualize any particular occurrence of *inflammation* with a degree of accuracy comparable to that of a human expert.

Conclusions

The services that are addressed by the semantic web are not fundamentally different from those traditionally addressed by database integration, requirement analysis, conceptual modelling, information integration, meaning negotiation, etc.

The current innovation that is emerging from ontology is based on the availability of languages, theories, and conceptual architectures that can be shared and reused beyond a local agreement. Therefore, the global nature of semantic services should be considered in its wide range of implications.

In particular, medicine is a very complex domain from the point of view of modeling and representing intended meaning. In such a discipline we find different activity domains (e.g. clinical vs. administrative knowledge), different scientific granularities (e.g. molecular vs. organic detail), different user requirements for the same service (e.g. physician-oriented vs. patient-oriented views), ambiguous terminology (polysemy).

Ontologies are used as basic infrastructures for modern interoperable systems and are the groundwork for implementing the semantic web. We have shown that such interoperability crucially depends on rich and carefully designed foundational ontologies, such as DOLCE+ used in our case study.

Detailed and context-sensible ontologies seem to impose an increased complexity to both the users that develop or reuse them, and the computational devices that exploit them. On the other hand, we envisage new tools that are able to 'hide' the complexity behind effective patterns and user-friendly interfaces, as well as methods based on an ontology maintainance life-cycle that considers both a *development procedure* for expressive ontologies used to negotiate and explicit domain assumtpions, versus a *runtime procedure* for lightweight ontologies used to carry out computationally heavy tasks.

References

[1] A Gangemi, DM Pisanelli, G Steve, "Understanding Systematic Conceptual Structures in Polysemous Medical Terms", in: J Marc Overhage (ed.), Proceedings of the 2000 AMIA Annual Symposium (2000).

[2] http://www.semanticweb.org.

[3] Webster Medical Dictionary; online at: www.intelihealth.com

[4] Steadman's Medical Dictionary, 26th edition, Williams and Wilkins, Baltimore, USA, 1995

[5] C Masolo, S Borgo, A Gangemi, N Guarino, A Oltramari, L Schneider. The WonderWeb Library of Foundational Ontologies. WonderWeb Deliverable 18, http://wonderweb.semanticweb.org (2003).

[6] A Gangemi, DM Pisanelli, G Steve, "An overview of the ONIONS project: Applying ontologies to the integration of medical terminologies", *Data and Knowledge Engineering*, 31(2):183–220, (1999).

[7] A Gangemi, P Mika "Understanding the Semantic Web through Descriptions and Situations", R Meersman et al. (eds.), Proceedings of ODBASE03 Conference (2003).

[8] Pisanelli DM, Gangemi A, Steve G, "A Medical Ontology Library that Integrates the UMLS Metathesaurus™", Lecture Notes in Artificial Intelligence 1620, Berlin, Springer Verlag, pp. 239-248, 1999.