Modelling Syllepse in Medical Knowledge Bases with application in the Domain of Organ Failure and transplantation

Christian Jacquelinet[1], Olivier Bodenreider[2] and Anita Burgun[3]

[1] Département Médical et Scientifique, Etablissement français des Greffes, 5, rue Lacuée, F–75010 Paris, France
christian.jacquelinet@univ-rennes1.fr
[2] National Library of Medicine, Bethesda, USA
olivier@nlm.nih.gov
[3] Laboratoire d'informatique Médicale,Faculté de Médecine, Rennes, France
anita.burgun@univ-rennes1.fr

Abstract. The Etablissement français des Greffes (EfG) is a public health agency in charge of organs, tissues and cells transplantation in France. Among EfG's missions is the evaluation of the organ retrieval and transplantation activities, which relies on a national information system (IS). In order to facilitate data recording, to improve information quality and homogeneity, to allow data interchange and semantic interoperability with hospital information systems and other registries, a specific work as been initiated dealing with ontological foundations of medical terminology in the domain of organ failure and transplantation. The aim of this paper is to describe how the syllepse appeared to us as a key figure that accounts for medical knowledge acquisition.

1   Introduction

Transplant teams periodically record data in the EfG-IS for their patients (registration on the national waiting list, follow-up). Because this work is time consuming and redundant with the data recording in the hospital information system and other database or registries, the evolution of the EfG-IS toward Electronic Data Interchange (EDI) has raised with a special focus on semantic interoperability. Interoperability denotes the capability of an IS to exchange data (functional interoperability) and its ability to utilise the exchanged information (semantic interoperability) [1]. Functional interoperability deals with the heterogeneity of database management systems, networks protocols and operating systems [2] whereas semantic interoperability involves the various way used to represent the real world entities, the differences concerning the meaning and the treatment of the data, and the relations between objects. Semantic heterogeneity concerns data structure (field names, attributes, and relations). It also deals with differences in covered domains, in underlying terminology and in exploitation goals. Thus, in order to facilitate data recording, to improve information quality and homogeneity, to allow data interchange, a specific work as been initiated dealing with ontological foundations of medical terminology in the domain of organ failure and transplantation.

The standardisation of the terms and thesauri is a first step toward semantic interoperability [3]. Therefore, a set of medical terminological systems such as ICD-10, SNOMED, UMLS [7] have been considered at the early phase of this work as knowledge information sources. Because of the specificity of the domain to cover, other specific thesauri relative to organ failure or transplantation [8] and the EfG-IS own ad-hoc thesaurus have also been used. As a second step, we chose the Conceptual Graph to support the deep semantic structure for knowledge representation in a home-built knowledge acquisition tool, called RIBOSOME [ref].

This system transforms natural language (NL) sentences and medical terms into a deep semantic structure according to the Conceptual Graphs formalism. During the acquisition process of sentences, unknown terms, Concepts and Conceptual Relations (CR) are integrated in the same semantic network. Each node of this semantic network is a context that associates lexical entries to its relevant concept or conceptual relation conditionally to the syntagmatic neighbourhood.

In analysing medical language, the syllepse, a classical rhetoric figure, appeared to us of importance as a common figure in medicine and a powerful tool for NLP and knowledge acquisition from Natural Language. We attempt in this paper to define the syllepse using the semiotic triangle with some extensions.

2 Adapting the semiotic triangle to model Homonymy, Polysemia, Synonymy and Syllepse

The classical semiotic triangle can be used to explicit the difference between Polysemia and Syllepse. In the following examples, we add some formal extensions such as markers to differentiate the context of occurrence of a term or a concept and two indicate the instance of an object. We use the following conventions:

{Term: $i} where term is a lexical entry (i.e. a word or a locution) encountered in the text or sentence and where $1 is a contextual marker that tags a context of occurrence of a given term;

[Concept_Type_label: referent_field] where Type_Label should be virtually arbitrary and meaningless and referent_field contains contextual markers and instance markers.

>Object: #i< where Object is a concrete or abstract object of the world and #i an instance marker of the object.
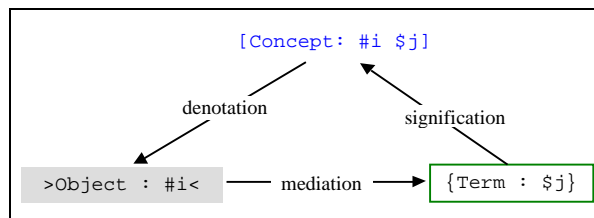

Figure 1: A concept, or a conceptual relation, denotes an object of the world, or referent, that is mediated by a term. This term has a signification that is the given concept or a conceptual relation. In the classical Saussurian approach, the linguistic sign unifies the term and the concept.

We can postulate that two distinct (class) objects of the world are denoted by two distinct concepts and conversely that two distinct concepts necessarily denote two distinct objects. But if we consider the world as a system and language as a model, we unfortunately have to notice that language performs some black box, at least grey box, approach of the world and that the direct links between the concepts and the objects are not or few accessible. In practice, links between objects and concepts are mostly accessible through the lexical entries that mediate the objects of the world.

If one consider that most Conceptual Relations support the signification of preposition, we have to assume that most if not all Conceptual relations have no object.

Let's define a context as a strand of words or locutions where a given lexical entry may occur with the same meaning. We can assume that each word of the given strand of words share the same contextual marker (Figure 2).
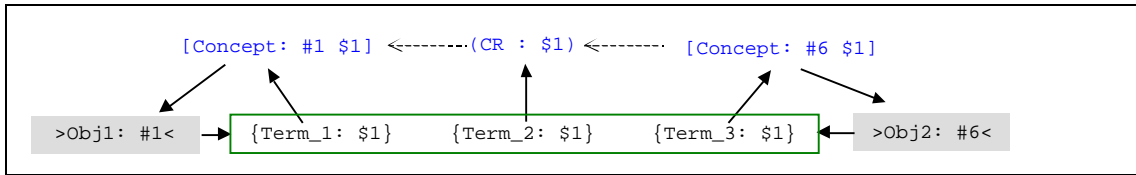
```
    [Concept: #1 $1] <-------(CR : $1) <-------  [Concept: #6 $1]

 >Obj1: #1<      {Term_1: $1}    {Term_2: $1}    {Term_3: $1}      >Obj2: #6<
```

Figure 2 : All connected terms of a given strand of words, say for instance a noun phrase, share the same contextual marker.

2.1   Homonymy, Polysemy and ambiguity

Two lexical entries that write the same but mediate distinct objects are homonyms. Because they mediate distinct objects, they have distinct signification.

```
    [Concept_1: #1 $1]          [Concept_2: #2 $2]

 >Object_1: #1<    {Term: $1}{Term: $2}    >Object_2: #2<
```
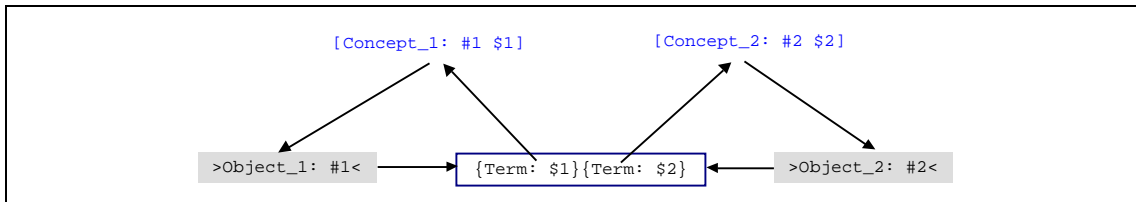
Figure 3: Homonyms mediate two distinct objects and have two distinct significations. They also appear in exclusive contexts.

For instance, the term *arm* as a weapon and *arm* as an anatomical member are homonyms. They mediate distinct objects and have distinct meaning, they don't share the same contextual marker.

Polysemy is sometimes assimilated to homonymy when it is applied to a single term. It may be a little more sophisticated when polysemy is applied to a strand of words, say a sentence or a phrase. Then polysemy denotes two or more competitive contexts where most often some Conceptual Relations have homonymic terms.

At a higher level, ambiguity is some generalised polysemy. Homonymy is to be considered as a simple type of polysemy.
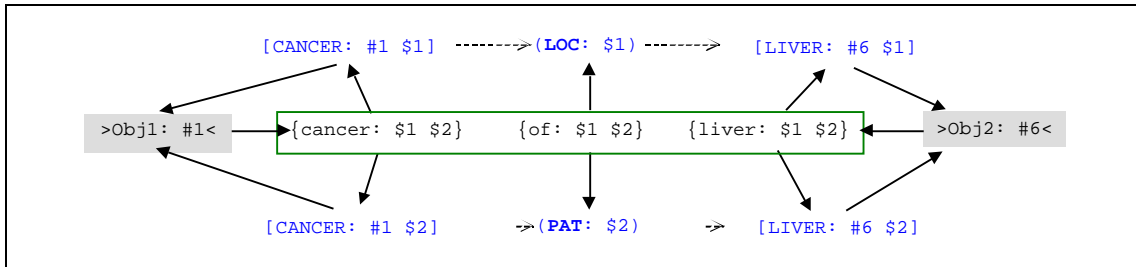
Figure 4 : A 'cancer of the liver' is stated here as a polysemic noun phrase because there are two possible roles for liver, one is the location of the cancer (LOC), the other is the patient of the cancer (PAT). This case is equivalent to the homonymy of the preposition 'of'. The two meanings are in competition and waiting for some disambiguation coming from the rest of the sentence.

At a morphological level, a sentence is a set of words separated by spaces. We can notice that a space between to words may be meaningless when there is no other relation between the words than contiguity. But when one or more conceptual relations may be instanciated by a space between two words, then the space can be considered as the hyperpolysemic lexical entry. Let's note it with the hyperpolysemic .

2.2   Synonymy, paraphrase, translation

Two distinct terms are synonyms when they have the same meaning in the same context. Two sentences or expressions are also synonyms or paraphrase when they have the same meaning in the same context. A word by word translation (when it exists) is also a paraphrase.
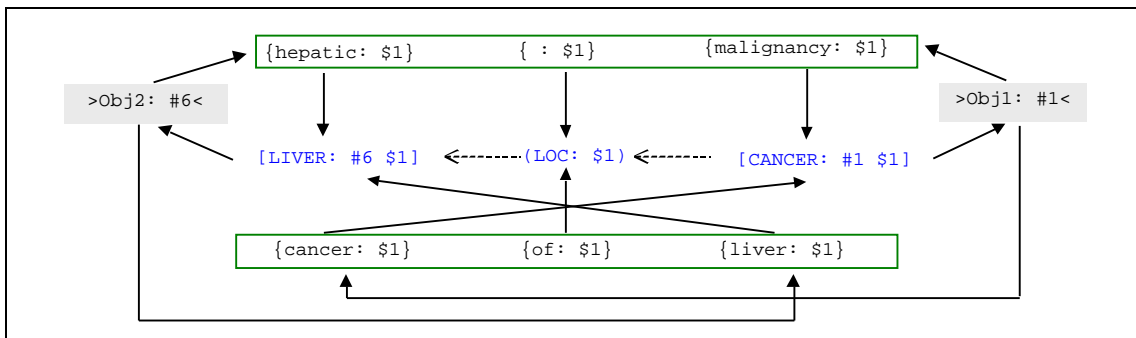

Figure 5 : An '*hepatic malignancy*' is stated here as a synonym noun phrase of '*cancer of the liver*' because *malignancy* and *cancer* as well as *of* and   share the same meaning in the same context.

```
  ┌──────────────────────────────────────────────────────────────────────────────────┐
  │          {transplantation: $12}    {des: $1}    {hertzens: $12}                    │
  │                                                                                    │
  │  >Obj2: #4< ←[TRANSPLATATION: #1 $12]  > (RESOU: $1) ---->[HEART: #45 $12]→ >Obj1: #45< │
  │                                                                                    │
  │          {transplantation: $12}    {du: $1}    {coeur: $12}                        │
  └──────────────────────────────────────────────────────────────────────────────────┘
```
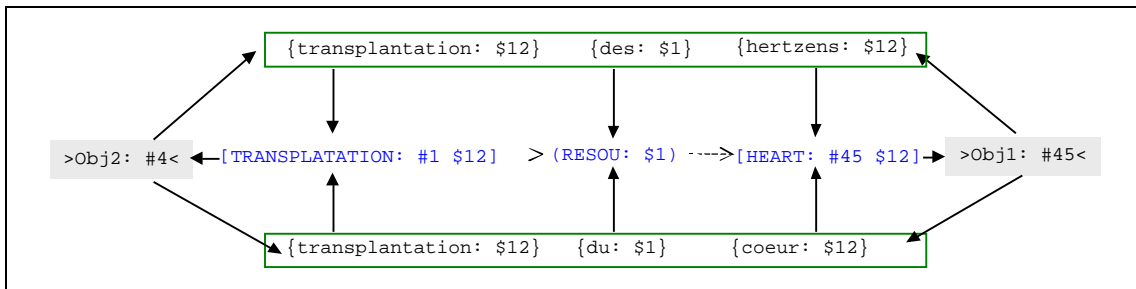
Figure 6 : A word by word translation of '*transplantation of the heart*' from French to German.

2.3 Syllepse

A syllepse is a rhetoric figure in which a lexical term has two co-occurrent meanings within the same sentence, generally a proper one and a metaphoric one. It differs from polysemous terms that can be distinguished by their context of occurrence and their different denotation. In medicine, the use of a lexical term to mean both the histo-pathological lesion and the disease process a very frequent phenomenon (tumour, cancer, fibrosis…). The same holds true for the lesion and the physio-pathological process (ischemia) or the clinical finding and the physio-pathological process (anaemia for instance).
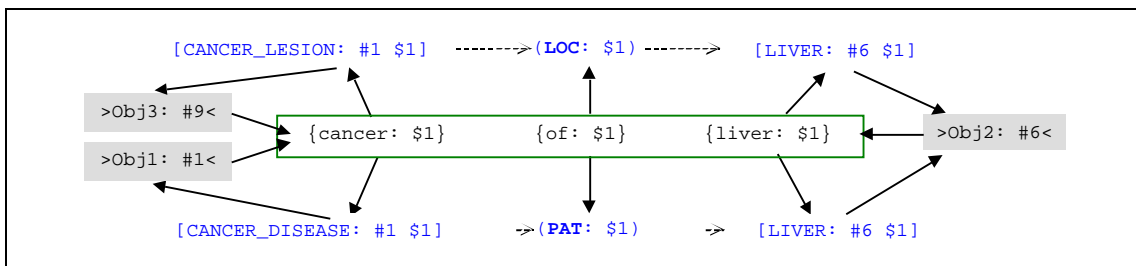
```
  ┌──────────────────────────────────────────────────────────────────────────────────┐
  │      [CANCER_LESION: #1 $1]  ------->(LOC: $1) ------->  [LIVER: #6 $1]            │
  │  >Obj3: #9<                                                                         │
  │              {cancer: $1}      {of: $1}      {liver: $1}          >Obj2: #6<       │
  │  >Obj1: #1<                                                                         │
  │      [CANCER_DISEASE: #1 $1]   -->(PAT: $1)    -->  [LIVER: #6 $1]                 │
  └──────────────────────────────────────────────────────────────────────────────────┘
```

Figure 7 : A '*cancer*' is stated here as a syllepse that mediates within the same context two different objects with two distinct meanings that occur together. Conversely, the dinstinct meanings are in competition in the polysemic representation given in figure 4.

2.3 Syllepse, a common figure and a powerful tool for NLP

During the analysis of a sentence, a syllepse simultaneously triggers the instantiation of two (or more) concepts or CR whereas a polysemous term triggers the instantiation of two (or more) mutually exclusive concepts or CR. Syllepse is more than a rhetoric figure. Instead of generating ad-hoc graphs to map the times for verbs (present, future, past), it is possible to consider the conjugated verb as a syllepse for the action type it denotes, and for the time type it denotes. In the following example, the verb *registered* is a syllepse for [Past] and [Registration].

```
    [REGISTRATION: #9 $6]    ------->(LOC: $6) -------> [WAITING LIST: #6 $6]

  >Obj3: #9<
                  {registered: $6}  {on: $6}    {waiting list: $6}          >Obj2: #6<
  >Obj1: #1<

    [PAST: #1 $6]                 >(TIM: $6)
```
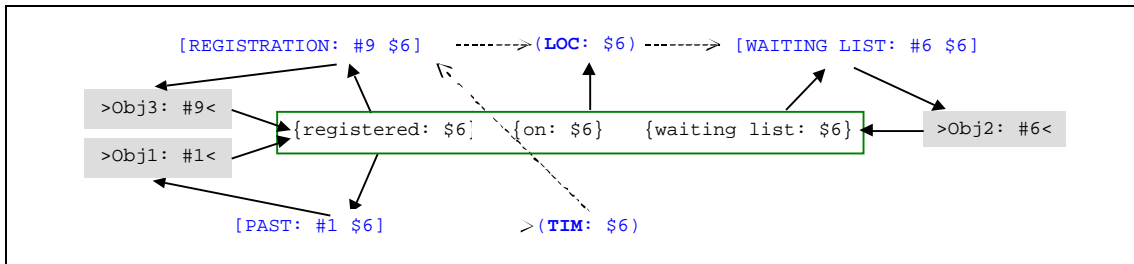
Figure 8 : *Registered* denotes simultaneously the past and the action of registration. In this context, the two concepts REGISTRATION and PAST can be linked by a temporal conceptual relation (TIM).

Discussion

Conclusion

With nearly 470 concepts, 680 contexts and 870 words and locutions, some RIBOSOME preliminary results looks promising because they confirm the accuracy and the strong efficacy of some underlying linguistic original considerations.

It appeared to us difficult to manage the semantic knowledge required for knowledge acquisition lonely with a top-down universal inheritance process of the semantic neighbourhood. A bottom-up syntagmatic existential inheritance process is indeed very useful and powerful tool. In this context, Syllepse is non lonely an interesting figure, but also a requirement to the inheritance bottom-up process.

Originally conceived to acquire knowledge from medical terminology, RIBOSOME is already capable to extract accurate CG representations from texts with a set of sentences.

In doing this, our experimental approach is also an attempt to fit the CG model to the NL data.

References

3   IEEE Standard Computer Dictionary : A compilation of IEEE Standard Computer Glossaries, IEEE (1990)

4   Sheth A, Larson J. Federated Systems for Managing Distributed, Heteregeneous and Autonomous Databases, ACM Computing Surveys, 22 (1990) 183-235

5   Dudeck J. Aspects of implementing and harmonising healthcare communication standards. Int J Med Inf. 48 (1998) 163-71

6   Rossi Mori A et al. Exploiting the terminological approach from CEN/TC251 and GALEN to support semantic interoperability of healthcare record systems. Int J Med Inf. 48 (1998) 111-24.

7   Mc Donald CJ et al. What is done, what is needed and what is realistic to expect from medical informatics standards. Int J Med Inf. 48 (1998) 5-12.

8   Beeler GW. HL7 version3 - an object-oriented methodology for collaborative standards development. Int J Med Inf. 48 (1998) 151-61.

9   Hammond WE. The role of standards in creating a health information infrastructure. Int J Biomed Comput. 34 (1994) 29-44.

10   Stengel B, Landais P. Data Collection about the Case management of end-stage renal insufficiency. Nephrologie, 20 (1999) 29-40.

11   Sowa JF. Conceptual Structures: Information Processing in Mind and Machine. Systems Programming Series. Addison Wesley Reading (1984)

12   Rassinoux A-M et al. Tuning-up Conceptual Graph Representation for Multilingual Natural Language Processing in Medicine. In: Mugnier M-L, Chein M. (eds.): Conceptual Structures: theory, Tools and Applications. Lecture Notes in Artificial Intelligence, Vol 1453. Springer-Verlag, Berlin Heidelberg New York (1998) 390-397.

13   Sowa J F. Conceptual Graph Summary. In: Nagle, Nagle, Gerholz, Eklund (eds): Conceptual Structures: Current Research and Practice. Ellis Horwood, New York London Toronto (1992) 3-66

14   Sowa JF. Knowledge Representation. Logical, Philosophical and Computational Foundations. Brooks/Cole (eds) Pacific Grove (1999).