A cooperative methodology to build conceptual models in medicine

Elena Galeazzi, Angelo Rossi Mori, Fabrizio Consorti, Anna Errera
Dottorato in Informatica Medica, Università La Sapienza, Roma
Reparto Informatica Medica, ITBM-CNR, Roma
IV Clinica Chirurgica, Università La Sapienza, Roma

Abstract

We designed a methodology to perform distribute activities on conceptual modelling among cooperating centers. Our methodology assigns responsibilities and tasks and regulates interactions preserving coherence; it passes through the construction of unambiguous paraphrases to make explicit the context within the original sources, and through their compositional representation in an intermediate language.

The process is intrinsically iterative, with continuous feedbacks and refinements, alternating analytic view on details and synthetic view on regularities and structures.

Our methodology is based on requirements and experience made in the first GALEN project, and was applied in the GALEN-IN-USE project to coordinate modelling activities of three teams of surgeons in Rome with activities of other partners, during the production of an extensive model of surgical procedures.

1. Introduction

Terminological systems used in healthcare include thesauri, nomenclatures, classifications, local controlled vocabularies, formal models [Rossi Mori 1993]. Diffusion of clinical information systems is shifting application of terminological systems to routine management of patient record with multiple re-uses including health care organization, evaluation and planning [Nowlan 1994; Rector 1995; Rossi Mori 1995]. Advanced methods, as *formal models*, are therefore required, providing adequate representation of terminological phrases within computer systems [Rector 1994, Galeazzi 1996].

The stream of UE-funded projects evolving from *GALEN* (1992-1995) to *GALEN-IN-USE* (1996-1998) is creating an environment for the development of methodologies, skills, formalisms, software and awareness about conceptual modelling in healthcare.

Bulding large concept models is an ambitious and expensive task: effort cannot be afforded by a single institution; it requires a large amount of experts in various domains — trained in compositional modelling and in the usage of the GRAIL language, ie. the formalism used in GALEN — both to build the model and to validate it. Decentralization of modelling activities is mandatory, and the issues on coherence, uniformity and integrability of the various contributions are crucial. The cooperative development of a model implies frequent revisions and reconciliations towards a *common modelling style*, with explicit decisions that affect the previous work of each center. The process *must be iterative*, with different layers of agreements, from general to specific; the work on more specific items will refine the working agreements among the centers at the more generic layers. Moreover, it would be hard to integrate cooperative efforts without a *unique conceptual framework, ontologically based*.

Cooperative modelling should be therefore supported by a *methodology* to extract and represent knowledge in an uniform way, based on

- early discovery of potential sources for conflicts among modellers, by focusing on anticipated issues and early reconciliation;
- minimal interaction among cooperating experts (ie. maximal autonomy), preventing incoherence by adequate structured discussions based on precise intermediate

documentation and by a consolidated set of rules and guidelines on a common modelling style.

We describe in this paper the methodology we worked out for the GALEN-IN-USE project; it was tested during 1996 in a cooperative effort by 4 "domain centers" in Europe (Italy — with 3 specialist teams —, Nederland, France, Sweden) interacting with a "GRAIL center" in Manchester, on more than 1000 phrases about surgical procedures. The final goal of the project is to demonstrate the feasibility of distributed modelling for a European nomenclature.

2. Working out a methodology for cooperative modelling

First in §2.1 we identify the kinds of skills of the various people involved in the analysis; then we outline in §2.2 the basic process of modelling an individual concept, from the expression in the original corpus to the canonical form in the GRAIL model. Finally, in §2.3 we describe the kind of activities to be perfomed in an iterative distributed process.

2.1. Define roles to assure effectiveness and quality

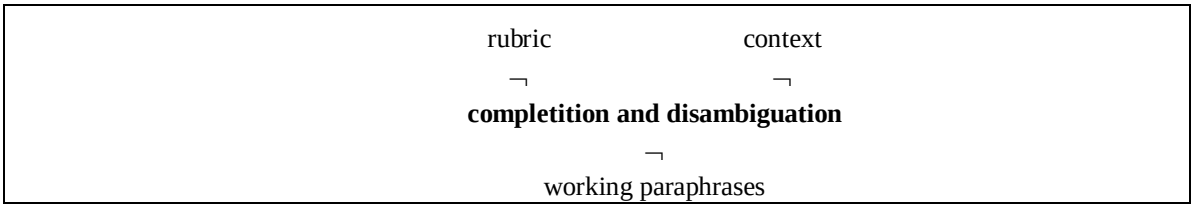
The first step was to identify the roles of the people that should interact; these roles correspond to skills that could be provided by one or more people; a person could have skills to perform different roles. We considered four different roles:

- the *specialists on the domain* that have to interpretate the corpora and gradually produce a structured representation of the rubrics;
- the *experts on terminologies and classifications* that have to organize, revise and homogenize the efforts of the specialists;
- the *GRAIL modellers* that produced or will produce the formal model (e.g. GRAIL model);
- the *coordinator* in each center.

We defined their responsibilities in a set of inter-related activities, and an iterative process of development, with products that they have to produce and gradually refine.

2.2. The basic process on individual rubrics

The modelling process should bring developers from a set of terminological phrases selected from an existing corpus to a set of representations of the related concepts into the formal model of GALEN, according to the GRAIL formalism used in the project (fig.1).



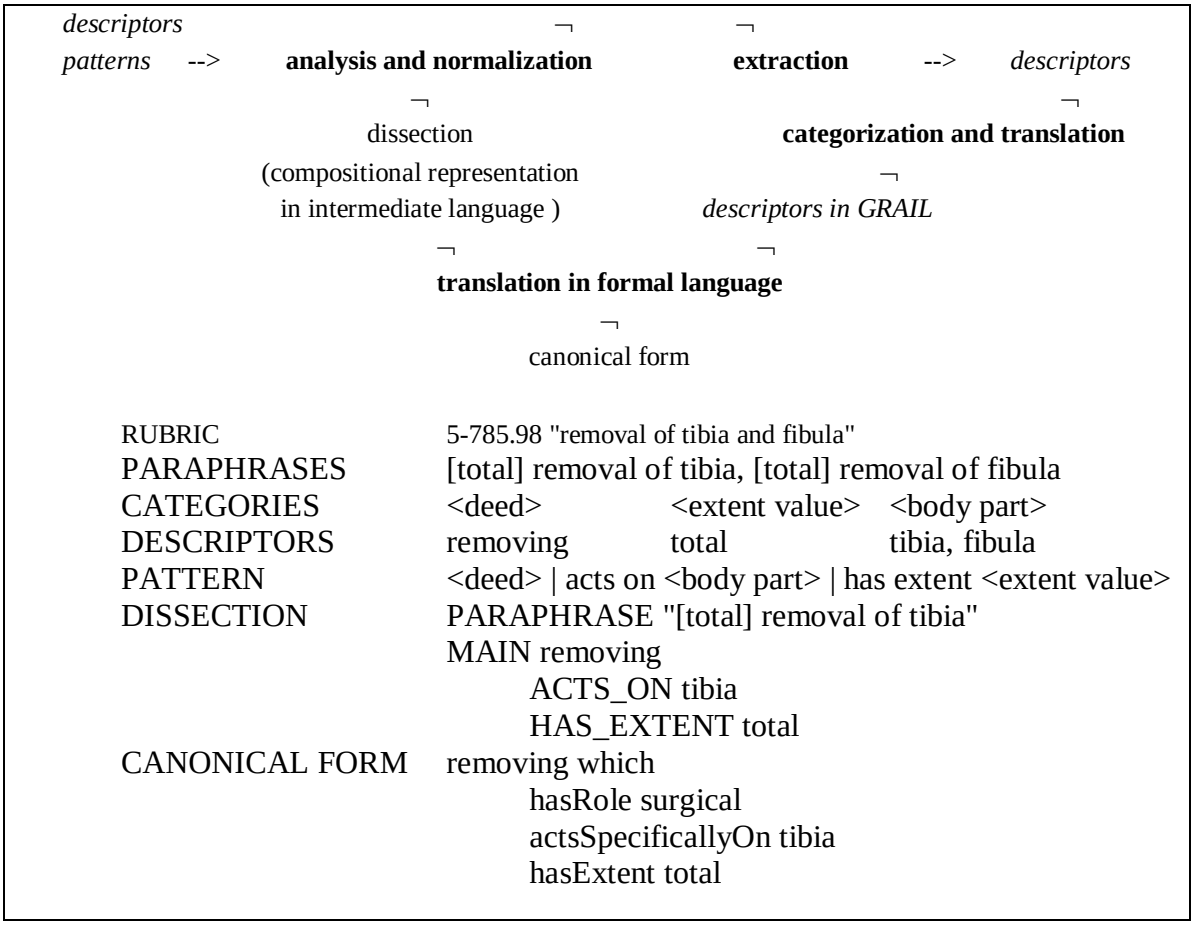


Figure 1. The modelling process and its application to an example
(items in italics are re-usable for other corpora)

2.3. Detailed description of the cooperative process

The basic process regarding an individual entry (Figure 1) implies three major activities :

- construction of unambiguous and explicit paraphrases that grasp hidden meanings and the context of the entry in the original corpus,
- their compositional representation in an intermediate language, and
- a semiautomatic translation into the GRAIL formalism.

In parallel, we extract and translate the atomic concepts needed in the compositional representation, ie. the building blocks (called "descriptors").

A preparatory phase of selection of sources — and of phrases within the sources — is also needed, with mutual awareness of the decisions among the cooperating centers.

In consequence, our methodology distinguishes five kinds of activities embedded in an iterative process; they can be schematized as in table 1.

Table 1. Description of activities in the iterative process of cooperative modelling

Activity 1. Prepare the subset of expressions to be analyzed
Each center specifies a corpus (one or more terminological systems) or collects a set of terminological phrases from patient files, textbooks, available paper forms or input layouts [Galeazzi 1994]. The domain specialist and the expert on terminological systems organize the corpus in "waves" (ie. they select a narrow subset of terminological phrases and harmonize them, by adding current phrases or removing too detailed or obsolete phrases). The results should be harmonized across centers.
Activity 2. Prepare working paraphrases

Each original terminological phrases in the wave should be checked for ambiguities, implicit information, errors, contextual information. From a rubric the expert can produce one or more "*paraphrases*", according to

- his/her interpretation of the meaning of the expression,
- the additional knowledge provided by the terminological system on that rubric, and
- the context of the rubric within the terminological system.

Paraphrases serve as reference for further modelling and do not replace rubrics in original sources. Each domain specialist should figure out — with the assistance of the expert — possible criteria to organize paraphrases of the current wave into medically meaningful clusters; each cluster suggests a concept ("*node label*" in the vocabulary of ISO TC46) that should be superordinate to all the elements of the cluster.

Activity 3. Extract candidate descriptors

Descriptors (activity 3) and dissections (activity 4) are strongly related. Within each working paraphrases, the domain specialist should separate the most general superordinate concept from differentiating characteristics that can be systematically expressed. Each superordinate concept is a candidate base concept (in our example, the descriptor "removing"). Each differentiating characteristic produces candidates for semantic links and associated concepts (in the example, the string "of tibia" produces the link "ACTS_ON" and the concept "tibia"). Descriptors belong to "*categories*" (eg. "removing" belongs to the category <deed>, tibia and fibula belong to <body part>). The domain specialist should systematize each descriptor under its own semantic category. The GRAIL expert should translate the set of descriptors using the GRAIL formalism.

Activity 4. Systematic production and harmonization of dissections

Starting from the paraphrase, the domain specialist uses (agreed) patterns and descriptors to dissect each paraphrase and to obtain an intermediate representation.The semi-formal representation in the intermediate language is called "*dissection*". A dissection is a semantic network made of a set of descriptors (eg. removing, tibia, fibula, total) related by means of semantic links (eg. ACTS_ON, HAS_EXTENT). Using the node labels (activity 2) domain specialists should verify that phrases within each cluster have similar dissections and harmonize the dissections in the whole wave. Local criteria for harmonization should be integrated into a common set of criteria across centers.

Activity 5. From dissections to canonical forms

The GRAIL expert translates the patterns into grammar-level statements in GRAIL; then, using the GRAIL descriptors already in the model (activity 3) he/she translates dissections into canonical forms. The feedback from this translation will assist not only in the discovering of errors and inconsistencies, but also in further harmonization among the representations of similar phrases (to increase the uniformity of style and to revise the common guidelines).

3. Discussion

The 'ideal' methodology should avoid as much re-modelling as possible by a preventive exercise (with timely reconciliations on problems), ie. it should:

- facilitate since the beginning interaction among teams working at different extensions;
- bridge between specialists and modellers.
- foster awareness and coherence in the modelling process.

To facilitate integration, the extensions produced by the individual teams have to use an explicit similar set of rules and the same "style", fully compatible with the ones already embedded in the model and compatible among them. These rules are partially enforced by the supporting software that is being developed by partners of the GALEN-IN-USE project, namely by the University of Nijmegen, the University of Manchester and CNR.

3.1. Requirements for a cooperative methodology

After our experience of direct GRAIL modelling in the first GALEN Project, it was clear that:

- the modelling effort requires a large amount of resources and different skills; therefore it had to be distributed among an adequate number of domain expert and terminological experts, and adequately coordinated;

- the GRAIL language has peculiar difficulties for "normal" physicians and cannot be used as the current formalism for distribute effort of analysis of expressions; therefore most of the experts should be enabled to focus on the issues of compositional modelling, independently from the additional difficulties of GRAIL modelling;
- the different subdomains are not homogeneous and the level of details that could be represented about each concept depends too much on the modelling style of the expert; therefore modelling activities in a field should be based on existing systematic corpora (in our case, mainly terminological systems on surgical procedures) and experts should use them to decide how many concepts they have to model, and how many details they have to represent about each concept;
- even if available corpora are systematic with respect to their needs, conceptual modelling requires a further systematization to obtain a set of phrase with homogeneous number of explicit details;
- issues and problems raised by the experimental work tend to increase to unmaneageable levels, because discussions tend to diverge on too many subtopics, and the amount of resources allocated to discussions must be balanced with the amount of resources to populated the model; therefore interaction among experts should be focussed on really crucial issues, and experts have to be aware of which decisions can be taken locally and which ones are for a common debate and consensus.

Our methodology was designed in order to provide an answer to these requirements.

3.2. Separate semantic issues from GALEN-specific implementation

Our methodology is based on the idea of an "intermediate representation" of rubrics of terminological systems by descriptors, initially developed and refined by two of the Authors (EG, ARM) during the first GALEN project. This attitude is intended to:

- involve as much as possible of specialists in the first phases of analysis;
- separate "what has to be there" from "how to express it in GRAIL".

The goal is to separate what is related to *any compositional representation* (eg. according to the CEN approach [Rossi Mori, 1997]) from the peculiarities of the GRAIL formalism.

This attitude is motivated by cultural, organizational and practical reasons:

- decisions are taken in the most appropriate context;
- more domain specialists can be involved, not exposed to GRAIL;
- it allows to exchange experience and data with other "non-GALEN" initiatives;
- it allows to re-translate the intermediate representations according to different releases of the formalism and the model.

The intermediate representation is also more "tolerant" about initial contradictions and irregularities, and can be used in preliminary phases of structuring and refining a raw model.

4. Conclusions

Advanced terminological systems are urgently needed. Conceptual modelling will be a bottle-neck for the diffusion of clinical information systems in healthcare.

A methodology to assign responsibilities and tasks and to regulate interactions preserving coherence is a prerequisite to distribute modelling activities among cooperating centers.

Our methodology exploits 5 different constructs:

1. paraphrases to decouple terminological systems with their context from the subsequent work on modelling;

2. descriptors	to detect issues of potential conflicts among centers;
	to prepare the translation into the formal model and to provide an early feedback to experts and domain specialists;
3. patterns	to facilitate uniformity of style among centers and to prepare the grammar-level statements in the formal model;
4. dissections	to manage a semi-formal intermediate representation, as a bridge between specialists and GRAIL modellers;
5. node labels	to refine the previous analytical work by comparative views, thus facilitating comparison of potentially similar dissections and the extraction of patterns.

Our methodology was applied successfully to coordinate the modelling activities of three teams of surgeons in Rome with activities of other partners in the GALEN-IN-USE project.

Acknowledgements.

Work partially supported by contract HC1018 "GALEN-IN-USE" from European Union.

5. References

Galeazzi 1994 — Galeazzi E, Agnello P, Gangemi A (et al). What is a medical term ? Terms and phrases in controlled vocabularies and continuous discourses. In: Barahona P, Veloso M, Bryant J (eds): Proceedings of the 12th Congress of the European Federation for Medical Informatics, Lisbon, 1994, 234-9

Galeazzi 1996 — Galeazzi E, Rossi Mori A. I servizi terminologici come elemento cruciale dei servizi informativi sanitari nell'era telematica. In "AIIM 96", Atti del IX Congresso Nazionale di Informatica Medica, Venezia ottobre 1996.

GALEN doc. — GALEN documentation, available from the main contractor AL Rector, Medical Informatics Group, Dept. Computer Science, Univ. Manchester, Manchester M13 9 PL, UK (e-mail galen@cs.ac.man.uk; URL=http://www.cs.man.ac.uk/mig/galen)

Nowlan 1994 — Nowlan W, Rector A, Rush T, Solomon W. From Terminology to Terminology Services. 18th Annual Symposium on Computer Applications in Medical Care (SCAMC94). Washington DC, 1994: 150-4.

Rector 1994 — Rector A. Compositional models of medical concepts: towards re-usable application-independent medical terminologies. In: Barahona P, Christensen JP eds. *Knowledge and Decision in Health Telematics*. Amsterdam: IOS Press, 1994; 109-14

Rector 1995 — Rector A, Glowinski A, Nowlan W, Rossi-Mori A. Medical concept models and medical records: An approach based on GALEN and PEN&PAD. Journal of the American Medical Informatics Association 1995;2(1):19- 35.

Rossi Mori 1993 — Rossi Mori A, Gangemi A, Galanti M. The coding cage. In: ReichertA, Sadan BA, Bengtsson S, Bryant J, Piccolo U eds. MIE 93 London: Freund Publishing House, 1993, pp 466-72

Rossi Mori 1995 — Rossi Mori A. Coding systems and controlled vocabularies for hospital information systems. *Int J Biom Comp* 39 (1995) 93-8

Rossi Mori 1997 — Rossi Mori A, Consorti F, Galeazzi E. Standards to support development of terminological services for healthcare telematics. to be presented at the Working Conference of IMIA WG6, Jacksonville, FL, January 19-22, 1997