Natural Language Generation of Surgical Procedures

Judith C. Wagner[a], Jeremy E. Rogers[b], Robert H. Baud[a], Jean-Raoul Scherrer[a]

[a]*Medical Informatics Division, University Hospital of Geneva, Switzerland*
[b]*Medical Informatics Group, Department of Computer Science, University of Manchester, UK*

Abstract

The GALEN-IN-USE project has developed a compositional scheme for the conceptual representation of surgical operative procedure rubrics. The complex representations which result are translated back to surface language by tools for multilingual natural language generation. Such generators can be adapted to the specific characteristics of the scheme by introducing particular definitions of concepts and relationships. We discuss how one generator uses such definitions to bridge between the 'style' of the GALEN scheme and natural language.

Keywords:

Natural Language Generation; Medical Concept Representation; Surgical Procedures; Description Logic.

Introduction

The GALEN-IN-USE project has developed tools and methods to assist in the collaborative construction and maintenance of surgical procedure classifications [1]. 7500 conceptual representations of individual surgical procedures across substantially the whole of the surgical domain have been authored so far; the goal is 10-15000 . The GALEN Representation and Integration Language (GRAIL), a description logic developed in the previous GALEN project, serves as the formalism for these representations. The medical knowledge model written in GRAIL - the GALEN Common Reference Model (CRM) - is a true conceptual model and therefore language independent.[2].

The representations of surgical procedure rubrics produced during the project will be integrated into the overall CRM. This integration allows for

1. an automatic classification of the procedures;
2. the support of refinement, extension and reorganisation of classifications by tools;
3. automatic generation of natural language expressions from the conceptual representation in several European languages.

To achieve this, the CRM is continually extended as required to include the concepts necessary for the representation of these procedures, as the need arises.

A Scheme for Surgical Procedures

The Technical Committee CEN/TC251 of the European Committee for Standardization CEN has proposed 'categorial structures of systems of concepts' in the medical informatics domain [3]. Building from this work, CEN ENV 1828 is a pre-standard scheme for the compositional representation of surgical procedures [4].

ENV 1828 provided the starting point for developing a GALEN scheme for representing surgical procedures [5]. Some adaptations and extensions were necessary in order to transfer CEN's representation to the GRAIL representation [6] and to enable a correct automatic classification of these procedures. The advantages gained through the use of GALEN techniques, but also the difficulties introduced by the additional complexity are discussed in [1].

The GALEN CRM is a deep, conceptual representation and is intended to be language-independent. Representations derived from even relatively simple natural language terminological phrases are necessarily complex; in the case of many of the surgical procedure rubrics they are almost unreadable (e.g. Figure 1). For human users, therefore, it becomes necessary to translate such a representation back to natural language for use within display applications. Furthermore, the evaluation of such a conceptual representation by its original authors also requires a translation back to natural language [7].

This paper discusses the techniques developed to present GALEN representations for display to humans - both knowledge authors and end users. We have developed specific operations to bridge between the conceptual representation and natural language. In particular, the knowledge which has to be introduced in order to deal with specific schemes such as the GALEN scheme for surgical procedures will be described.

Generation of natural language from a deep representation

As part of the GALEN project, we have implemented a multilingual natural language generator [8] which is based on a Conceptual Graph formalism [9]. It makes particular use of some of the operations defined within this formalism, as described in [10].

The graphs are transformed by a series of contractions and expansions of concepts or links. The transformations operate on definitions of concepts found within the CRM or additional (language oriented) definitions of concept types and relations. This enables the level of granularity to be adapted to the availability of terms in each language, and to the desired language style.

The language generation services provided are:

- as part of the GALEN Terminology server to generate natural language terms for concepts to be used within applications.
- to translate GALEN representations of terms of a terminology into several languages.
- to generate descriptions of concepts from their definitions.

To enable translation between a conceptual model and a linguistic model, structures for linking both are necessary. The main structures we use are:

- ‘concept-annotations’, i.e. annotations of concepts by corresponding words or terms in several languages including their syntactic properties;
- ‘statement-annotations’, i.e. annotations of semantic link relationships by syntactic structures qualified to express these relationships.

These annotations furnish the knowledge necessary for a translation of concepts to words and of links to syntactic structures. The generation results, however, can not be obtained by such a one-to-one translation. This is not only because annotations do not exist at every level in each language, but also because the conceptual representation usually includes information, not required for acceptable language generation, which ensures coherency of the overall conceptual model or the automatically derived classification. If the generated language is not to become too verbose, the GRAIL expression must be contracted before it can be translated in language - redundant information is discarded or ‘condensed’ into more precise words or structures.

However, our experience has been that this contraction is dependent on the language as well as on the modelling ‘style’, expressed by the scheme. An adaptation to the characteristics of the particular scheme is necessary, and is realised through the introduction of particular definitions. These definitions accommodate the specific concept types and combinations of attributes met in the scheme.

An Example

The example presented below is selected from a French catalogue of procedures, the ‘Nomenclature Commune des Actes Médicaux’ (Figure 1) [11]. This ‘rubric’ is manually decomposed into a ‘dissection’, expressed in an intermediate representation specifically designed for facilitating the modeling task as described in [1,5]. The dissection is then automatically expanded into a final GRAIL representation, according to an algorithm described by Rogers [1]. The final representation is clearly much more complex (Figure 1).

‘Rubric’

"Résection endoscopique cervico-prostatique, sans urétrotomie à l’aveugle, avec urétrocystoscopie et mise en place d'une sonde urétrale"
(endoscopic resection of the neck of the urinary bladder and prostate gland without urethrotomy, including urethrocystoscopy and insertion of urethral catheter.)

‘Dissection’

```
MAIN resecting
   ACTS_ON Anatomy:neck_of_urinary_bladder & prostate
   BY_MEANS_OF endoscope
   BY_TECHNIQUE inspecting
      ACTS_ON Anatomy:urinary bladder & urethra
      BY_MEANS_OF endoscope
WITH installing
   ACTS_ON Device:catheter
      HAS_LOCATION Anatomy:urethra
WITHOUT incising
   ACTS_ON Anatomy:urethra
   HAS_APPROACH closed approach
```

Conceptual representation: ‘GRAIL expression’

```
(SurgicalDeed which <
isMainlyCharacterisedBy (performance which isEnactmentOf
   ((Resecting which playsClinicalRole SurgicalRole) which<
      actsSpecificallyOn (ArbitraryBodyConstruct which <
         hasStructuralComponent UrinaryBladderNeck
         hasStructuralComponent ProstateGland>)
      hasPhysicalMeans Endoscope
      hasSpecificSubprocess (Inspecting which <
         actsSpecificallyOn (ArbitraryBodyConstruct which <
            hasStructuralComponent UrinaryBladder
            hasStructuralComponent Urethra>)
         hasPhysicalMeans Endoscope >)        >))
isCharacterisedBy (performance which isEnactmentOf
   ((InstallingProcess which playsClinicalRole SurgicalRole) which
      LocativeAttribute (Catheter which
         LocativeAttribute Urethra)))
isCharacterisedBy (nonPerformance which isEnactmentOf
   ((Incising which playsClinicalRole SurgicalRole) which <
      actsSpecificallyOn Urethra
      hasSpecificSubprocess (SurgicalApproaching which
         hasSurgicalOpenClosedness (SurgicalOpenClosedness
            which hasAbsoluteState surgicallyClosed))>))>).
```

Figure 1. Example of a French surgical procedure ‘rubric’, its ‘dissection’, and its GRAIL representation.

The first step in generating a natural language phrase from this complex GRAIL structure is to transform it into a Conceptual Graph. This serves as input to the natural language generation tool which produces phrases in several languages, as illustrated by Figure 2.

English:
'endoscopic cervico-prostatic resection under endoscopic urethro-vesical inspection with installation of a urethral catheter without closed urethra incising'
French:
'résection cervico-prostatique endoscopique sous inspection urétro-vésicale endoscopique avec mise en place d''un cathétère urétral sans incision urétrale fermée'
German:
'endoskopische Resektion des Blasenhalses und der Prostata unter endoskopischer Inspektion der Harnblase und der Urethra mit Einlage eines urethralen Katheters ohne geschlossene Inzision der Urethra'

Figure 2. Generation results in three languages.

Handling a scheme of surgical procedures

The GALEN scheme for representing surgical procedures has particular solutions for representing the 'objects' of a procedure, the combination of several procedures and the enactment of a procedure.

Direct and Indirect Object

The CRM expresses the relationship between a surgical deed (e.g. opening or cutting) and its primary, or direct object (the thing opened or cut) using the [actsSpecificallyOn] relation. For example, 'removing of kidney' is expressed as:

(SurgicalRemoving which actsSpecificallyOn Kidney)

Figure 3. Representation of a 'direct object'.

Such a deed on an object may also correspond to the definition of a specific procedure more commonly known by a specific term. For example, the concept defined in Figure 3 is more commonly termed 'Nephrectomy' in English.

(SurgicalRemoving which actsSpecificallyOn Kidney)
En: 'excision of kidney'
Fr: 'excision du rein'
Ge: 'Exzision der Niere'
(with contraction on concept):
En: 'nephrectomy'
Fr: 'nephréctomie'
Ge: 'Nephrektomie'

Figure 4. Generation results for the 'direct object' representation without / with use of concept definition for contraction.

Two strategies are possible: finding ['SurgicalRemoving' which actsSpecificallyOn Kidney] in a graph, the generator can either translate it back to natural language using the annotations on each individual concept and attribute; or it can use the concept definition for contracting it to ['Nephrectomy'] and generating a single word by using the annotations of this latter, composite concept.

In the same way the generator can deal with nested structures, such as the representations of the 'indirect object', usually the anatomy in which the direct object is located. Thus, 'removal of renal cyst' has direct object the cyst itself and indirect object the kidney from which the cyst is removed (a contraction is applied to the indirect object only for German):

('SurgicalRemoving' which actsSpecificallyOn
(Cyst which hasSpecificLocation Kidney)).
En: 'excision of kidney cyst'
Fr: 'excision d'un cyste du rein'
Ge: 'Exzision einer Nierenzyste'.

Figure 5. Generation results for 'direct object' and 'indirect object' (nested). The application of the type contraction operation is language-dependent.

The example shows how the translation output may be switched between different levels of verboseness using concept definitions of the model and respective annotations. It also shows that the contraction can only be applied if a word/term is available in the respective language.

Surgical Procedure and Surgical Deed

The basic GALEN scheme for surgical procedures shows how an individual surgical deed - excising, implanting etc. - is to be linked to the anatomy, devices and lesions involved in the deed. Within the CRM, the ability to nest or embed surgical procedures recursively is required, so that a complex surgical procedure might be defined as being the combined enactment of a number of (less complex) surgical procedures. Although not important for the conceptual representation, a main surgical procedure is designated by using the attribute 'isMainlyCharacterisedBy' in place of 'isCharacterisedBy' (Figure 6).

(SurgicalDeed which <
isMainlyCharacterisedBy
(Enactment which isEnactmentOf SurgicalDeed1)
isCharacterisedBy
(Enactment which isEnactmentOf SurgicalDeed2)>)
(SurgicalDeed which <
isMainlyCharacterisedBy (performance which
isEnactmentOf (Resecting ...))
isCharacterisedBy (performance which
isEnactmentOf (InstallingProcess.....)) >)

Figure 6. Representation of a surgical procedure as a combination of several nested surgical deeds.

Figure 1 provides a real example of such embedding. Its representation would translate to «surgical deed with performance of resection ... with performance of installation ...». However, this translation is not the desired result.

'Resection' has to become the main deed, the other deeds depending on it («resection... with installation»). This transformation is enabled by introducing a general concept type definition (scheme). The applied operation is a combination of a relation contraction (see next section), a concept type contraction (see previous section), an inversion of a relation, and a further concept type contraction. As a result the concept marked by the variable x (abstraction parameter) becomes the 'main concept' (Figure 7).

ling_definition(general,
[cl_SurgicalDeed,
[rel_isMainlyCharacterisedBy([cl_performance,
[rel_isEnactmentOf([**cl_SurgicalDeed:x**,
[rel_playsClinicalRole(cl_SurgicalRole)]])])])]).

Results:

((**Resecting** ...)
isCharacterisedBy (performance which
isEnactmentOf (InstallingProcess.....)) >)

En: 'resection ... with installation....'
Fr: 'resection ... avec installation ...'
Ge: 'Resektion... mit Installierung...'.

Figure 7. Concept type definition for the main surgical deed.

Such a contraction only applies to the surgical deed which 'isMainlyCharacterisedBy' another deed. It not only collapses a subgraph to one concept, it replaces the subgraph by a particular concept (instantiation) of it.

Performance and nonPerformance

The GALEN scheme also provides for capturing notions of cancellation, partial completion and other similar modalities of operative procedures which are found commonly in the real clinical world and in act management. This is done by particular complex links, which have a different instantiation of the 'Enactment' concept. For example, the links [isCharacterisedBy (performance which isEnactmentOf X)] and [isCharacterisedBy (nonPerformance which isEnactmentOf X)] are used for describing the 'performance' or the 'nonPerformance' of a particular deed as part of the complete procedure (Figure 8, Figure 9).

(SurgicalDeed which <
isCharacterisedBy (performance which isEnactmentOf X)
isCharacterisedBy (performance which isEnactmentOf Y)
isCharacterisedBy
(nonPerformance which isEnactmentOf Z)>).

Figure 8. 'Performance' of surgical deeds.

((Resecting ...) which <
isCharacterisedBy (performance which
isEnactmentOf (InstallingProcess ...)
isCharacterisedBy (nonPerformance which
isEnactmentOf (Incising ...) >)

Figure 9. Example of representation of the performance of surgical deeds.

In Figure 9, an initial transformation has already been applied to make 'Resecting' the main deed. The next operation transforms the relation [isCharacterisedBy (**performance** which isEnactmentOf SurgicalDeed)] into a relation which can be expressed in natural language by the preposition 'with'. In contrast to this, the relation [isCharacterisedBy (**nonPerformance** which isEnactmentOf SurgicalDeed)] has to be transformed to a different relation, expressed as 'without'. This is done using relation definitions: the two links which differ by the concept used are replaced by two different relations. These distinct simple relations can then be translated to different prepositions, such as «with» or «without».

rel_definition(reld_**isCharacterisedBy**,
[cl_SurgicalDeed:x,
[rel_isCharacterisedBy([cl_**performance**,
[rel_isEnactmentOf(cl_SurgicalDeed:y)]])]]).

rel_definition(reld_**not_isCharacterisedBy**,
[cl_SurgicalDeed:x,
[rel_isCharacterisedBy([cl_**nonPerformance**,
[rel_isEnactmentOf(cl_SurgicalDeed:y)]])]]).

Results:

((Resecting ...) which <
reld_isCharacterisedBy (InstallingProcess ...)
reld_not_isCharacterisedBy (nonPerformance which
isEnactmentOf (Incising ...) >)

En: 'resection ... with installation... without incision'
Fr: 'resection ... avec installation ... sans incision'
Ge: 'Resektion .... mit Installierung .... ohne Inzision'

Figure 10. Separate relation definitions for performed / unperformed deeds.

Results

A first large-scale experiment on representing surgical procedures in GRAIL and re-generating natural language phrases has been done on the Urology part of the French 'Nomenclature Commune des Actes Médicaux'. This part includes 522 'rubrics' (surgical procedures) prepared by the St. Etienne group [11]. Results have been generated in French, English and German.

Most of the linguistic effort is in introducing annotations and definitions. However, we found that the effort for annotation significantly diminishes once a domain is covered. For the generation of these rubrics and other rubrics from other centers and in other languages (Dutch, Italian) ample use of graph operations has been made.

Furthermore, our experiments showed that once a domain has been handled and the necessary definitions have been introduced it is much more straightforward to have results in an additional language by adding the respective annotations. We interpret this as confirmation of the multilingual approach, realised through the distinction of the domain model, linguistic model and their link by well defined structures, as well as the introduction of the language-dependent application of several operation types.

A persisting problem remains the representation of knowledge which is dependent on semantics, but is not conceptual knowledge. For example, there is no ordering of attributes by which to determine the main procedure or the relative order of adjectives to be generated. Some of this information is superimposed on the conceptual model as a linguistic meta-model (e.g. using the attribute 'isMainlyCharacterised' for the main procedure), but a more general solution seems to be desirable.

Conclusion

In this paper we show three examples of how language-dependent graph operations can be used to translate a particular conceptual modelling scheme back to natural language. Tailoring to the specific characteristics of a modelling scheme is required. This is achieved through the introduction of particular concept definitions, more general linguistic concept type definitions and relation definitions.

Experience shows the importance of well-defined linking structures between a conceptual model and linguistic model. It seems useful to pursue the approach of adapting the generator to specific modeling styles - such as the 'scheme' for surgical procedures - by introducing specific definitions which mirror this scheme and allowing a language dependent contraction of concepts and attributes which have been introduced for conceptual reasons.

Acknowledgments

GALEN-IN-USE is funded as part of Framework IV of the European Union Telematics for Healthcare programme. The Swiss part is funded by the Office Fédéral de l'Education et de la Science, Bern.

References

[1] Rogers JE, Solomon WD et al. Rubrics to Dissections to GRAIL to Classifications. In Pappas C et al. (Eds). *Medical Informatics Europe '97.* IOS Press, 1997; 241-245.

[2] Rector A. Compositional Models of Medical Concepts: Towards Re-usable Application Independent Medical Terminologies. In: Barahona P, Christensen JP (Eds). *Knowledge and Decisions in Health Telematics.* IOS Press, 1994; 109-114.

[3] CEN ENV 1828. *Health care informatics - Structure for classification and coding of surgical procedures.* Brussels: CEN, 1995.

[4] CEN ENV 12264. *Medical Informatics - Categorial structure of systems of concepts - Model for representation of semantics.* Brussels: CEN, 1995.

[5] Rogers JE, Rector AL. Terminological Systems: Bridging the Generation Gap. J Am Med Informatics Assoc 1997; Symposium Supplement: 610-614.

[6] Rogers JE, Rector AL. Deliverable 5.1 *Extended Core model for representation of the Common Reference Model for procedures.* The GALEN-IN-USE Consortium. (Available from Medical Informatics Group, University of Manchester).

[7] Baud RH, Rodrigues J-M, Wagner JC et al. Validation of Concept Representation Using Natural Language Generation. J Am Med Informatics Assoc 1997; Symposium Supplement: 841.

[8] Wagner JC, Solomon WD, Michel P-A et al. Multilingual Natural Language Generation as Part of a Medical Terminology Server. In: Greenes et al. (Eds). *MEDINFO 95.* Proceedings. Amsterdam: North-Holland, 1995; 100-104.

[9] Sowa JF. *Conceptual Structures: Information Processing in Mind and Machine.* Reading, MA: Addison-Wesley Publishing Company, 1984.

[10] Wagner JC, Baud RH, Scherrer J-R. Using the Conceptual Graphs Operations for Natural Language Generation in Medicine. In: Ellis G et al. (Eds). *Conceptual Structures: Applications, Implementation and Theory.* Berlin: Springer-Verlag, 1995; 115-128.

[11] Rodrigues J-M, Trombert-Paviot B, Baud R et al. An EU project applied to the development of a new national coding system for surgical procedures: NCAM. In Pappas C et al. (Eds). *Medical Informatics Europe 97.* IOS Press, 1997; 897-901.

Address for correspondence

Judith Wagner
Medical Informatics Division, University Hospital of Geneva
24, rue Micheli-du-Crest / 1211 Geneva 14, Switzerland
email: Judith.Wagner@dim.hcuge.ch