Ceusters W, Michel C, Penson D, Mauclet E. Semi-automated encoding of diagnoses and medical procedures combining ICD-9-CM with computational-linguistic tools. Ann Med Milit Belg; 1994;8(2):53-58.

Semi-Automated Encoding of Diagnoses and Medical Procedures Combining ICD-9-CM with Computational - Linguistic Tools

W. Ceusters (1, 2 ,3), Cl. Michel (1,4), D. Penson (1, 5), E. Mauclet (1)

(1) Department of Informatics, Military Hospital Queen Astrid, Brussels
(2) Office Line Engineering NV, Sint Lievens Houtem
(3) Department Of Medical Informatics, University Hospital, Gent
(4) Department of Medical Informatics, University Hospital, Woluwe
(5) Staff of the Medical Service

Abstract
The rationale for clinical coding is to associate concepts with precise meanings to patient data. One of the most widely used coding systems is ICD-9-CM. Due to several limitations associated with this system, coding cannot adequately be performed without additional computational - linguistic tools. In this paper, the development of a clinical thesaurus and an appropriate set-theoretic algorithmic search facility that assists physicians in finding a set of possibly relevant ICD-9-CM codes starting from a diagnostic statement in natural language, is described. Although very powerful, the approach as such cannot account for all possible problems. An increase in performance can be realised by analysing not only the semantic contents of a user query, but also the syntactic realisations through which meaning is conveyed.

Samenvatting
Het coderen van klinische gegevens komt neer op het associeren van patient-gebonden informatie aan concepten met een precieze betekenis. Een van de meest gebruikte codeersystemen is ICD-9-CM. Verscheidene tekortkomingen aan dit systeem staan het efficient coderen in de weg, tenzij men beschikt over computer-linguistische hulpmiddelen. In dit artikel worden de ontwikkeling van een thesaurus en van bijhorende zoek-algorithmen gebaseerd op de verzamelingentheorie beschreven. Deze algorithmen stellen de arts in staat om op basis van een uitdrukking in natuurlijke taal, de relevante ICD-9-CM codes terug te vinden. Ondanks de performantie van het gerealiseerde systeem kunnen niet alle problemen opgelost worden. Daarvoor moet niet alleen de semantiek, maar ook de syntactische kenmerken van de gebruikersvraag aan een analyse onderworpen worden.

Résumé

L'encodage de données cliniques permet d'associer des informations liées au patient avec des concepts précisément définis. Un des systèmes de codage les plus utilisés repose sur la classification ICD-9-CM. Différentes limitations de cette approche rendent indispensable l'utilisation de technologies informatiques exploitant les acquis récents en linguistique. Cet article décrit le développement d'un thésaurus médical et d'un algorithme de recherche approprié en vue d'assister le clinicien dans la sélection des codes ICD-9-CM correspondant à un diagnostic clinique exprimé en langage naturel. Bien que l'approche soit prometteuse, différents problèmes résiduels persistent, qui pourront être résolus par une analyse approfondie des dimensions syntaxiques et sémantiques des propositions formulées par le clinicien.

Introduction

The **medical record** is the cornerstone of the clinical practice as it permits the combination of information from different sources and provides the basis for diagnostic, therapeutic and management decisions. The more precise the information the record contains, the more reliable the conclusions based upon these data will be.
Part of the medical record deals with **diagnostic expressions**, statements describing what diseases or pathologies the patient is suffering from or has experienced previously. Diagnostic expressions need to be as clear as possible and may not be subject to misinterpretation. Only this can guarantee that proper and continuous care can be given to patients by the various physicians and healthcare workers by whom they are treated.
Another part of the medical record describes the actual care or therapy that has been given to the patient. A specific subset of expressions that can be found in this part of the medical record deals with **procedures**, statements describing what interventions have been carried out, or are planned to be carried out. Once again, procedure statements need to be very precise, not at the least to have an accurate description of what happened to a specific patient.

However, medical natural language, as any language, is not free of ambiguities. Depending on the cultural models shared by the speakers of medical language, it is the result of influences and conflicts among the medical language varieties of clinical communities, characterised by medical specialty, task, nationality, school of thought, etc. (1). Physicians not always are capable to resolve ambiguities in messages uttered by their colleagues, especially not when there are too many differences between the clinical communities they are living and working in. Standardised coding and classification systems in medicine are designed to reduce this ambiguity in meaning to a maximal degree, such that efficient communication between different clinical communities can be achieved. This approach has been (and currently still is) favoured by the Council of Europe (2), CEN\TC251 (the Technical Committee 251 on Medical Informatics of the European Standardisation Committee) (3, 4), and the Commission of the European Community (5, 6) at various occasions. Indeed, since the recent interest in Medical Telematics, terminology work in medicine has gained much importance. Research in this field has shown that medical data mostly are built around concepts that only can be described with sufficient detail by using representations built around multiple axes (7, 8). However, concept systems that are able to fulfill the different requirements of various healthcare workers at the same time, are very difficult to design. Major efforts are needed in this domain (9, 10).

The coding task

The rationale for clinical coding is to associate concepts with precise meanings to patient data. A *concept* is defined as a unit of thought independent of any particular language (11), a definition we take for granted in order not to be led into discussions that would bring us back to the time of Aristotle, and perhaps even earlier. Concepts with precise meanings are to be looked for in standardised concept systems. A *system of concept* is a structured set of concepts established according to the relations between them, each concept being determined by its position in the set (11).

Concepts in concept systems are represented by a term, expression or rubric, and uniquely identified by means of a code. The concept of a medical procedure by which a surgeon removes a portion of the liver is designated in ICD-9-CM by the term *partial hepatectomy*, and identified by the code *50.22*. The codes are assigned to the concepts following a coding scheme. As a result, a coding system can be viewed as a specific type of concept system in which the concepts are identified by means of a code according to a coding scheme. A *classification* is a system of concepts where the relations between the concepts are generic in nature. Usually, this hierarchical relationship is expressed in the codes themselves. E.g.: 50. = operations on the liver, 50.2 = local excision or destruction of liver tissue or lesion, 50.21 = marsupialisation of lesion of the liver, 50.22 = partial hepatectomy, ...

Coding patient data means that a physician (or professional encoder) has to describe the patient data by means of the concepts available in the coding system to be used. The requirements to be met in order to perform the coding task adequately , are outlined in table 1.

1. a perfect understanding of the meaning of the patient data (the source concepts), 2. a perfect understanding of the meaning of the concepts available in the concept system (the target concepts), 3. a minimal level of similarity and coherence between the source concepts and target concepts, 4. facilities to search the concept system for the target concept(s) that match(es) a given source concept as closely as possible.

Table 1: Requirements for adequately performing the coding task.

ICD-9-CM

The **International Classification of Diseases, 9th Revision, Clinical Modifications (ICD-9-CM)** is a modified version of the World Health Organisation's ICD-9, designed for the classification of morbidity and mortality information for statistical purposes. The origin of ICD goes back to 1893 when the International List of Causes of Death, also known as the Bertillon Classification was formalised. The latest version of this continuously updated system is ICD-10 of which Volume 1 recently has been published. A classification of diseases may be defined as *a system of concepts to which morbid entities are assigned according to established criteria* (modified from (12)). ICD-9-CM is a *clinical* modification. The term clinical is used to emphasise the intent of the modification: to serve as a useful tool in the area of classification of morbidity data for indexing of medical records, medical review, and ambulatory and other medical care programs, and not merely for basic health statistics (13). To describe the clinical picture of a patient, the concepts defined in the system must be more precise, i.e. having a smaller grainsize, than those needed only for statistical grouping and trend analysis.

ICD-9-CM currently is one of the most widely accepted coding systems for diseases on an international level. In many countries, it is used for the indexing of hospital records by disease and procedure, for data storage and retrieval.

The system consists of a structured arrangement of concepts, each concept being uniquely identified by means of a code, and defined by a term or expression in natural language. The relationships between the concepts are of pure generic, hierarchical nature (IS_KIND_OFF relationships). This relationship is also expressed in the codes.

Coding with ICD-9-CM as such is not trivial a task. First, it should be clear that ICD-9-CM codes or expressions **cannot replace** the original expressions uttered by the physicians. Payne et al. showed that physician satisfaction with an exact ICD-9-CM representation of phrases in the medical problem list of the medical record, was disappointing in 45 % of cases (14). ICD-9-CM is just not expressive enough for this purpose. What needs to be done, is that physicians **translate** their original statements into ICD-9-CM only for **registration purposes**, but not for **documentation purposes**.
ICD-9-CM provides some limited functionalities to assist the physician in this coding or translation task. A first functionality is, of course, the intrinsic hierarchical classification scheme itself. A clinician wanting to code a ***meningitis due to cryptococcus*** (a non-viral, non-bacterial infectious agent) may start his search in the chapter *Diseases of the Nervous System and Sense Organs* (codes 320-389), look further in the sub-chapter *Inflammatory Diseases of the Central Nervous System* (codes 320-326), to finally arrive on the pages covering *meningitis* (codes 320-322). One serious problem with ICD-9-CM is that one never can be sure that all types of meningitis have to be coded between 320 and 322. Some specific types of meningitis indeed need to be coded differently. Examples are *tuberculous meningitis* (013.x) or *meningitis caused by Leptospira* (100.81)

A second utility provided by ICD-9-CM is the Alphabetic Index. In this index, keywords are associated with a set of relevant codes. When looking in the Alphabetic Index for the word *meningitis*, a physician quickly will see that this pathology appears in quite various chapters. Within the Index, some clues are given to restrict the search required, but nevertheless, the coder has to go to the classificatory part of ICD-9 to see what code is applicable. Apart from its user-unfriendliness, this facility also has a number of inconveniences. Usually, when an upgrade of ICD is developed, the Alphabetic Index is the last part to be completed. This means that the system must be used for a long time without the availability of the Index. Moreover, as ICD is being developed in English, translations of the Index may last even longer. Another problem is the occurrence of synonyms, homonyms, quasi-synonyms and lexical variants denoting the primitive concepts making up ICD-9-CM expressions. To a certain extent, some additional structure to deal with them is provided in ICD (the "dagger-asterisk" links), but the meaning of these links is highly variable and requires human interpretation.

These problems call for computational - linguistic tools that are able to assist the physician in encoding diagnoses and procedures properly. How this is achieved in several applications developed by the authors, is described in the next paragraphs. First the development of a thesaurus, linking terms to concepts and concepts to ICD-

9-CM, is covered. Second, the search-algorithms to propose relevant ICD-9-CM codes on the basis of natural language expressions, are described.

Linguistic oriented basic terminology work on ICD-9-CM

Prior to developing the tools, each word used in ICD-9-CM has been analysed semantically. As such, basic concepts have been identified. A basic concept is defined as an abstract entity having one single meaning in the field of medical diagnostics. This meaning is expressed by (a number of) terms, a term being (part of) a linguistic representation of a concept. After this extraction of words, all terms have been mapped to the relevant concepts. The identification of concepts based on the terms encountered has primarily been based on linguistic characteristics: pre-, suf-, and affixing at the one hand, and the occurrence of compound words at he other hand. Only in very specific cases, domain knowledge has been used to associate terms with concepts. Similar approaches can be found in the literature (15, 16, 17, 18).

The relationships between terms and their associated concepts, are classified on the basis of a degree of coverage. Three types of relationships are used.

The *covers* relationship is used when the term covers the concept completely. It is used in case of synonyms, lexical variants, adjectives, nominalisation of verbs, etc., i.e. when the syntactic realization of a concept does not affect the meaning. For each term, there can be no or only one concept having this relationship with the term. At the contrary, each concept has at least one term bearing this relationship.

The *Is-Broader-Than* relationship (IBT) is used when the term is more general than the concept it is associated with. This occurs in two situations: 1) when the concept linked to the term in the IBT-relationship is a physical part of the concept *covered* by the term (in this case, there is a PARTITATIVE relationship. E.g.: the meninges are parts of the cerebrum), and 2) when the set of objects referred to by the concept linked to the term in the IBT-relationship constitutes a specific subset of the objects referred to by the concept covered by the term. In this case, there is a GENERIC or, inverse IS_KIND_OF relationship. We call this a relation of hyperonimy. E.g.: an ascaris is a kind of worm.

Finally, the *presumes*-relationship is used when the concept covered by the term presumes (includes) the concept in the relationship. It is mostly encountered in case of compound terms, or when pre-, suf- or affixes change the meaning of the basic concept. Table 2 gives some examples.

Term	Concept	Relationship
encephalitis	encephalitis	covers
encephalitis	brain	presumes
encephalitis	inflammation	presumes
meninges	meninge	covers
medulla	brainstem	IBT

Table 2: term - concept relationships.

The bridge to ICD-9-CM is made by associating the concepts (not the terms) with the ICD-9-CM expressions.

Set - theoretic algorithmic search

The purpose of our system is to allow an efficient access to ICD-9-CM. It must be possible for users of our system to get a relevant set of ICD-9-CM expressions, starting from a query in natural language.
The steps to be performed by the system are the following. First, the query of the user is analysed and transformed into a number of concepts. This is done by removing the function words, i.e. words without a specific meaning in the domain of diagnosis and medical procedures, from the query, and by searching the thesaurus for each concept that has a relationship with the terms not removed from the query. The set of concepts is then used to restrict the complete list of ICD-9-CM expressions to a relevant set of expressions. Finally, the user has to perform a (manual) selection procedure to select the ICD-9-CM expression looked for initially.

Due to the different semantics of the various term-concept relationships as described above, the search procedure is different for each set of concepts associated with the individual terms in the user query. Hence, a two step approach is needed to find the relevant set *relevant_codes* of ICD-9-CM expressions for a given query.

Firstly, for each term, the relevant ICD-9-CM expressions have to be found separately, and temporarily stored in a list $LIST\_TERM_i$. To do this, for each concept associated with that term, the relevant ICD-9-CM expressions need to be found and stored in a list called $LIST\_CONCEPT_k$.

The following rules apply:

- If the relationship is of kind COVERS, $LIST\_CONCEPT_k$ **equals** $LIST\_TERM_i$, In this case, each term found represents precisely the concept asked for.

- If the relationship is of kind IS_BROADER_THAN, $LIST\_TERM_i$ is the **union** of the lists $LIST\_CONCEPT_k$, $LIST\_CONCEPT_{k+1}$, ... E.g.: if the concept looked for is *inflammation*, all the words ending with “-itis” are valid terms.

- If the relationship is of kind PRESUMES, $LIST\_TERM_i$ is the **intersection** of the lists $LIST\_CONCEPT_k$, $LIST\_CONCEPT_{k+1}$, ... E.g.: if the concepts looked for are *stomach* and *inflammation*, then *gastritis* would be a valid term.

Secondly, *relevant_codes* can be defined as the **intersection** of the various LIST_TERM's found.

Results

The approach described above has been implemented (with minor variations) in various applications. It is currently used in a number of military medical centers, and will be integrated in the future hospital information system of the military hospital of Brussels.

The main benefit of this terminological approach is that the classical *string-search* in text retrieval systems is replaced by a true *semantic search*. Semantic search makes it possible to arrive at the same list of items by entering queries that are semantically equivalent but of which the meaning is expressed syntactically different. As a consequence, queries such as ***retinitis***, ***inflammation of the retina, retinal inflammation,*** ... all lead to the same set of ICD-9-CM codes.

Semantic search also allows the user to query the knowledge base for more general concepts than those that are really looked for. Indeed, sometimes distinct terms such as **retino*pathy*** and **retin*itis*** are used for denoting the same concept. Physicians tend to do this, but also in ICD-9, the distinction is not always made in a consistent way. A physician looking for the code for **retinitis caused by diabetes,** will not find relevant codes. But when using the keyword combination **retina diabetes,** the system will give him all the possible expressions.

As the system uses set - theory applied to semantic entities for its basic operation, a number of limitations are predictable, and indeed do occur. However, it turns out that in practice these limitations are not of great importance, or are not difficult to address.

The main objective of the system is to provide physicians a tool with which they can use their own words to find specific ICD-9-CM expressions. The current version however is not based on a sublanguage analysis of physicians' utterances. As such, the dictionary does not contain all the synonyms and terms that are used in various clinical linguistic communities. Eponyms such as Kuru disease, Laplace syndrome, etc., are a typical example of this. This is not a serious problem as missing concepts and terms easily can be added in subsequent versions of the system.

Another problem is that ICD-9-CM does not provide codes and expressions for each single disease. Some diseases need to be classified under a more general heading. The relation between the disease and the heading can only be expressed using world knowledge, not linguistic knowledge. A possible solution is to use the BROADER-THAN relationship, but this will increase drastically the required number of primitive concepts.

In medicine, the meaning of terms frequently can be analysed based on the meaning of the stem of a word, and the associated pre-, and suffixes. Dujols for instance found for French 58 suffixes with a specific meaning in medicine and 14 prefixes. Using morphosemantic analysis of words, it would be possible to restrict the number of dictionary entries. However, much research in medical natural language processing has been carried out on this topic, without giving stable results. This has to do with the many exceptions that exist for the meaning of given suffixes (19, 20, 21). The suffix -osis may be used for an abnormal condition or non-inflammatory disease, e.g. nephrosis. But what is the meaning of -osis in hypnosis ? Morphosemantic analysis also could be useful to classify various syntactical realizations of the same concept under one dictionary entry, e.g. in Dutch: bacterie, bacteriën, bacterieel, bacteriële, ...

Queries are processed by the system with the aim to arrive at a limited number of meaningful keywords. No attempt however is made to analyse the syntactic relations

between the terms in the expression. Both the queries ***diabetes in renal diseases***, and ***diabetes without renal diseases,*** would result in the same set of candidate ICD-9-CM expressions. Another example is ***cardiac disease caused by hypertension***, and, ***hypertension caused by cardiopathy***.
As the current applications are not used for automatic encoding purposes, this is not an important problem. The physician is able to select the correct expression by himself.

A last problem is associated with concept-explosion. Concept-explosion is the phenomenon encountered when mapping a compound term on the individual concepts by means of the PRESUMES-relationship. If a number of terms, coming from one expression, are exploded, it is not possible anymore to detect where the resulting concepts have come from. This has some peculiar results. The expression ***cerebral arteritis*** (meaning inflammation of an artery of the brain), explodes in the concept's **brain, inflammation** and **artery**. If the user enters the query ***encephalitis*** (meaning inflammation of the brain), then this term is exploded in **brain** and **inflammation**. As a result, the set-theoretic approach selects the expression ***cerebral arteritis***, as a valid candidate, not knowing that the concept of **inflammation** in this expression is not associated with **brain**, but with **arteria**.

Conclusions

Coding diagnoses and medical procedures is an important topic for understanding the activities of healthcare personnel in- and outside hospitals. A number of requirements need to be met in order to perform the coding task adequately, i.e. to associate patient data with standardized concepts agreed upon by the clinical community in which the task has to be carried out. One of these requirements is the need for adequate facilities to search the concept system for the target concept(s) that match(es) a given source concept as closely as possible. In this paper, we described the development of a clinical thesaurus and an appropriate set-theoretic algorithmic search facility that assists physicians in finding a set of possibly relevant ICD-9-CM codes starting from a diagnostic statement in natural language.
Although very powerful, the system revealed to have a number of drawbacks due to the terminological methodology used. The problems mentioned do not affect the usefulness of the system when used in interactive mode, but make it impossible to have the coding task performed by a machine without human interaction.

The main reason for this non-perfect behaviour is that three relationships that explicitly have been defined in the thesaurus bridge exclusively the gap between terms and concepts. The only relationship that is used between concepts, is a hierarchical one, and it is implicitly defined in the ICD-9-CM codes. Actually, the term-concept relationship IS_BROADER_THAN ontologically may be viewed (and rather "is") as the combination of the term-concept relationship COVERS, and an inter-concept relationship IS_BROADER_THAN. Introducing inter-concept relationships means structuring the knowledge domain. Although standardisation of medical classifications would extremely be valuable in view of the increasing need for information exchange in health care on a national and international level, and of the increasing usage of data processing tools, standards in this area are very rare, even non existing. Another important issue has to do with the representation of terminological systems in health care. Traditional schemes heavily rely on human interpretation. As a consequence, issues of the precise language used for referring to concepts are intimately associated with and difficult to separate from the concept system itself. For this reason, it is argued that in the approach to medical terminologies, a paradigm shift is needed (22): *This shift is away from the enumerative classification of concepts and terms towards compositional models of medical concepts with formal properties and away from single structures of concepts and terms, and towards language independent concept systems which are interpreted through separate grammars and lexicons of terms.*

Finally, increase in performance can be realised by analysing not only the semantic contents of a user query, but also the syntactic realizations through which meaning is conveyed. Indeed, a close interaction between the conceptual knowledge and the linguistic knowledge must be accounted for. For this reason, we strongly belief that a more functional approach of medical language should be looked for. By functional, we mean that natural language is used - and as a consequence should be analysed as such - as an adequate instrument for medical knowledge description and manipulation. The functional dimension of natural language has been convincingly introduced by Dik in theoretical linguistics (23), and successfully applied and adapted by Deville for the modelisation of administrative language (24). The same approach seems promising with respect to medical diagnostic and surgical procedure language analysis and representation.

Both tasks will be addressed in the research project ANTHEM (Advanced Natural Language Interface for Multilingual Text Generation in Healthcare) funded by the Commission of the European Communities in the framework of the Linguistic and Research Engineering Programme. Together with research teams from Belgium, Luxembourg and Germany, the Medical Informatics Department of the Military Hospital of Brussels will work out a system in which encoding can be performed fully automatically. The final results are expected in 1996.

References

1. Rossi-Mori A, Bernauer J, Pakarinen V, Rector AL, De Vries-Robbé P, Ceusters W, Hurlen P, Ogonowski A, Olesen H. Models for Representation of Terminologies and Coding Systems in Medicine. In: De Moor G, McDonald Cl, Noothoven van Goor J (eds.). Progress in Standardization in Health Care Informatics, 92 - 104, 1993.

2. Rodrigues J.M., Hansen R., Pakarinen V., Roger F.H., Scicluna M., Ruijs J., Wiley M. Computerisation of medical data in hospital services, including university hospitals. Coordinated Medical Research Programme 1985, Council of Europe, (1985).

3. CEN\IT\PT001. Problems and needs in relation to standardisation in Health Care Informatics (as per mandate BC-IT-SI-05A). CEN\TC251 15-01-1991, (1991).

4. CEN\TC251. Directory of the European Standardisation Requirements for Health Care Informatics and Programme for the Development of Standards. Version 1.7, (1993).

5. Roger F.H. Standardisation and Harmonisation of Medical Data Classification Systems in Europe. CEC DG XIII, BICEPS, (1986).

6. Commission of the European Communities. R&D on Telematic Systems in Health Care, AIM Workplan (1991).

7. Gabrieli E. Computer-based Medical Terminology and Knowledge Representation. In: De Moor G., McDonald C., Noothoven van Goor J (eds.) Progress in Standardization in Health Care Informatics, IOS Press, Amsterdam, 81 - 86, 1993.

8. Olesen H. A Model for Coding and Multiaxial Classification Examplified. In: De Moor G, McDonald Cl, Noothoven van Goor J (eds.). Progress in Standardization in Health Care Informatics, IOS Press, 86 - 91, 1993.

9. Ceusters W. Deville G, Buekens F. The Chimera of Purpose- and Language Independent Concept Systems in Healthcare. In: Proceedings of Medical Informatics Europe 1994. In Press.

10. Ceusters W, Bonneu R, De Moor G, Lapeer R, Thienpont G. The Challenge of the Nineties: Bringing Multimedia Healthcare Records to Life. In: Proceedings of MIE'93. Reichert A, Sadan BA, Bengtsson S, Bryant J, Piccolo U (eds.), Freund Publishing House, England, 594-599, (1993).

11. International Standards Organisation. Terminology - Vocabulary (ISO - 1087), Geneva, ISO.

12. World Health Organisation. ICD-10: International Statistical Classification of Diseases and Related Health Problems, tenth revision, WHO, Geneva, 1992.

13. Comission on Professional and Hospital Activities (eds). ICD-9-CM, International Classification of Diseases, 9th Revision, Clinical Modifications. 1th Edition 1978.

14. Payne TH, Murphy GR, Salazar AA. How well does ICD-9 represent phrases used in the medical record problem list ? In: Proceedings of SCAMC'92, AMIA Inc, 654 - 657, 1992.

15. Frutiger P. Prospective encoding, transcoding and pragmatic representation of medical language. In: Scherrer JR, Côté RA, Mandil SH (eds.) Computerized Natural Medical Language Processing in Knowledge Engineering. Elsevier Science Publishers BV, North-Holland, 83 - 94, (1989).

16. Dujols P, Baylon P, Chein M. Project LIME: linguistiqe et langage médical. Informatique & Santé, 5: 126 - 138, 1992.

17. Gabrieli E. Criteria for a Computer-Oriented Medical Nomenclature. Non published communication, MEDINFO 92, Geneva, 1992.

18. Milan J. The READ Clinical Classification Version 3 File Structure; a developer's needs. Report prepared for the NHS Centre for Coding and Classification, march 1993.

19. Wolff S. The Use of Morphosemantic Regularities in the Medical Vocabulary for Automatic Lexical Encoding. Meth Inform Med 23: 195 - 203, 1984.

20. Wolf-Terroine M. Terminology and Nomenclatures. In: SchneiderW, Sagvall Hein AL (eds.) Computational Linguistics in Medicine, North-Holland, 55-61, 1977.

21. Dujols P, Aubas P, Baylon C, Grémy F. Morphosemantic Analysis and Translation of Medical Compound Terms. Med Inform Med, 30: 30-35, 1991.

22. Rector A, Nowlan A, Glowinski A, Matthews G. GALEN Deliverable 6, The Master Notation, version 1, (internal document) 31/03/93.

23. Dik S. Functional Grammar. North-Holland, Amsterdam, (1978).

24. Deville G. Modelization of task-oriented utterances in a man-machine dialogue system. PhD thesis, University of Antwerp, (1989).