Ceusters W, Deville G, Buekens Ph. The Chimera of Purpose- and Language Independent Concept Systems in Health Care. In Barahona P, Veloso M, Bryant J (eds.) Proceedings of the XIIth International Congress of EFMI, 1994;:208-212.

The Chimera of Purpose- and Language Independent Concept Systems in Health Care.

Ceusters W [a], Deville G [b] and Buekens F [c]

[a] *Dept of Medical Informatics, University Hospital, De Pintelaan 185, 9000 - Gent, Belgium*

[b] *Institute of Modern Languages, FUNDP, Rue de Bruxelles 61, 5000 - Namur, Belgium*

[c] *Dept. of Philosophy, University of Tilburg, P.Box 90123, The Netherlands*

Wide integration of health care information systems and unlimited exchange of patient data and medical knowledge call for the development of language- and purpose (e.g. collection of morbidity data) independent concept systems. If those systems are intended to be used by computers in an automatic way, formal representation and computational tractability are additional requirements. Physicians' (and other health care workers') acceptance of such systems can only be realised if the expressiveness of those systems is satisfactory. Finally, the elaboration of such systems should follow the guidelines and directives of relevant ISO and CEN standards. In this paper, we argue that no system can satisfy all these requirements at the same time. By identifying the contradictions that justify this claim, we bring to light the concessions that have to be made in order to arrive at workable systems.

1. Introduction

The transmission and comparison of heterogeneous data is recognised as a major issue in health care informatics. At a European level, a multi-purpose representation of knowledge and information, based on explicitly stated principles within a uniform framework, is essential in order to allow the integration of health care systems throughout Europe. CEN\TC251 has recognised this need, and has outlined the approach to be taken [1]. Central in this approach is the development of concept systems that are language- and purpose independent, highly expressive, fully formalised, and when implemented on a machine, computationally tractable. In order to evaluate whether this is really feasible, a careful analysis of these requirements should be carried out.

2. The structure of a concept system

Concept Systems are defined by ISO as *structured sets of concepts established according to the relations between them, each concept being determined by its position in this set,* a "concept" being defined as *a unit of thought constituted through abstraction on the basis of properties common to a set of objects* (i.e. any parts of the perceivable or conceivable world) [2]. As a consequence, to develop a concept system, the first activities that must be undertaken are: (i) identifying the concepts that must be represented in the system, and (ii) outlining the relationships between them.

Already at this point, major problems may be encountered. The first one deals with the criteria that should be adopted when deciding whether a given concept needs to be represented in the system or not. To clarify this issue, we can use Rector's example (be it in a slightly different context) of what should be understood by *clinical information* (i.e. falls in the medical domain), and what shouldn't [3]: *that the pain is aggravated by cold is 'clinical', that it comes on when walking past the freezing compartments at the local supermarket is 'anecdotal'.* Saying that a given patient caught a cold when walking along the freezing compartments of a given store is indeed not sensible when having only therapy in mind, but surely is sensible when preventive health care policies are

discussed. What may be viewed as 'anecdotal' (thus irrelevant) in a certain context, might turn out to be relevant information in another application field.
A number of problems are associated with the structural properties of a concept system, more specifically with the nature of the relationships between the concepts, be it subsumption, part-kind, or pragmatic relationships. It is not always clear what should be introduced in a concept system as a concept (e.g. what should become a node in a semantic network) and what should be a role (e.g. the links between nodes in a semantic network). For instance, we could start by introducing in a concept system the concepts *disease* and *symptom*. Then we could proceed by representing a *meningitis* as a *kind of disease*, and headache as a *kind of symptom*. But what should we do with something such as *hypertension*. Hypertension can be a disease, but at other occasions, just a symptom of a disease such as nephrosis. When detecting such a problem, it might be an indication that one (or both of the concepts) could be introduced as a role or relationship. One solution in the example given could be that symptom should be declared as a role. But what 'kind of thing' is then a *headache*, and how can we express the notion that concepts such as *headache*, *dizziness* and *diarrhoea* belong to some kind of higher-order concept without referring to them as being potential symptoms for a disease ? We can say they are all *abnormal conditions*, but also diseases are abnormal conditions. And what is then the difference between a disease and a symptom ? Issues like this can be discussed (and in fact are) for ages. The problem can be resolved when being 'pragmatical', but this again requires reference towards a certain context, and as such, purpose - independence remains a beautiful dream.

3. On the definition and meaning of concepts

An issue repeatedly encountered in the literature is whether or not a rigorous distinction should be made between a concept system, and the real world for which this concept system is intended to be some kind of representation of. This issue is closely connected with the notion of meaning, while meaning itself bears some relationships with definitions. ISO does not provide a definition for what is *the meaning of a concept* (not to be confused with *the meaning of a term*, which, of course, is the concept designated by the term), or how the meaning of a concept in a concept system can be captured. Classically, i.e. in the Aristotelian point of view [4], *meaning equals definition*. A concept is defined by a genus and a set of necessary and sufficient conditions that differentiate it from other species of the same genus. As such, intensional definitions may be interpreted as meanings. In Wittgenstein's view however, a concept is defined by a prototype, such that an object is an instance of a concept if it resembles the prototype of that concept more closely than the prototypes of other concepts in the concept system [5]. This view can be combined with Aristotle's view, as is done by Putnam [6]. First you define a concept through its essential properties, and second, you think of this definition as not only fixing the 'reference' of the concept (defining its Fregean 'sense') but functioning as a prototype. In the light of this approach, meaning cannot be separated from the real world. Also, this view contradicts partially the claim that conceptual knowledge (i.e. the structure of sensible medical concepts and their relationships independent of the language in which they are expressed) should be distinguished from the knowledge of clinical definitions and criteria (i.e. the knowledge to determine whether or not a concept applies to a particular patient) [7].
This issue on meaning is not only of language philosophical nature, but has practical implications in the elaboration of any concept system, as it can again be demonstrated by the hypertension example. What is the meaning of the concept *hypertension* ? We could define it as *a blood pressure that is too high*. If in a given concept system the concepts *blood pressure* and *too high* are defined, *hypertension* can unambiguously be represented as well. Although this definition is correct, very few clinicians will feel happy about it, just because the definition gives no clues on how to evaluate whether a given patient has hypertension or not. What indeed is the meaning of *too high* ? In Rector's view [7], the meaning of *too high* relies on criterial knowledge, and as such, should not be represented in the concept system. This is perfectly acceptable when the concept system is to be used for describing what symptoms patients with hypertension can present, or what kind of drugs should be given for specific pathological conditions. But as a vehicle for exchange of patient information between heterogeneous systems, it is not adequate. Here again, purpose - independence is lost !
However, research in terminological knowledge representation has shown that it is possible to work around this problem. Brachman for instance proposes the *structural inheritance networks* formalism, in which concepts come in two flavours [8, 9]. The meaning of *defined concepts* is completely determined by their description in the concept system using a generic hierarchy of concepts at the one hand and roles as potential relationships between concepts at the other hand. *Primitive concepts* are only partially described, i.e. some part of the concept is undefined in the concept system. But this does not prevent it from being used as part of the description of a

concept with complete definition, e.g. *nephrogenic hypertension*: hypertension caused by a renal disease. The undefined part stands for a description that is irrelevant for the use of the concept system.
A first interesting aspect of working with undefined parts is that there is a clear relationship with non-monotonic forms of reasoning or non-monotonic classification. A rule might state that *disease(x)* $\rightarrow$ *has_cause(x)*. This can be a standard rule. However, if we add to the antecedent *essential_hypertension(x)*, the previous conclusion does not follow anymore as, by definition, essential hypertension has no cause. These discrepancies are accounted for in non-monotonic reasoning.
It is also worth noting that the existence of undefined parts does not prevent a concept system to be formally defined by using a formal syntax and model-theoretic semantics. Nebel for instance introduces primitive components as a syntactic category to be used in the description of concepts [10]. The notion of undefined part of a concept could be used as a kind of warning that when the concept system is used for another purpose than the one for which it has been designed, some additional actions need to be undertaken.

4. Language, interest-relativity and concept systems

Concept systems should maintain a clear distinction between the concepts itself, and the terms used to designate the concepts. It is estimated that a concept system bearing this property could be used in the exchange of information across linguistic borders. Whether this is achievable, remains an open question up to know. From a cognitive perspective, elaborating a concept system in health care (as in any other specialised domain) comes down to organising universally experienced and observed facts into categories by means of natural language.
In the traditional formulation of what is called the *criteria-attribute model*, a category is defined by a fixed set of properties or features. These attributes are necessary and sufficient conditions for category membership, affording absolute predictability. Hence, class membership is an all-or-nothing affair. A category does not display any significant internal structure. This is partly explained by the fact that the membership-relation is *extensional*: a member belongs to a class independently from the description we use to identify it. This 'objectivist' perspective is strongly reflected in the structure of concept systems encapsulated as knowledge sources of information or expert systems in the field of Artificial Intelligence. Categories are characterised here independently of the understanding of the 'categoriser', and the system is objectively constructed by the manipulation of abstract symbols that are given meaning only via conventional correspondences with things in the external world. This view on knowledge representation advocates unambiguous and language-independent concept systems (or nomenclatures) in specialised domains such as medicine.
However, recent advances in cognitive linguistics seriously question the cognitive validity of a strict 'criteria-attribute' model of linguistic categorisation [11, 12]. A viable alternative advanced by the cognitivists is the *prototype* model as first introduced by Wittgenstein. In this model, a category is defined with reference to prototypes, i.e. schematised representations of typical instances. Entities that conform to this prototype are accepted as 'central' members of the class, non- conforming members may be assimilated to the category as 'peripheral' members provided that they are judged as being similar to the prototype in certain aspects. A class is structured internally by virtue of its organisation into central and peripheral members, and class membership is a matter of degree, reflecting the distance of a member to the prototype. Because there is no specific checklist of criterial features, membership in a category is not subject to absolute predictability. Whether an entity qualifies as a member depends completely on the judgement of the categoriser.
Fortunately, medicine is not a *worst case* domain. Despite a number of exceptions, the medical sublanguage tends to be monosemous (each term tending to have one particular meaning) and highly conventionalised. These properties allow medical knowledge engineers to elaborate conceptual models that are mainly characterised by prototypicality and are based on a large consensus. However, from the prototype model perspective, any conceptual model cannot be totally language independent as it necessarily reflects the judgement of a 'categoriser'.
So, in order to structure concepts in a formal system, one has to rely on the language that is used by experts to reason about these concepts. This position does not exempt concept system designers from serious methodological pitfalls. Let us suppose that a concept is not yet well defined during the process of developing (formalising) a concept system, and that different alternative expressions or terms are admittedly used for describing such a concept. One can then reasonably assume that the concept can be further subcategorised in a number of different concepts. For instance, when reasoning about human anatomy, the concept of *disrupture* of an anatomical structure very quickly comes in mind. If we then look at the various terms that can be used to speak about disruptures, we can come up with terms such as *fracture*, *dislocation*, *cut*, etc. Consequently, we will notice that *fracture* exclusively is used for bones, *dislocation* for joints and *cuts* for structures such as the skin or organs such as the liver or the spleen. In this case, we follow Rector e.a. when he says that *fracture of the leg* is sensible,

and *fracture of the eyebrow* isn't due to facts describing legs and eyebrows [7]. It is clear here that the concept disrupture is further subcategorised in English. Notice however that the hyponyms *fracture, dislocation* and *cut* do not have necessarily equivalent translations in other European languages. In other words, one can reasonably assume that other languages may subcategorise the concept of disrupture differently (for example by using an additional category). This hypothesis (that should be empirically validated by a contrastive lexical semantic study) advocates also the language dependent character of any conceptual system.

At other occasions, it is not entirely clear whether or not different terms denote different concepts, or whether or not utterances cover what Rector e.a. call *conventional knowledge* or *descriptive knowledge*, both being specific subtypes of conceptual knowledge [7]. Brachman excludes drastically descriptive knowledge from a terminological system [13]. His distinction can also be accounted for in terms of the distinction between monotonous and non-monotonous knowledge as has been explained earlier. Coming back to the *essential hypertension* example, statements such as *essential hypertension is a kind of hypertension without a cause*, is a definition, and as such terminological or conventional knowledge. *Diseases have causes*, is a descriptive statement. However, we can transform this statement by introducing the concepts *causable disease*, and *uncausable disease,* such that it becomes part of conventional knowledge. The fact that no single English word exists for a *causable disease* might make it suspicious whether this a good idea or not, but one can never be sure that in another language such a term does exist. Fortunately, it has been shown that at implementation level, by using a proper formalism, both views are completely compatible: the difference does not really exist [10]. But this observation will not prevent physicians for advocating one view or the other, such that giving an account for this distinction and its implications will remain one of the challenges for conceptual representations in health care.

Finally, we can question whether concepts such as *fracture*, *dislocation*, etc. really are needed in a given concept system, or whether it would suffice to just mention disruptures. Stated otherwise, does it always make sense to distinguish at a conceptual level between fractures, dislocations and cuts? It does if in a certain context something particular (and outside the definition of the concepts) needs to be said about each of them or when the difference between "fracturable", "cutable" and "dislocationable" things is needed for some other purpose. If this is not the case, we could just make abstraction of the existence of these concepts. Although this approach is opposed by Davidson [14] who claims that conceptualisation should make knowledge explicit, and not reduce or domesticate it, it would not do harm to a computer application if the conceptual knowledge only is to be used by some kind of reasoner for making inferences. But if information needs to be displayed to the user, system generated terms such as '*disrupture of the femur*' sounds odd. Here again, the purpose of the concept system plays a central role!

5. On the formal representation of concept systems

If concept systems are to be used by machines, they should be represented in a formalism that leaves no room for discussion on the exact meaning of their constituents. Meanwhile, some machinery (a reasoner) has to be developed that can make (terminological) inferences using the knowledge expressed in the concept system. Both the conceptual knowledge and the reasoner should have some particular properties. According to McCarthy and Hayes [15], a terminological formalism should be *epistemological adequate* (i.e. it should be possible to represent all the aspects of the world we are interested in), while at the same time *heuristically adequate* (i.e. it must be suitable to be used efficiently by a reasoner). Unfortunately, it is very hard to have these two requirements fulfilled at the same time. First Order Logic for instance is generally judged as a very expressive formalism, but it is only semi-decidable. As such, for time-critical applications, it cannot be used. With respect to terminological systems, things are even worse, at least theoretically. By reducing the co-NP-complete problem of deciding whether two nondeterministic finite state automatons that accept finite languages are equivalent to concept-equivalence in a terminological system, Nebel showed that subsumption in terminologies is also co-NP-complete [10]. The complexity of the problem is related to the depth of the terminology. As a consequence, it is feasible to design time-efficient reasoners for concept systems if one reduces the depth of the terminology. So, this leaves us with the question what criteria should be used to limit that depth. Again, purpose-dependent criteria will be the answer.

6. Conclusions

Nowadays it is generally accepted that concept systems in health care must be language- and purpose independent, and that they should be formally described in a powerful and expressive formalism on which computationally tractable algorithms can be applied. Our analysis of the relevant literature in the domains of medical informatics, computational linguistics and philosophy has shown that these requirements cannot be fulfilled at the same time. Purpose - independence seems to be the most problematic goal to achieve as orientation towards a purpose is required for (1) identifying what concepts should be represented, (2) deciding on what should be introduced in the concept system as a concept or as a role, (3) eliminating unnecessary complexity of the concept system's structure by avoiding unneeded subcategorisations, and (4) limiting the depth of the terminology in order to avoid the problems associated with the computational intractable property of formal terminological systems. The interest-relativity of conceptual systems is due to the fact that descriptions tend to have a particular explanatory role. When describing objects, answers to particular questions are implicitly given. What is accepted as an interesting answer, is usually a context-sensitive matter [16]. Language - independence also cannot completely be achieved as structuring the knowledge domain and building the concept system is a matter of thematic sublanguage analysis and of subcategorisation which itself only can be performed by using the information provided in a given language. In different languages, the same concept may be subcategorised on different criteria or features. Finally, there is a well-recognised trade-off between expressive power of the formalism, and computational complexity. Although these restrictions are well documented theoretically, in practice things might not be as dramatic as they appear. In medical terminology and standardisation, a reasonable amount of consensus has been reached in various subdomains. Therefore, a tentative final conclusion of this paper is not that research in formal terminological systems in medicine should come to an end as the requirements that are put forward are not realistic, but rather that the identification of the pitfalls in this field indicates that adequate concessions to the ultimate requirements can, and should be made.

References

[1] CEN/TC251. Directory of the European Standardisation Requirements for Health Care Informatics and Programme for the Development of Standards. Version 1.7, 1993-06-01, 1993.

[2] International Standards Organisation. Terminology - Vocabulary. ISO 1087: 1990 (E/F).

[3] Rector AL, Nowlan WA and Kay S. Foundations for an Electronic Medical Record, *Meth Inform Med* 1993, 30: 179-186.

[4] Aristotle. *The Categories*, *On Interpretation*, *Prior Analytics*, *Posterior Analytics*, *Topica*. Cambridge MA: Loeb Classical Library, Harvard University Press, 1975.

[5] Wittgenstein L. *Philosophical Investigations*. Oxford: Basil Blackwell, 1953.

[6] Putnam H. *Philosophical Papers*. Cambridge MA: Harvard University Press, 1975.

[7] Rector AL, Nowlan WA and Kay S. Conceptual Knowledge: the core of medical information systems. In: *Medinfo 92.* Lun KC et al. (eds). North Holland: Elsevier Science Publishers BV, 1992: 1420 - 1426.

[8] Brachman RJ. On the epistemological status of semantic networks. In: *Associative networks: Representation and use of knowledge by computers*. Findler NV (ed.). New York: Academic Press, 1979: 3-50.

[9] Brachman RJ, Schmolze JG. An overview of KL-ONE, *Cognitive Science* 1985, 9 (2): 171 - 216.

[10] Nebel B. Reasoning and Revision in Hybrid Representation Systems. In: *Lecture Notes in Artificial Intelligence 422.* Siekmann J (ed.). Berlin: Springer Verlag, 1989.

[11] Langacker R. *Foundations of Cognitive Grammar*. Standford: Standford University Press, , 1987.

[12] Lakoff G. Woman. *Fire and Dangerous Things. What categories reveal about the mind*. Chicago: University of Chicago Press, 1987.

[13] Brachman RJ, Gilbert VP and Levesque HJ. An essential hybrid reasoning system: knowledge and symbol level accounts in KRYPTON. In*: Proceedings of the 9th International Joint Conference on Artificial Intelligence*. Los Angeles, 1985: 532 - 539.

[14] Davidson D. Truth and Meaning. In: *Inquiries into Truth and Meaning*. Davidson D (ed.). Oxford: 1984.

[15] McCarthy J, Hayes PJ. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: *Machine Intelligence.* Meltzer B, Michie D (eds.). Edinburgh: Edinburgh University Press 1969: 463 - 502.

[16] Buekens F, Ceusters W, De Moor G. The Explanatory Role of Events in Causal and Temporal Reasoning in Medicine, *Met Inform Med* 1993, 32: 274 - 278.