Chapter 11
Clinical Terminology

11.1 Why Clinical Terminology Matters

When, in the fifteenth century, Gutenberg's invention of the movable type led to the mass production and dissemination of books and written information, language was still relatively unformalized. It took until the eighteenth century before the great dictionaries and nomenclatures such as Dr Johnson's English Dictionary and Linnaeus' biological taxonomy were produced. Medical terminology escaped formalization, leading to problems that are now recognized as significant risks to patient safety.

The lack of agreed medical terminology has been recognized as an issue for at least 250 years. There is even an old word, "nosology," to describe the development of medical terminology, but the need has increased with the use of computers. Modern nosologists call themselves terminologists.

Sciences such as biology and chemistry have an internationally agreed formal structure for their terminology. Every living organism has a generic and specific Latin name expressed within a comprehensive biological taxonomy, which in many ways anticipated the full understanding of the evolution of life. All chemical structures are expressed in internationally standardized ways. Medical terminology, in contrast, lacks any formal structure.

Scott Blois in his classic work *Information and Medicine*, showed how medical concepts can be classified into a series of connected levels. At one extreme are subatomic concepts, which are key to radiation physics and molecular biochemistry. At the other extreme are concepts relating to societies, impacting population health and wellbeing. In between are more levels relating to the structure and function of cells, organs, body systems, individuals, personal relationships and so on (Fig 11.1).

The day-to-day vocabulary of medicine relates to every level. Radio-therapists use subatomic particles, clinical chemists measure molecule concentrations, haematologists study blood cells, microbiologists grow bacteria, radiologists review images of internal structures and organs, physicians are concerned with abnormal body functions, psychiatrists with unusual behaviour and interpersonal relationships and epidemiologists and public health doctors study the spread of disease in populations (Blois 1984).

T. Benson, *Principles of Health Interoperability HL7 and SNOMED*, HI,
DOI 10.1007/978-1-84882-803-2_11,

Healthcare mixes multiple overlapping theories, each with its own sub-terminology. Any classification system is inevitably just one way of slicing up a very complex reality and is made more difficult because key medical concepts such as diseases are abstractions, defined using information from a variety of information levels; diseases are not objects which can be seen or touched.

The historical, eclectic, and ad hoc origins of medical terminology have encumbered anyone interested in healthcare with the need to learn a whole new language, replete with homonyms (where the same term means different things depending on context); synonyms (where there is more than one term for exactly the same concept); eponyms, named after people; acronyms; and abbreviations. Nobody, who has not learnt the eponym, can guess Hodgkin's disease(lymph node cancer), Bright's disease (kidney disease), and von Recklinghausen's disease (hereditary neurofibromatosis).

People use terms in the way that they and their immediate colleagues understand. Each user of a term assumes that everyone else understands precisely what he or she intends it to mean; over time, groups develop their own local dialect. Medical records staff can often identify the institution in which a doctor was trained from the way he or she uses certain terms.

Lewis Carrol expressed the same problem in an exchange between Alice and Humpty Dumpty in *Through the Looking Glass* (Carrol 1871):

> 'I don't know what you mean by "glory"' Alice said.
> Humpty Dumpty smiled contemptuously. 'Of course you don't – till I tell you. I meant "there's a nice knock-down argument for you!"'
> 'But "glory" doesn't mean "a nice knock-down argument"' Alice objected.
> 'When I use a word,' Humpty Dumpty said in a rather scornful tone, 'it means just what I chose it to mean – neither more nor less.'
> 'The question is,' said Alice, 'whether you can make words mean so many different things.'
> 'The question is,' said Humpty Dumpty, 'which is to be master – that's all.'

The representation of written information has become more and more specific over the centuries. The first way of representing information was by a picture or drawing, such as in Stone-Age cave paintings. The earliest writing was based on pictograms, such as Egyptian hieroglyphics and Chinese characters, but the need for cheap and quick writing materials led to the development of cruciform characters on wet clay blocks in Mesopotamia and the development of phonetic alphabets such as those of Greece and Rome. Modern English uses just 26 letters and 10 numerals.

11.2 Computers Need Codes

Computers hold information as sequences of binary bits and work by matching strings; they need precisely coded data. A computer can instantly check if two strings are the same, but if any difference is detected, it cannot tell whether that

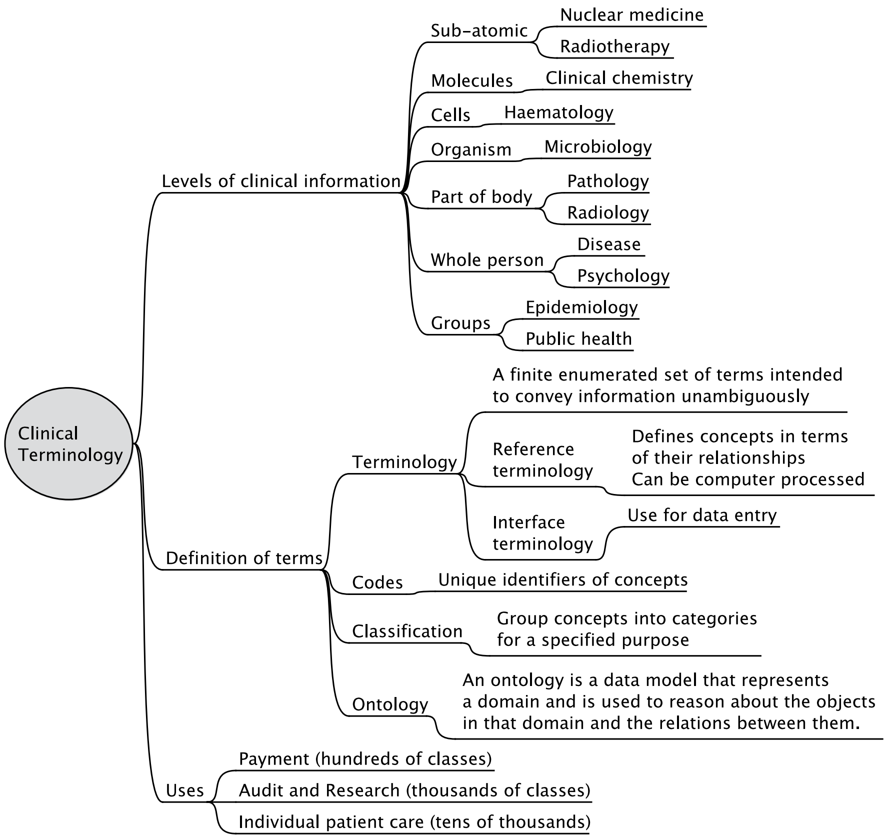

Fig. 11.1 Clinical terminology

difference is important. In spite of decades of effort, we do not yet have computers that cope well with the ambiguity inherent in natural language.

ASCII (American Standard Code for Information Interchange) was the code most widely used for representing the English alphabet in computers; it uses seven bits to provide 128 characters. ASCII was not finalized in 1967 and has since evolved into Unicode, which covers most of the world's writing systems. The ASCII coding system is important, because when lists are sorted in a computer, ASCII is the most natural order to sort them.

In ASCII (the American Standard Code for Information Interchange, which is the basis of the way that computers hold data), the numerals 0–9 precede upper case letters A–Z, which precede all of the lowercase letters a–z. Some printable characters, such as "!", "*," and ".", come ahead of the numerals; some, such as: "?" and "@," come between the numerals and uppercase letters; others, such as "\," "^," and "_," come between the upper and lower case letters; and a few, such as "|" and "~," come right at the end.

11.3 Coding and Classification

People often confuse the terms coding and classification and use them almost synonymously. This may be because the process of classification involves recording the codes used to name specific classes. However, coding schemes and classifications do different jobs. Classification allocates things into groups or classes, while coding is the allocation of identifiers, which can apply to anything (including classes in classifications).

A code is just sequence of symbols, usually digits or letters, which designate an object or concept for identification or selection purposes. It is just an alternative name for something, an identifier, designed for computer processing. Coding systems are an indispensible part of healthcare computer applications and specifications for exchanging data between computers.

The primary challenge for the designers of coding schemes is to produce something that will be widely and willingly adopted and endorsed by clinicians and managers. However, clinicians and managers have no more interest than a retail customer has in the bar code on a packet of cornflakes. Codes are needed and used by computers, not humans.

Clinicians need to record information in the form, language, and detail that is of most benefit to them when treating individual patients. Clinical records require precise and comprehensive detail about each individual patient, but statistical analysis requires patients to be classified into a relatively small number of discrete and mutually exclusive groups.

Clinicians and managers need to be interested in classification; classification is the basis for most statistical analysis, quantitative management, accountancy, and research.

Classification is the systematic placement of things or concepts into categories or classes, which share some common attribute, quality, or property. There is no limit to the number of ways that any set of objects can be classified.

The choice of what classification system to use is often determined by payment agencies, insurance companies, and national governments who control whether or not a doctor or institution gets paid or not. Such bodies usually specify the precise classification system that they require, often in collaboration with representatives from the professional and trade associations, medical colleges, and educational bodies. Once chosen it has to be accepted by users and implemented in computer software.

In *The Endangered Medical Record*, Vergil Slee and his colleagues argue that the choice of coding system used for electronic patient records represents a serious, real threat to the truthfulness and completeness of medical record content. Use of broad category codes, such as those specified by the International Classification of Diseases (ICD), rather than precise diagnoses means that we throw away detail, which should be preserved permanently. His plea is for detailed, permanent, and unambiguous codes (Slee et al. 2000).

For example, a trauma surgeon might describe a typical skiing accident as: a closed spiral fracture of the shaft of the right tibia with fractured fibula. In ICD-10, the code for fracture of shaft of tibia has the following logical structure:

Chapter XIX: Injury, poisoning, and certain other consequences of external cause (S00-T98)
Block: Injuries to the knee and lower leg (S80-S98)
S82: Fracture of lower leg, including ankle
S82.2 Fracture of shaft of tibia (with or without mention of fracture of fibula)
S82.2.1 Closed fracture of shaft of tibia

ICD-10 does not specify whether the leg is left or right, whether the fracture is simple, spiral, or compound or whether the fibula is also fractured.

Most healthcare computer systems use nationally prescribed coding systems such as the ICD, the Read Codes in UK and CPT-4. These all use a position-dependent hierarchical coding structure.

The internal structure of a position-dependent hierarchical code specifies its meaning relative to other codes. The structure of the code increases in detail from left to right, with the first character of the code specifying the chapter, the second the main subdivision, and so on until down the branches of the tree until the final leaf codes are reached.

One of the technical problems of hierarchical changes is that they cannot be modified easily without changing the meaning of codes in different versions, creating problems when, as inevitably happens, one version needs to be replaced by another.

A hierarchical classification can be thought of as an inverted tree with its trunk or root at the top. For example, biological classification places animals and plants into a hierarchical classification (a taxonomy) according to similarities in structure, origin, etc. that indicate a common relationship. The main levels in the hierarchy are Kingdom, Phylum (animals) or Division (plants), Class, Order, Family, Genus, and Species.

11.4 First Generation Clinical Coding Systems

Any coding system has various components.

- Concept: The fundamental idea is that of a concept. Each concept is identified by a concept code.
- Coding Scheme: Each concept code originates from a coding scheme. A coding scheme defines a set of concept codes, which are unique within the namespace of the coding scheme and are globally unique when coupled with the name of the coding scheme itself.
- Display Term: This is a human readable term. In some cases more than one display term may be provided for the same concept, to cover true synonyms, such as translations into different languages. One display term is usually designated as the preferred term.
- Relationship: Concepts may be related to other concept via a relationship, which allows the generation of hierarchical structures. One concept may be part of

more than one hierarchical structure. Often these relationships will be defined as part of original coding schemes, but other relationships are also possible.
- Value Set: This is a set of values that is allowed for a particular data item. Message specifications refer to value sets as the allowed values for a field.

Each concept is part of a coding scheme, but there is not a one-to-one relationship between value set and coding scheme.

The simplest code is a single code value, which needs to be unique for the field where it is used. The combination of field and concept code is unique. These codes need to be part of agreed code tables and there is no requirement to exchange the code meaning.

Simple codes are referenced using a value-set table, which has a heading which includes meta-data such as: value-set name, unique identifier, coding scheme, author, time validity, version, and other notes. Each entry in the table contains concept code value, display term, and notes about applicability.

Computer systems need unique identifiers, which have similar properties to codes. One way of achieving uniqueness is to treat each identifier as a pair, comprising a unique name for the assigner plus a value for the identification number, which is unique within assigner. It is the responsibility of the assigner to ensure that all such values are unique.

11.4.1 User Requirements

A key design requirement for any coding and classification system is to satisfy the needs of the different stakeholders.

Roger Côté views this as a pyramid with three levels of use:

1. At the tip, case-mix classifications such as DRGs, used for payment.
2. In the middle, classifications of diagnoses and procedures used to monitor and audit clinical activities.
3. At the base, clinical terminology used for individual patient care.

Healthcare managers and researchers need classified data, which enable comparisons and data exchange with existing data sources. Links between classifications must be explicit with one-to-one or many-to-one links. A many-to-one link involves loss of information, the extent of which is determined by how closely one classification is based on the other.

A multilevel classification with both course and fine granularity may allow two-way mapping from another classification. High levels of compatibility can usually be obtained only by basing a new classification directly on the target, using the same class boundaries. This requirement for cross-mapping with existing classifications inevitably drives the developers of clinical classifications to build on existing schemes, even if they are not suitable for the need in hand. For example, the ICD is organized around body systems, which is helpful in some circumstances, but not in others. Early versions of SNOMED reflected its origins as an extension of the Systematized Nomenclature of Pathology (SNOP) giving a forensic and medico-legal slant.

Doctors and nurses will not take the trouble to learn how to use any system unless it is quick and easy to use and provides information in the form and language that best helps them treat individual patients. Automatic or semi-automatic encoding software is required. Clinical records need to be as specific as possible. Hence, clinicians require a comprehensive nomenclature of medical terms covering everything that could occur within any patient's medical record. That is, all of clinical medicine and health service administration, but not the whole of biomedical science.

In 1984, the IMIA working conference on clinical terminology concluded:

> In future healthcare information systems, the user interface should be based upon natural language. The generation of numerical or alphanumeric codes should occur within the computer. Automatic encoding of natural language should be used. The morbidity and mortality statistical classification requirements of national and international groups should be the by-product of medically-based healthcare information systems.

It was not anticipated that clinical coding in hospitals would continue to be done by coding clerks.

11.4.2 Oxmis Problem Codes

Development of computing in primary care in the UK can be traced back to the formation of the Oxford Record Linkage Project in 1962, which connected birth, hospital data, and mortality data for a whole community and led directly to the Oxford Community Health Project, which maintained a central disease and encounter register for over 100 GPs. This required a coding scheme and Dr John Perry developed the Oxmis problem codes for this purpose.

The Oxmis codes were based on extending the four-digit ICD-8 classification for diseases and the OPCS classification of surgical operations. Prefixes were added to indicate F, family history; H, history from patient; K, surgical operation; L, locally defined codes; and T, items not within either ICD or OPCS classifications. Two-character alphabetic suffixes were added to provide greater specificity. Oxmis was one of the first multidimensional coding schemes, developed at around the same time as the first SNOMED.

Although no longer in use, the majority of GPs who used computers during the 1970s and 1980s in the UK used Oxmis, which also demonstrated that one person single-handedly could develop a coding system to meet the needs of many GPs.

11.4.3 The Read Codes

The Read Codes are widely used by GPs in the UK and New Zealand and are a direct predecessor of SNOMED CT.

The development of the Read Codes began in 1983, when, with colleagues James Read and David Markwell, I helped design a new computer system for use in general practice. An early design decision was to use a development tool that used fixed-length fields, requiring all codes, terms, and look-up keys to have a fixed predefined length.

The original design used alphanumeric codes with four characters (later extended to five characters) and terms up to 30 characters long. Another key requirement was that the coding scheme should be as comprehensive as possible covering everything that might be entered into a patient's computerized record. No existing coding scheme could be found which met these criteria so we chose to write one from scratch (as did many of our competitors during that period).

The motivation was commercial. We sought to commoditize GP computing, so that systems would be useful straight from the box. In earlier systems, GPs had to develop their own local coding schemes, which deterred prospective customers. It was also important to have a system that would be quick to use, so that GPs could use it themselves in the consulting room and could generate reports almost instantly.

Our idea was to take existing classifications and convert these into the appropriate format. The obvious candidates at the time were ICD-9 for diseases, the British National Formulary (BNF) for drugs, the International Classification of Procedures in Medicine (ICPM), and the national coding scheme for operations OPCS-4.

Dr James Read, a GP in Loughborough and one of my early customers, undertook the editing task and developed new sections for examination findings, preventive care, administrative procedures, and other subjects for which no suitable model could be found. Dr David Markwell developed the software. What was originally planned to take 3 months took almost 3 years and the scheme was finally launched as the Read Codes in 1986 (Read and Benson 1986). In the meantime, our customers continued to use homegrown codes.

As the work evolved, we found that we had improved on earlier classification and coding systems in several respects.

They were designed primarily for use by GPs in their surgery, not for epidemiology and international comparisons.

No paper version was ever published, facilitating regular updates and extensions.

The simple position-dependent, unidimensional hierarchy was easy to implement in software.

Each code has a uniform alphanumeric structure with five levels and 60 possible child codes at each node, by using numerals 0 to 9, upper and lowercase letters A–Z, with a couple of exceptions such as O and I that are easily confused with numeric digits, and a few printable characters such as "." (dot). This provides a large potential code space – 777 Million possible codes (60^5).

The first character of any code shows the chapter heading, the second the main subdivision, and so on. The following example shows the five-byte version (Version 2).

A…. Infectious and parasitic diseases
A1… Tuberculosis
A13.. Tuberculosis of the meninges and CNS
A130. Tuberculosis meningitis

Diseases were allocated codes with the first character in the range A–Z, so infectious diseases have codes starting with A, corresponding to the first chapter in ICD-9. The scheme was cross-mapped to ICD-9 and later to ICD-10.

Drugs were given codes starting with lower-case letters a–z, corresponding to the chapters in the first edition of the BNF.

Other concepts including occupations, history, and symptoms, examination findings, diagnostic and laboratory procedures, prevention, radiology/imaging, surgical operations, and administrative procedures were allocated first digits in the range 0–9.

The system included an index to enable quick semiautomated encoding by typing the first few letters of any term; the system responds with a list of terms to choose from, which may be context-specific.

The Read Codes combine the features of a classification and a coding scheme. However, hierarchical coding schemes can never be truly multipurpose, because they are built around a single hierarchical axis and each code is classified in one way only.

The Read Codes proved successful in General Practice, for which they were designed. However, attempts to use the original versions in hospitals proved impracticable, primarily because the simple hierarchical scheme could reflect only one view, namely the general practice perspective. Hospital doctors did not understand why information retrieval in one dimension was easy, but in another dimension was difficult and slow.

Once a concept has been placed in the classification, it is not practicable to move it, even if it has been placed in a location that is later regarded as wrong.

Another problem is the inherent multidimensionality of medicine. Tuberculosis meningitis is a type of tuberculosis, which is an infectious disease and is given code A130., but it is also an inflammatory disease of the central nervous system and has another code F004. It has two separate Read codes, creating code redundancy, which can cause inaccuracies in hierarchy-based analysis of clinical data, stored using the codes.

Being restricted to only four levels (later extended to five levels) in the hierarchy causes another problem. Consider the following hierarchy:

7…. Operations, procedures, sites
71… Endocrine system and breast operations
713.. Breast operations
7130. Total mastectomy operations
71304 Subcutaneous mastectomy

It is not possible to add a more detailed variant of this operation, such as subcutaneous mastectomy for gynecomastia in the appropriate position because there is no sixth level. A possible solution is to add it as a sibling alongside subcutaneous mastectomy in the fifth level with a code such as 71307. However, this creates the danger that when retrieving cases of subcutaneous mastectomy (71304), those recorded using 71307 would be missed.

The Read Codes were purchased by the Department of Health in April 1990, leading to the establishment of the NHS Centre for Coding and Classification (Chisholm 1990).

The NHS Clinical Terms project was started in 1992, as a major attempt to address these issues. The resulting scheme, which is known as Clinical Terms Version 3 (ctv3), was merged with the College of American Pathologist's SNOMED RT during 1999–2002 to create SNOMED CT (see the next Chapter). First we consider the history of SNOMED (Fig. 11.2).

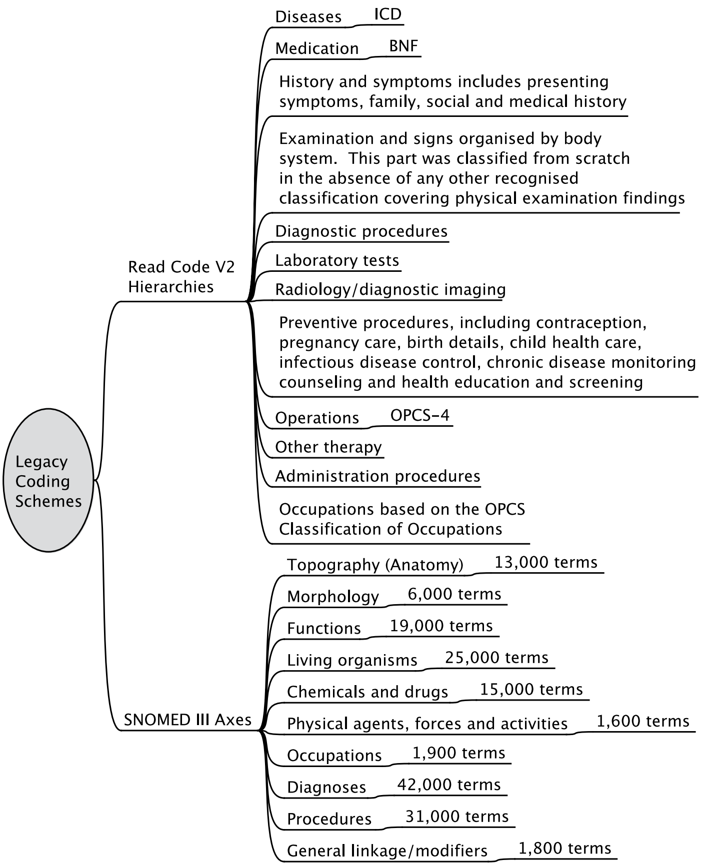

Fig. 11.2 Legacy coding schemes

11.4.4 SNOMED

SNOMED has a long history.

Back in 1955 the College of American Pathologists (CAP) established a committee to develop a nomenclature for anatomic pathology. In 1965, they published the *Systematized Nomenclature of Pathology* (SNOP), which describes pathology findings using four axes:

- Topography (anatomic site affected)
- Morphology (structural changes associated with disease)
- Etiology (the cause of disease) including organisms
- Function (physiologic alterations associated with disease).

SNOP was the first multi-axial coding system used in healthcare. By 1975 Roger Côté and colleagues had extended SNOP by adding additional dimensions covering diseases and procedures to give it a broader scope with the name *Systematised Nomenclature of Medicine* (SNOMED).

SNOMED was developed around a model of illness that started with normal structure (topography) and function. Sickness typically involves some abnormal function and abnormal structure (morphology). This has some cause (aetiology), which may be internal or external. Medicine seeks to reverse the process from the sick state to the healthy state by using administrative, diagnostic, and therapeutic procedures, which act on function or body structure. Disease was added to give easy mapping to ICD.

Occupations and organisms were added later. SNOMED III, published in 1993 had ten axes and 156,000 terms.

In 1999 the NHS and the College of American Pathologists (CAP) agreed to merge SNOMED with the NHS Clinical Terms Version 3 (also known as CTV3 and the Read Codes Version 3) to produce a single joint clinical terminology – SNOMED CT (Clinical Terminology). The merger was completed in 2002 with the first release of SNOMED CT.

While SNOMED has its origins in North American pathology laboratories, the Clinical Terms Version 3, formerly known as the Read Codes Version 3, was developed from the earlier versions of the Read Codes (see Chapter 11). SNOMED CT is a true merger. Every Read Code and previous SNOMED code ever released is present in SNOMED CT so migration to SNOMED CT should not result in loss of information.

In 2007, the International Health Terminology Standards Development Organization (IHTSDO) acquired all of the IPR of SNOMED (See Chapter 5).

Over a period of 40 years SNOMED has evolved from a pathology-centric terminology distributed and used in print format to a comprehensive clinical terminology, which is only available in electronic format and needs to be integrated with clinical applications software.

11.5 Desiderata

In 1997, Jim Cimino produced a paper, *Desiderata for Controlled Medical Vocabularies in the Twenty-First Century,* which brought together a number of common requirements for clinical terminologies, which had been developed in leading terminology projects such as GALEN (Rector et al. 1994), UMLS (Unified Medical Language System) (Lindberg et al. 1993), SNOMED RT (Reference Terminology) (Spackman et al. 1997), and the NHS Clinical Terms Project (O'Neil et al. 1995). This paper was particularly influential in influencing the design of SNOMED CT (Cimino 1998).

The desiderata are (Fig. 11.3):

- Vocabulary Content
- Concept Orientation
- Concept Permanence
- Non-Semantic Concept Identifiers

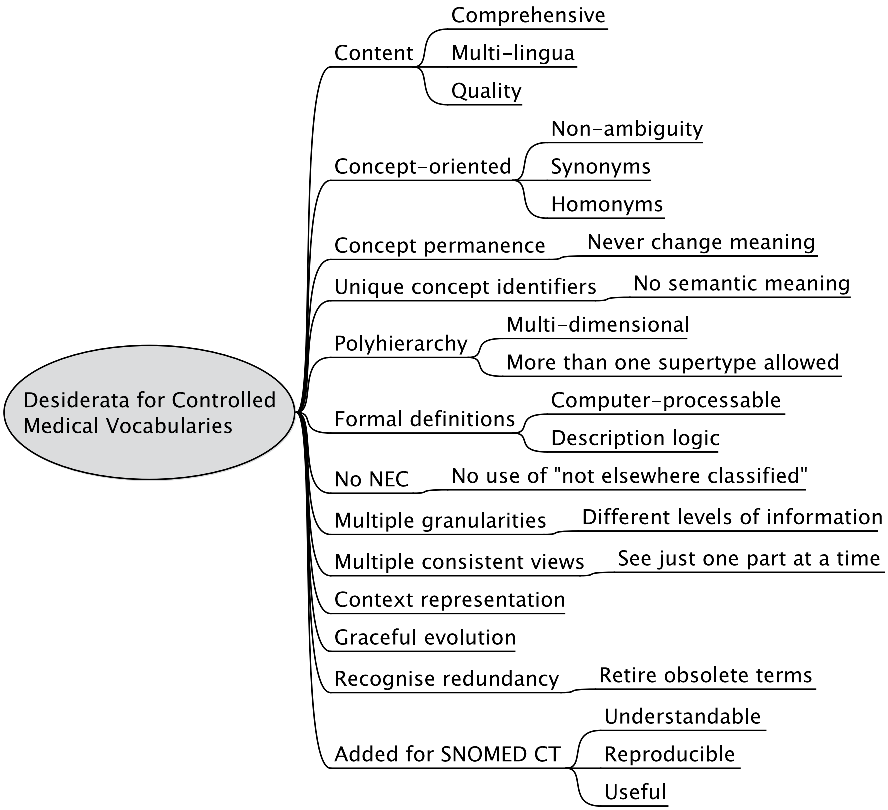

Fig. 11.3 Desiderata for clinical terminologies

- Polyhierarchy
- Formal Definitions
- Rejection of "Not Elsewhere Classified" Terms
- Multiple Granularities
- Multiple Consistent Views
- Context Representation
- Graceful Evolution
- Recognize Redundancy (Fig. 11.3)

Content, in terms of scope and quality, is paramount. Any practical clinical terminology needs to be comprehensive in terms of both domain coverage (concepts) and human readable terms (descriptions and synonyms). A methodology is required to allow the content to be expanded as and when required, including translation into other languages and dialects, while retaining high quality.

Concept orientation means that each concept term has one meaning (non-vagueness) and only one meaning (nonambiguity). A concept can be described by several terms (synonyms) in the same language and different terms in each language or dialect. Note also that one term can have several meanings (homonyms) each relating to a different concept.

Once a concept is created its meaning is persistent. It must not be changed or deleted by updates. However, a concept may be marked as retired where its meaning is found to be ambiguous, redundant, or otherwise incorrect.

Each concept should have a unique identifier, which should be meaningless. Semantic information should be handled as attributes of the concept, not as part of its identifier. Some of the problems with position-dependent hierarchical coding schemes have been reviewed above.

It is widely accepted that it is useful to organize medical concepts in a hierarchical way, but many clinical concepts are naturally multidimensional, with more than one super-type (parent) concept. For example, a fractured tibia is both a type of fracture and a type of leg injury.

The means of classifying a concept is independent of the means of identifying it. The development of formal, descriptive logic to define and classify clinical concepts is a major development away from the traditional position-dependent coding schemes and dictionary type of definition. For example, *pneumococcal pneumonia* may be defined using a hierarchical (*is a*) link to the concept *pneumonia* and a *caused by* link to the concept *streptococcus pneumoniae*.

Many existing classifications include one or more catch-all categories for concepts not covered. The problem with such *not elsewhere classified* or *NEC* categories is that they tend to change their meaning over time, as and when a new category is added which covers some of the NEC scope. The meaning is not permanent, which was a previous criterion.

Different users require different levels of granularity. Different levels of granularity are needed for defining concepts, navigation, decision support, and reporting. For example, a manager may only need to know that a patient has a broken leg; the finance department that it is a fractured tibia, but the clinician needs to know that it is a spiral fracture of the shaft of the right tibia. In principle, there should be no limitations on the number of levels in the display tree hierarchy.

When a concept has multiple hierarchical parents, the view of that concept should not depend on whether it was reached by following the hierarchy from a particular parent. Different requirements need different views on the terminology. The complete structure of a terminology, including all hierarchies and relationships can be complex and perhaps unusable. Each user needs to be able to see one or more views that reflect his or her own needs and understanding, but each of these views needs to be consistent with the underlying model.

Part of the problem of medical terminology is that information is usually recorded in a particular context and cannot be interpreted without that understanding. The context needs to be computer-processable. One approach to this problem is to provide a means of recording context explicitly within the terminology.

Terminologies change over time, which can create major problems for users if the meanings of aggregated time series data change in an uncontrolled manner. Care is needed to design the whole structure to support graceful evolution of concepts, terms, and relationships.

When terminologies change, some components will become redundant and so it is important to recognize explicitly that this has happened.

Three other criteria were added for SNOMED CT:

- Understandable: definitions should be understandable by average clinicians, given brief explanations.
- Reproducible: retrieval and representation of the same item should not vary according to the nature of the interface, user preferences, or the time of entry.
- Usable: we can ignore distinctions for which there is no use in healthcare.

11.5.1 The Meaning of Meaning

The Guidelines for Translation of SNOMED CT (Høy 2009) provide a short introduction to terminological principles, on which the following section is closely based.

The basic idea in the science of terminology is an onomasiological approach (concept-based approach) as opposed to a semasiological approach (term/word-based approach) that is applied in lexicography. With an onomasiological approach, the starting point is the concept; with a semasiological approach, the starting point is the linguistic expression, i.e., the word/term. A lexicographer would ask the question: "How many meanings could this term have/how many different concepts could be reflected by this term?" But a terminologist will ask: "Which terms could reflect this particular concept?"

The semasiological (term-based) approach will reveal the existence of homonyms and/or polysemes: homonyms are identical designations representing different concepts (e.g., *race* = taxonomic distinction of human beings and *race* = competition of speed). If the designations have the same origin they are referred to as polysemes (e.g., *bed* = piece of furniture and *bed* = the ground under a body of water).

The onomasiological (concept-based) approach will reveal the existence of synonymy: synonyms are different designations representing the same concept. The recommended approach to use is that applied within terminology, i.e., the onomasiological approach.

A concept is a "unit of knowledge created by a unique combination of characteristics." In other words, the concept corresponds to the image or idea created in our brains when we are presented with an object in our surroundings. The object may be physical, such as a car, or abstract, such as speed.

Any concept may be represented by a designation, which, in this context, would be a term. In other contexts, the designation could be a drawing or a photograph. The concept, i.e., the unit of knowledge/idea/thought, forms the connection between the object and the designation. This designation is defined as a "representation of a concept by a sign, which denotes it," and a term is the "verbal designation of a general concept in a specific subject field." The term will denote a concept, and a concept will refer to a particular object.

Traditionally, these principles are represented in the Ogden-Richards "Semiotic Triangle," which differentiates between three separate dimensions: the conceptual domain - thoughts that are in our minds; the symbolic domain - words and symbols that we use to communicate with others; and the real world - things in the real world that we refer to in our thoughts and with symbols (Ogden and Richards).

Whenever we are presented with an object, we automatically perceive its position in some kind of organized system – provided, of course, that experience tells us where it belongs. In the case of an abstract concept such as *democracy*, it would automatically be placed in our mind as a "type of government" – i.e., we would conceive it as belonging to the level below the concept *government*. Therefore, utilizing concept systems is useful in many contexts.

Concept systems allow us to place an unknown concept in a semantic context and give us a good idea of the importance, or, the "size" (magnitude) of specific concepts in relation to other concepts. Therefore, for didactic purposes, in connection with translation work as well as for storing and retrieving information in a systematic way, concept systems are extremely useful.

There are various principles that may be used when establishing a concept system: typology, partition, chronology, to name a few. The most common systems are based on generic relationships (IS-A -relationships) and partitive relationships (PART OF - relationships). In these systems, each concept belonging to the hierarchy is a TYPE OF and respectively a PART OF the immediate superordinate concept. In a generic system, a *metacarpal bone* could be considered as "a type of bone of hand," whereas in a partitive system, a *metacarpal bone* could be considered as a "part of the bone structure of hand."

In a generic system, the subordinate concept will be differentiated from its superordinate concept by means of at least one particular, distinguishing characteristic.

In a representation of a concept system, one will always find the generic concepts at the top levels and the more specific or "granular" concepts further down.

It is possible to establish "combined" concept systems that contain both generic and partitive relationships.

A definition is a "representation of a concept by a descriptive statement which serves to differentiate it from related concepts." To define a concept may take just a few words or it may entail a long phrase. Ideally, the definition will be based on the immediate superordinate concept in the concept system.

For example, a sedan could be defined as a "closed car having two or four doors and front and rear seats" (i.e., a type of *car*), and the *transmission* could be defined as a "set of mechanical parts in a car that transmits power from the engine to the wheels" (i.e., a part of *car*). In both cases, there is a reference to the immediate superordinate concept, and in the case of the generic definition, the particular, distinguishing characteristic(s) is/are added.

Instead of such narrative definitions, the "descriptive statements" which serve to differentiate concepts from one another may be expressed in description logic. This principle is applied in SNOMED CT where the concepts are defined by their hierarchical and defining attribute relationships.