Universal Terminology

Chapter 14: Definitions

This chapter is about definitions of entities of the domain of anatomy. TA98 was published without definitions of entities, speculating that all users share a common knowledge of anatomy and are able to define by themselves the entity they are manipulating. This is quite a common situation: most terminologies are published without definitions. The common reason for this absence of definitions is the lack of sufficient manpower. But we have a different opinion: the authors of past terminologies are not convinced about the need of quality definitions and they prefer to allocate their resources to other tasks.
We will try in this chapter to demonstrate the major role of good definitions in a terminology. Then we will develop a strategy for developing definitions. Finally we will open a window on the future where a formal model of definitions could automatically generate the expected definitions in several languages.
This document is the chapter 14 of the book Universal Terminology which presents a global documentation on the $\mathbf{T}_{logy}$.

Contents

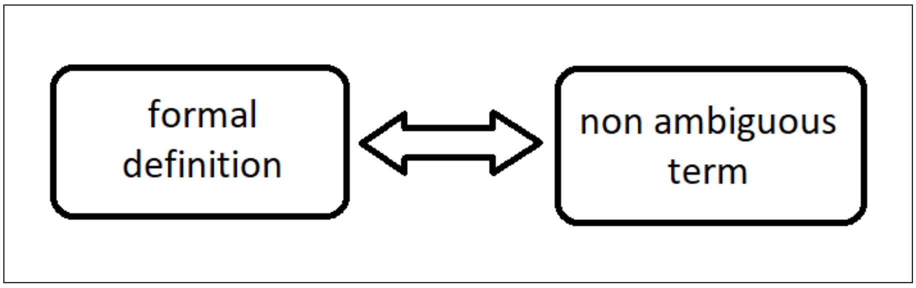

Figure 14.1: The duality of terms and definitions. Any weakness on one side is rapidly propagated to the other side. Formal definition means machine-readable definition.

14.1 The role of definitions

Let first consider a simplistic approach of the problem of definitions. One could say that it is not necessary to define what is the heart because everybody knows about this organ. The same is true for certain body parts that are known in the general population. By health practitioners one can imagine that a consensus definition does exist, possibly for 75 percent of all anatomical entities. Even if this number would be augmented to 90 percent, what about the rest of entities?

The role of the terminology is precisely to worry about those entities which are problematic because they are not well-known or ambiguous or coming with different interpretations, all this in a single population. And as soon as these criteria are examined relatively to different populations using different languages, the problem is increased.

Indeed there is a perfect duality between the term denoting an entity and a definition of this entity, see figure 14.1. What would be the use of a given term if we do not know precisely about what entity in reality we are speaking of? What would be the benefit of an exact definition if we do not have a good term allowing to communicate on this entity? The final precision is simultaneously proportional to the quality of the term and to its precise definition. The ultimate goal of the $\mathbf{T}_{logy}$is to satisfy the needs for clarity for the problematic entities, because the more common entities can be sufficiently controlled without a terminology!

Several authors have pointed the need for definitions.

[**Cimino, 1998**]
The list of desiderata contains as its sixth part the mention of formal definitions. This article shortly considers the possibility to make formal the definitions through the development of some model and it mention the need of sufficient manpower resources for any development of definitions.
The value of this paper is that it was published early and that it already mentioned the need of definitions among its basic desiderata.

[Michael et al., 2001]
This article is similar and complements the article of Rosse 2003 below. In addition, it presents a set of 10 relevant desiderata for writing definitions.

[Rosse and Mejino, 2003]
Citations from this article:
1) The disciplined modeling approach employed for the development of the FMA relies on a set of declared principles ... the Aristotelian definitions.
2) The purpose of definitions is to align all concepts in the domain in a coherent inheritance type hierarchy or taxonomy.
3) Paraphrasing Aristotle, the essence of an entity is constituted by two sets of defining attributes; one set, the genus, necessary to assign an entity to a class and the other set, the differentiae, necessary to distinguish the entity from other entities also assigned to the class.
This article supports the principal arguments that we have adopted for the $\boldsymbol{T}_{logy}$.

[Rosse and Mejino, 2007]
This article, in its section 4.4.5, summarizes the position of FMA about taxonomic definitions, to which we fully adhere. However, the FMA implementation of definitions is limited to a few hundreds entities after more than 20 years of development, and this considerably limit the validity of their proposal. And they do not foresee any automatic generation of definitions. See the appendix of this article for a list of definitions of the top level entities of the taxonomy.

The form of definitions is as large as there are authors of definitions in any domain. The most common definitions, to which any of us is accustomed, are the so call encyclopedic definitions, made exclusively of free text. In reality, it is more than a definition: it is a set of properties relevant to the object to be defined. Dictionaries and encyclopedia usually do not distinguish the definition from the properties: saying that *the kidney is an organ of 120 to 160 grams filtrating the blood and producing urine* is correct, but the weight of the kidney is a property not a definition, and the enumerated functions are not exhaustive. Indeed, the role of a definition is to distinguish an entity from another, in other words to differentiate the entity.

When building an ontology, our task must be delimited to the strict needs of this discipline. Because the manpower resources are limited (during creation, but also during maintenance), we must be scarce of any unnecessary effort: we do not want to rewrite another encyclopedia, not even a part of it. This immediately leads us to contingent the corpus of definitions by strong guiding principles. Such a constraining context is in particular obtained with the taxonomic definitions. In addition, it is based on the core of the ontology, the taxonomy of the domain.

14.2 Taxonomic or Aristotelian definitions

As said in [Rosse and Mejino, 2003], one can capture the essence of an anatomical entity under the form of two attributes: the genus and the differentia according to Aristotle principle.

We extract a citation from the above reference: "The purpose of definitions is to align all concepts in the ontology's domain in a coherent inheritance type hierarchy or taxonomy." This is precisely our orientation about definitions.

The genus is by definition the taxonomic father entity. But the father entity itself has a genus and so on, in such a way that the full genus is the set of taxonomic ancestors. In practice we limit the genus to the taxonomic father, but the essence of the definition may be to be searched for on other ancestors. See the examples below about this aspect of the taxonomic definition.

This means that the genus is totally constrained and can be computer generated at will. For example we have as first part of a taxonomic definition: "the kidney is a corticomedullary organ which ...". This text is easily computer generated. If we look at the upper ancestors, we get the following information: *parenchymatous organ* and *solid organ.* This later information will not be displayed in the definition, but it is available to the reader or a computer application.

The $\mathbf{T}_{logy}$has implemented the generation of the genus part of definitions. Because our implementation is prepared in 4 modern languages, the generated text is multilingual.

The differentia of a definition should be found in the corpus of properties which is available in the domain. But it must be limited to the documentation of the essence of the defined entity. Strictly speaking, the differentia must distinguish the defined entity from its siblings in the taxonomy. Nothing more is necessary. In other words, this means that some properties can be qualified of *defining properties.*

Some general rules are applicable for creating new definitions in the present context of taxonomic definitions. One rule is about avoidance of functional arguments, knowing that it is not always possible or at least desirable. We are in the domain of anatomy and the functional arguments are borrowed from the domain of physiology. The arguments issued from the morphology, histology or even embryology are to preferred.

Another guiding principle is to keep the definition simple and as easy to understand. Each definition, as much as possible, must be fully understood at the first reading by any casual user. But of course if a conflict exists between the precision of the definition and this above requirement, the precision has the priority.

Some definitions are trivial and may seem stupid. For example: *a gyrus orbitalis anterior is a gyrus orbitalis which is in anterior position.* There is no need to add anything more, this definition is good, precise, not ambiguous. What else? In fact we have a situation where the term itself contains the definition. This is far from being true in 90 percent of all entities. But it does exist.

A possible criticism of the taxonomic definitions is due to the fact that the imposed discipline and the formalism does not favor a user friendly didactic presentation. The risk is that we display the definition and the user discard it because it is too complicated and then, he refers to an encyclopedic defini-

A kidney is a corticomedullary organ [organum corticomedullare] which has as its part renal pelvis [pelvis renalis] shared with and connected to ureter [ureter]. (Foundational Model of Anatomy)

Un rein est un organe corticomédullaire [organum corticomedullare] qui a pour partie le bassin rénal [pelvis renalis] partagé avec et connecté à l'uretère [ureter].

Uno reno es uno órgano corticomedular [organum corticomedullare] que tiene como parte la pelvis renal [pelvis renalis] compartida y conectada al uréter [ureter].

почка - это корково-медуллярный орган [organum corticomedullare] который имеет в своей части почечную лоханку [pelvis renalis], общую и связанную с мочеточником [ureter].

Figure 14.2: The definition of kidney is available in 4 languages. The semi-automatic generation tool guarantee that the texts are equivalent in all languages: it is very important to provide precisely the same definition to all users of the $\mathbf{T}_{logy}$. The definition is built with two references to other entities: the *renal pelvis* and the *ureter*. On the website of the $\mathbf{T}_{logy}$, these references are hyperlinks to these entities, which themselves have a definition.

tion. This is certainly a part of the reality and it cannot be ignored. Anyway, encyclopedic definitions do exist at large and they are not to be considered as an exclusive alternative. On the contrary, why not mixing the sources, adding also on-line atlases of anatomy? It must be understood that the discipline of taxonomic definitions is just another facet of the anatomical terminology.

14.3 Examples of definitions

The idea of this section is to illustrate to guiding rules when creating new definitions by significant examples from the domain of anatomy. Each example is presented as a subsection and is coming with comments.

It should be noted that the corpus of definitions of the $\mathbf{T}_{logy}$is currently in a evaluation process and that a formal validation by experts of the domain would be organized in due time.

14.3.1 Cervical vertebra

Cervical vertebra
```
A cervical vertebra is a vertebra [vertebra] which
forms the upper part of the vertebral column [columna
vertebralis] and supports the head [caput].
```

This is classical definition based on the location. The second assertion concerning the head is not necessarily present and this is the choice of its author.

14.3.2 Vertebra

Vertebra
```
A vertebra is a bone of vertebral column [os columnae
vertebralis] which is the unit of construction of its
articulated part above the sacrum [os sacrum].
```

Here the definition is derived from the fact that the defined entity is a constituent parts of the whole. When going one step higher in the taxonomy, one learns that the vertebra is an irregular bone.

14.3.3 Decussatio pyramidalis

Decussatio pyramidalis
```
A decussation of pyramids is a decussation of neuraxis
[decussatio neuraxis ] which is found in the caudal
rhombencephalon [rhombencephalon caudale ] where the fibers
of the lateral corticospinal tract [tractus corticospinalis
lateralis ] cross from the ipsilateral to the contralateral
side of their origin.
```

This is a definition with two assertions: first the entity is localized in the caudal rhombencephalon, second the tract involved in the decussation is determined.

14.3.4 Thalamus

Thalamus
`A thalamus is a component of organ of neuraxis [componens organi neuraxis] which acts predominantly as a relay station for afferent sensory fibers and efferent motor fibers between the cerebral cortex [cortex cerebri] and the other segments of the neuraxis [neuraxis].`

The role of the taxonomy is important here: the thalamus is not an organ, but a component of neuraxis that is itself an organ. The FMA model is clear on this aspect. The definition is oriented on the white matter connections and insist on the main role of thalamus acting as a relay station between the cerebral cortex and the other component of the central nervous system.

A complement on the location of the thalamus could be added to this definition, something like *below the telencephalon.*This is not strictly speaking necessary, because the cerebral cortex is mentioned and further indication is optional. This is open to discussion.

This example shows the large liberty of the authors to select one aspect of this entity and to ignore other properties. Other authors insists on the location between telencephalon and midbrain or the proximity of the third ventricle.

14.3.5 Arteria basilaris

Arteria basilaris
`The basilar artery is a systemic artery [arteria systemica] which is formed by the confluence of the left vertebral artery [arteria vertebralis sinistra] and the right vertebral artery [arteria vertebralis dextra].`

This is quite a particular artery, being formed by an anastomosis of two branches of the subclavian artery. Indeed, it is not a branch of another artery. The definition is explicit on that situation.

14.4 Semi-formal definition

The differentia is more complicated to automatically generate and remains today dominated by free text. It starts with a verb at the third person and can be of any length. However, we try to restrict it to the task of differentiation of the specified entity related to all its siblings. In order not to be too formal and to facilitate the reading of the resulting definitions, we may accept some additional text based on properties or functions, but this must be an exception. On this basis we get for the above example: *Corticomedullary organ which has as its part renal pelvis shared with and connected to ureter* (source is FMA).

But, we observe that the differentia has two links to other anatomical entities: the EN:*renal pelvis* and the EN:*ureter*. It is relatively easy to locate these two entities in the free text and to replace them by the actual link to the entities, using their identifier. In addition, because the links gives access to the whole entity, we can automatically insert the Latin term. The result in 4 languages is visible on figure 14.2.

This approach is semi-formal, because a part of the differentia is automatized and the rest remains as free text. This is the current status of envelopment. We will present the future plans for creation of fully formalized definitions in a section below.

We have seen than the generation of the genus is automatized and that the reference to other entities in the differentia of the definition is also automatized. That's important steps, but the remaining tasks are difficult. When examining the form of the differentia in more than 1000 examples, we discover that it is made of a set of assertions in the form Relation + Entity, with a maximum of 3 assertions.

The next step of the automation process would be the modelization of the differentia. Our current evaluation is that the diversity of the texts is important and that a too simple model would not be adequate. The referenced objects in the differentia may be outside of the strict domain of anatomy and the number of defining properties to create is large. A extension of the model of the domain is necessary, providing new objects like structures, relations or functions, all present in a new branch of the taxonomy.

It is interesting to quote here a similar initiative, with the same conclusion: only a partial coverage of all definitions has been modeled. Full formal definitions have not been reached today.

[**Mungall, 2004**]
This article review an experiment by the author to capture the definitional part of the terms themselves and to use this information for validation purpose of the taxonomy. As expected, the success was real but partial because not all terms are directly significant. However this experiment shows that the modeling of anatomical texts is a source of knowledge to consider in the future.

14.5 Formal definition

There is no doubt: the ultimate goal of a modern ontology is to precisely name all objects of its concerned domain and to identify them in a machine-readable manner. This means that the terminological aspect must be completed by a strict definitional aspect. Moreover, definitions must be shared between ontologies: they have to be understandable by computers. As long as this goal has not been achieved, our ontology is only a partial work.

At the time where information is permanently shared between computer applications of quite different origins, there is major need for a mean to identify what we are speaking of and what we are communicating about. Without such an identification principle, no serious communication or sharing of information is possible. Today, when two applications are communicating, they simply assume that they share a common understanding of their constituting objects, but this assumption is not proved. In reality it is always partial and certainly far from the final expected situation. This is the fundamental limitation of computer communication. Human communication is also limited, but on the contrary of the computers, the humans may know when the limits are reached and they stop their exchange there, before major errors do occur.

As we have seen above, semi-formal definitions are already possible thanks to the availability of a taxonomy of the domain: the genus of the taxonomic definitions are automatically generated. Then, the problem is the automation of the generation of the differentia. Where do we have to search for additional information to be used as the background knowledge at the source for the generation of computer readable definitions?

14.5.1 A tank of properties

The first version of a terminology is usually naked and only the vertical hierarchies are present: the taxonomy and the partonomy. Other horizontal links between entities are poorly developed, if not missing. Let review the horizontal links already developed in the $\mathbf{T}_{logy}$.

At first, the vocabulary has been entirely formalized and entities using common words in their terms can be easily retrieved. Each entry of the vocabulary is a set of words sharing a common etymology and meaning. This set has two extensions: 1) by syntactic type of words, nouns, adjectives and prefixes; 2) by language used in the presentation of the terminology. In general, each term of the $\mathbf{T}_{logy}$is generated from a universal formula, language independent. The formula is made of pointers to the adequate vocabulary entry. A word in any language, even if it appears a few hundreds of times, is declared only once in the source terminology. This approach allows the easy display of all occurrences of a given word in the $\mathbf{T}_{logy}$like *testis*.

Secondly, the $\mathbf{T}_{logy}$has developed an expansion mechanism for the formulation of its terms: each term built from another term is explicitly declared and the link is made available for a computer application.

Let now have a regard to the vast field of properties of the anatomical entities. Indeed, the current state of the art of biological sciences has accumulated a sum of knowledge about these entities: most of them have several dozens of properties related to their position, their shape, their functions, their physical aspect, their typology, their dependencies, etc. This is what we call a tank of properties. In general, this knowledge is not formally represented, but is available thanks to the scientific literature and any other associated medias. Any casual anatomist is supposed to master this source of information, which represents his global understanding of anatomy. But all properties do not bring the same level of significance to our understanding of the domain. Some properties are immediately relevant, other are of low general interest, other are fundamental.

An indicated strategy is the following: do select some relevant properties as *defining properties.* This means that one, two or more properties are sufficient to build unequivocally the definitions. On a manual basis, this exercise has been already attempted with hundred of properties and has proved that in a majority of cases, no more than two properties are sufficient. Let consider an example going in this direction.

This example is a rather trivial definition: its term alone contains the definition, at least for a human reader, but certainly not for a computer reader! The core of this example is the relation is_face_of which must be formally defined. To do that, we first define the general relation is_surface_of which applies

selected entity:	THIS is LA:*facies lateralis testis*.
taxonomic ancestor:	THIS LA:*isa* LA:*regio superficialis testis*.
first property:	THIS LA:*is_face_of* LA:*testis*.
English pattern:	"is an oriented face of".
second property:	THIS LA:*has_direction* LA:*lateral*.
English pattern:	"in * direction".
formal definitions:	3276 1053; 3276 16482 3273; 3276 16483 10.
English definition:	a lateral face of testis is a superficial region of testis which is an oriented face of testis in a lateral direction.
French definition:	une face latérale de testicule est une région superficielle de testicule qui est une face orientée du testicule en direction latérale.
Spanish definition:	una cara lateral del testículo es una región superficial del testículo que es una cara orientada del testículo en dirección lateral.

Table 14.1: Example of definition. The * in a pattern is a placeholder for the generated value corresponding to the right entity in the property (by default, it is the rightmost position).

between a surface and a material entity. Then, our specific relation is defined as a taxonomic descendant with a property is_oriented with the value lateral. The whole definition is available either as a set of formal statements, or as textual sentences in all agreed languages. The formal definitions must be specified using an adequate mathematical formalism, in order that they can be transferred to foreign computers.

If we adopt this strategy of collecting specific relevant properties as the ground basis of definitions, we are naturally faced with an essential problem: how to move from properties as expressed in free text to a corpus of formal representation of the same properties? This is the main problem of knowledge representation.

14.5.2 Formal defining properties

As a working hypothesis, we have decided that binary relations of the type *A rel B* are sufficient to represent the necessary knowledge for building the differentiae of definitions. Possibly, unary relations could be considered. This hypothesis may not be sufficient on the long term and is at risk to be reconsidered at some point of development with the inclusion of more complex relations. But we have good reasons to think this hypothesis is sufficient and - second argument - we are not ready to manage more complex relations! This second argument is not scientific, but pragmatic: only positive results in this approach could justify it.

In the relation of the form *A rel B*, A represents the class of the entity to be defined. B represents either another anatomical entity, or a non-physical entity describing a structure present in the world of anatomy, like a duct, a tract, a morphological aspect, a form, etc. These later entities are explicitly created with the purpose to help preparing definitions. They represent a finite recurrent set of structures, each of which is reused several times. The *rel* itself may be any relevant link between A and B, but it is restricted by *B isa C*, with C being any entity of the taxonomy that is specified as part of the definition of *rel*. This allows to limit the use of this link to a specific subset of entities (the subset is due to the fact that *isa* is transitive). Any number of links may be created, but if an important reuse of links is not observed, our working hypothesis would be a failure.

14.5.3 Taxonomy of anatomical structures

An anatomical structure is a non-physical entity, which acts as model for the representation of the structures of anatomical entities. The anatomical structures are hierarchically organized in the taxonomy under the head entity LA:*structura corporea* or *bodily structure*. This entity extends on all material structures that are necessary for the description of the material entities of a human body. They are not body parts, but apply as model of the observable structures of the material anatomical entities.

The children for the bodily structure are the following:

1	**supporting structure:** The structures acting as support, protection or consolidation of entities within the human body.
2	**vector structure:** The structures acting for the transport of substances within the human body, either the material substances or the signals.
3	**operative structure:** The structures acting as producer of some substance or some motor actions.
4	**sensory structure:** The structures acting as detector of different signals.

Table 14.2: The different classes of structures.

14.5.4 Taxonomy of relations

The properties to be considered in the corpus of properties about the human body and its constituting entities are numerous, certainly a few hundreds. They must be identified and formalized. The identification is performed through a branch of non-physical entities of the taxonomy. This branch was already present for the identification of subparts of the *part_of* relation. It must be extended for all property relations.

The children for the relation top entity are the following:

The taxonomic relations are well-known in the design of the taxonomy. The main relation is the isa relation, which essential characteristic is the conservation of properties. A chapter of this book is devoted to the taxonomic relations.

The structural relations are devoted to the description of anatomical properties and they are the recipient figured above as the tank of properties. They are split into three subgroups for the partonomic relations, the spatial association relations and the anatomical structure relations.

The partonomic relations are all relations that subsume the part_of relation. They are numerous, depending on the type of entities they are related. A chapter of this book is devoted to the partonomic relations.

1	**taxonomic relation:** All relations for the specification of the taxonomy.
2	**structural relation:** All relations specifying the properties of anatomical entity.
2a	**partonomic relation:** All relations which are a specialization of the part_of relation.
2b	**spatial association relation:** All relations for the specification of a spatial association.
2c	**anatomical structure relation:** All relations specifying a model representing an anatomical structure.
3	**physical relation:** All relations specifying a physical property of anatomical entity.

Table 14.3: The different classes of relations. Any relation in use in the $\mathbf{T}_{logy}$is necessarily identified as a non-physical entity of the taxonomy. It is also formally defined elsewhere.

The spatial association relations deal with the spatial representation up to 3 dimensions of the anatomical entities.

The anatomical structure relations are the main subgroup with hundreds of items. They are the core of a model for anatomical entities, up to a detailed level of granularity. They contains the *supporting structures* like a bone or a membrane, the *transporting structures* like a tract or a duct, the *operative structures* like a muscle or ligament. An anatomical structure relation is designed for the representation of the different models of structure, which are realized on a large number of body parts.

For example, let consider a supporting structure, typically, a bone tubercle that is a bone projection, an eminence or an outgrowth present on a bone. This model of bone structure can be referenced by a relation like *A is_bone_tubercle_of B*, declaring that A is a realization of the tubercle model on some bone B, for example, LA:*tuberculum orbitale ossis zygomatici*, or explicitly *tuberculum orbitale is_bone_tubercle_of os zygomaticus.*

A variant of the anatomical structure relation is about the modelisation of pathways: a path in the human body for the transport of body substances. This is an example of a transporting structure. Typically, the LA:*gastrointestinal tract*. In general, such a structure is made of multiple *stages* in a defined direction: here there are four stages, the LA:*oesophagus*, LA:*gaster*, LA:*intestinum tenue* and the LA:*intestinum crassum*. This pathway transports the *bolus* from the mouth to the rectum. It seems natural to define the organs of the digestive system in relation to such functional pathways, with a property like *intestinum tenue is_distal_of gaster*, which is only understandable in the context of the gastrointestinal tract.

The physical relations are a separate group of relations, due to the fact that they are largely defined outside of the present terminology and they are valid in several domains.

14.5.5 Taxonomy of body substances

The body substances are present everywhere in a human body, but they are not parts of it. For this reason, they are generally absent of the past anatomical terminologies. In fact, they are closely linked to the human body or any of its anatomical entities by the relation *contained_in*, not to be confused with *part_of*. The body substances are necessary for a number of definitions. They need to be formally defined.

The children for the body substance are the following:

1	**material substances:** The substances, solid, liquid or gaseous, being entered, produced, transported or exited of the human body.
2	**immaterial signals:** The signals or energy which are transported between two locations, either the exterior of the body or any internal location

Table 14.4: The different classes of body substances.

14.5.6 Automated differentiae

The next step, given an entity for which a taxonomic definition must be created, having collected the *isa* link to its direct taxonomic ancestor for the genus of the definition, and the set of the defining properties under the form *A rel B* - possibly one, two or more properties - is the automated generation of the differentia in favor of the human readers. To do that, it is necessary to associate to each link a preset model of text for each language of the $\mathbf{T}_{logy}$. Such a model is part of the specification of the link.

For the transfer to another computer, the above mentioned links are sufficient. Of course the receiving computer has already collected the whole ontology, including the formal definitions of each *rel*. Such formal definitions have been prepared with care, following the guidelines of Smith2015.

Concerning the last example of table 14.1 above, the defining properties are the following:

1	facies lateralis testis isa regio superficialis testis
2	facies lateralis testis is_face_of testis
3	facies lateralis testis has_direction lateral
4	3276 10545 16413; 3276 16482 3273; 3276 16483 10

Table 14.5: The collection of defining properties. In this example, we display the three collected properties, all retrieved thanks to their common left entity *3276 facies lateralis testis*. Then the encoded form 4 of this set of properties is shown on a fourth line: this is the machine readable form of the definition. It is simply made of the identifiers of the entities and relations above.

The generation of the genus is performed thanks to 1 of the form A isa B. The natural language is based on the following schema: "a A is a B [] which ..." where [] displays the main Latin term of B.

The generation of the differentia is a sequence of contributions, one from each defining property that is present. It is assumed that most definitions can be adequately built as the concatenation of successive chunks of texts, this being true for all languages. This is again a working hypothesis to be confirmed by a successful implementation: if it does not work for a majority of all cases, a new strategy would be necessary. It is possible to think about two generations strategies, the second being activated in case of failure of the first. If this is true, we can continue with our working hypothesis, even if we currently have no idea about the second strategy.

Under our working hypothesis, our task is now the generation of a partial sentence for each available property regarding the definition. This task can be performed in total independence of the possible other properties. Let now consider the two properties in our example in turn.

The first property of the form A *rel* B links the property to be defined LA:*facies lateralis testis* to another entity, here the LA:*testis*. The relation *rel* is is_face_of and constitutes our leading argument. It must be constructed with two purposes in mind: 1) the generation of our expected chunk of text, 2) the formal definition of what it means in a computer readable form.

The text generation is based on a textual schema representing this relation. This schema is presently the following: "is an oriented face of". What we say is that we represent a face and that this face is oriented, in other word it has a principal direction, that will be precised by the second property. Such a schema must always be started with a verb at the third person of singular. By grouping the textual schema with the right member of the property gives us our awaited text: *is an oriented face of testis []*. The pair of square brackets is a placeholder for the Latin term of testis. This placeholder, when on a website, can be used as an hyperlink to the mentioned entity. It should be noted that our example is in English, but that in reality the work can be performed in any language of the $\mathbf{T}_{logy}$.

For the second property, the relation is has_direction, linking our entity to be defined to a value of direction. Here we get the value "lateral". The textual schema being "in * direction", we have to insert the value in the place of the asterisk, giving the final chunk of text "in lateral direction".

The expected differentia is now fully built, giving *is an oriented face of testis [] in lateral direction* and it can be appended to the available genus. This demonstrate the generation of the human readable definition, guaranteed to be in conformity to the formal information about the entity.

Finally, we consider the creation of a formal representation of each relation implied in the properties. This task must be done each time a new relation is created. This is a rather complex process partly based on mathematical formalisms, not necessarily mastered by all readers. We provide here the main steps without all the details. A complete presentation of these formalisms will be presented elsewhere.

The property 2 uses the relation *is_face_of*. It is applicable to a material entity only, here the testis. Any material entity can be linked to the volume it occupies with the relation *has_volume*, giving: *testis has_volume spatium testis*. Similarly, each volume is linked to its external surface with the relation *has_surface*, giving *spatium testis has_surface surface testis*. It suffices now to define the following relation *facies lateralis testis is_face_of surface testis*. It appears immediatly that the left argument is a subsurface of the right argument, with a restrictive condition. Such a condition says that all subparts of left argument are oriented in the same direction. All this formulation can be transformed in a mathematical formula. This finalizes our process to automatically generate the differentiae of our definitions.

14.5.7 Another example

Let now consider a more complex definition, involving four defining properties, for the LA:*testis*. Here four defining properties are necessary. First, the testis is the starting stage of the male spermatic tract, this tract involving all the organs defined under male internal genital organs. Second, the testis is the generator of the spermatozoids. Third, the testis is also the generator of the testosterone. Fourth, the location of the testis in the scrotum must be specified in the definition. This gives:

selected entity:	THIS is LA:*testis*.
taxonomic ancestor:	THIS LA:*isa* LA:*organum lobulare*.
first property:	THIS LA:*is_first_stage_of* LA:*tractus spermaticus masculinus*.
English pattern:	"is the initial stage of".
second property:	THIS LA:*is_generator_of* LA:*substantia spermatozoida*.
English pattern:	"producing".
third property:	THIS LA:*is_generator_of* LA:*substantia testosteron*.
English pattern:	"producing".
fourth property:	THIS LA:*is_contained_in* LA:*scrotum*.
English pattern:	"producing".
formal definitions:	3273 10545 10404; 3273 16482 16459; 3273 16487 16485; 3273 16487 16486; 3273 10566 3414.
English definition:	A testis is a lobular organ [organum lobulare] which is the initial stage of the spermatic pathway [tractus spermaticus masculinus] producing the spermatozoids [substantia spermatozoidi] and the testosterone [substantia testosteron] and contained in the scrotum [scrotum].
French definition:	Un testicule est un organe lobulaire [organum lobulare] qui est l'étape initiale de la voie spermatique [tractus spermaticus masculinus] produisant les spermatozoïdes [substantia spermatozoidi] et la testostérone [substantia testosteron] et localisé dans le scrotum [scrotum].
Spanish definition:	Un testículo es un órgano lobular [organum lobulare] que es la etapa inicial de la vía espermática [tractus spermaticus masculinus] que produce espermatozoides [substantia spermatozoidi] y testosterona [sustancia testosteron] y ubicada en el escroto [escroto].

Table 14.6: Example of definition. In this case, four defining properties are necessary, in order to precisely define what is the testis.

When examining this definition, it appears that not a single defining property can be dropped: each of the four properties is necessary. One may think that the two properties with the *is_generator_of* are not necessary, but in fact the global model must be able to say where any body substance is coming from. The last property about the localization in the scrotum is necessary, because otherwise no localization would be neither present nor deductible. About the first property, it is a significant contribution to the semi-functional approach that is actively developed for the definitions. The existence of the several passageways describes the logical circuitry in place in the human body and delivers the knowledge about the natural internal connections, where the body substances are moving.

In this definition, the LA:*tractus spermaticus masculinus* is invoked. This passageway describes the circuit of the spermatozoa in the male body. The spermatozoa are created in the testis, they are stored and maturated in the epididymis, they are moved through the ductus deferens up to the junction of the seminal gland where they receive a protective substance for their travel outside of the body, they continue through the ejaculatory duct internal to the prostate gland where they receive an alkaline substance, they continue in the male urethra and are finally ejected out of the body.

14.6 Discussion on definitions

This section is a discussion about the definitions, their role, their implementation, their complexity, etc.

14.6.1 Human versus computer

It has been stated that definitions should be understandable by humans and by computers. This is a very basic condition for a sound terminology. Of course, not the same information would be sent to the humans or the computers. In both situations, the sent messages must be adapted. However, the deep meaning about the domain of anatomy transported by the definitions in any form, must imperatively be the same. In other words, the form may be different but not the content. This desiderata is evidently to be also understood in a multilingual environment.

This target dependency rises several problems and it is certainly difficult to satisfy. A major problem is the difference of humans regarding computers to match the content of a definition to existing concepts in their mind using mechanisms which are poorly understood. Moreover, such a process by humans is not error prone and this is the responsibility of the authors of definitions to be sufficiently explicit and to avoid ambiguities. On the opposite side, the computers need quite extensive and precise details which may be considered by humans as unnecessary, pedant or embarrassing. The dilemma is there and the solution must be deep enough to overrule this gap.

Let consider a simple example of definition targeted to the humans: *the raphe of scrotum is a decussation which lies on the mid-line of the scrotum.* Here, the decussation is the taxonomic ancestor of the raphe of scrotum and as such is defined elsewhere. When one receives this definition, one is supposed to know what is precisely a LA:*decussation*. This means that the actual definition should not display any information which is available at another level. The same is true for the LA:*scrotum*, which is separately defined. Another comment is about the presence of the term *mid-line*: strictly speaking it is not necessary because there is only one raphe in the scrotum, making this information not necessary. But it is the initiative of the authors of the definitions to bring more information than what is strictly needed. We consider this definition as a good example satisfying the discipline of the taxonomic definitions: it is precise, it is short, it answers a human request.

What becomes this definition when the target is a computer? As usual with taxonomic definitions, the genus has no discussion and for the computer it is under the form of a relation: *raphe scrotum isa decussation* or in its coded form: *3415 10545 9649*. The computer knows that all properties of decussation apply to raphe scrotum. For the differentia, one needs a defining property expressing the fact that the raphe is mid-line on the scrotum. This can be done with a property like: *raphe scrotum is_midline_of scrotum* or in its coded form: *3415 xxx 3414*. Indeed, with such a property, one simply transfers the difficulty to the specification of this new relation *is_midline_of*. This relation must be defined as an inherent part of the terminology: it must be computer understandable and it must be translatable to any human language.

The relation *A is_midline_of B* is a spatial relation: it specifies that A is located in the mid-sagittal plane at the surface of B. In a detailed form, the plane in the scrotum at intersection of the mid-sagittal plane defines on the external surface of the scrotum a line, which is followed by the material raphe of scrotum. There is no doubt that the intersection of the mid-sagittal plane and the scrotum also exists on some internal surface of the scrotum, but we do not care about it! This is trivial for a human, but not for a computer. The external surface must be part of the specification of this relation, but this information should preferably be hidden to a human reader.

The formulation of the relation can be viewed in more details:

A is_midline_of B if:

- there is a sagittal subplane C defined as B is_intersection_of S, where S stands for the mid-sagittal plane,
- there is a line L defined as L external_boundary_of C,
- there is a property of A defined as A has_line L.

In this presentation, the used relations are all primitive spatial relations defined elsewhere.

What conclusions can we draw from this example? It appears clearly that the degree of details needed for a valuable computer representation is much more elaborated than our natural vision as human beings. The computer cannot manage an approximate view of the reality: it must look at all possible exceptions or unexpected irregularities, because they will occur soon or later, and if they are not anticipated, an error will impair the process.

However, at the cost of preparing a large collection of specific relations - probably a few hundreds for the whole terminology - the construction of a sound basis for automatically generated taxonomic definitions is feasible. It accounts for an acceptable effort in terms of manpower resources.

14.6.2 What the computer understand

The computer has three sources of information at disposal:

- the defining properties,
- the knowledge of the terminology,
- any outside general knowledge, encoded in the computer code or as data.

From that knowledge, and exclusively from that, the computer must understand the definitions and must be able to make logical deductions about anatomical entities. The third branch above is an open source of knowledge: we do not consider such a development and therefore we restrict ourselves to the first two sources. The defining properties have been defined above and represent a limited amount of knowledge. But the second axis is the adjustable part of anatomical knowledge. This second axis may be either limited to the strict necessary information, or extended at will with any collection of general properties.

The minimal knowledge to be transferred from the terminology to any foreign computer with the goal of understanding the definitions is about the relations used in the defining properties, as well as any other entities or relations used when specifying those initial relations. This bulk of knowledge is the core of the definitions. Even if a casual human user is not necessarily aware of this knowledge, it is assumed that he implicitly knows it. This assumption is from the authors of the definitions. Of course, if it is not true, the transfer of knowledge from the terminology to its users is impaired. This knowledge from the terminology is globally available as entries in the taxonomy.

The extension of the development of the knowledge of the terminology is certainly a goal of any specific terminology. Whether it is limited to the strict necessary information, with or without the knowledge about definitions, is a choice by the authors of the terminology. It must be said here that most terminologies are minimal and that even the knowledge for definitions is not available. In any case, the computer will only understand by what it has been fed: no miracle! And this must be added to the fact that most terminologies are dependent on one or two languages. In other words there are huge areas of development still open and partially explored (December 2023).

14.6.3 Text generation

This subsection examines how to generate free text definitions from the retained defining properties. The typical situation is a collection of two or three defining properties, from which the differentia of the taxonomic definition has to be built.

The differentia is always expected to be a single sentence started with a verb at the third person of singular, to be appended after the sentence of the genus. Asking for a single sentence is an arbitrary constrain, that has been decided because it gives a simple frame of presentation to all definitions and because it limits de facto the number of defining properties, which would give serious difficulties if too large. However, this constrain is certainly acceptable and not expected to weaken the quality or precision of the definitions. Most definitions should be realized with three or less defining properties and a few of them may come with four defining properties.

The common generating schema will be a sequence of generated texts appended one after the other, each being issued from a single defining property. The order of the generated texts is certainly not arbitrary. A weighting factor from the relations used for the defining properties should be sufficient to manage this aspect, but other strategies could be adopted if necessary. Generated texts issued from a pair of properties are not considered. Each generated text should be prepared as a text ready to be directly appended to the already generated part of the definition.

The text for a single property is essentially dependent on the relation used in the defining property. Each relation defined in the terminology is associated to a model of text expressing the meaning of this relation. This binding of a relation to a model of text must be rephrased in all languages of the terminology. The generated text is expected to be followed by the name of the right entity in the defining property. However, this name can be positioned inside the model text: an asterisk in the model text would act as a placeholder when the default rightmost position is changed.

When a generated text does not correspond to the first property, it must be coordinated with the previously generated text. This may be realized in different ways. The most common situation is the usage of the comma or the conjunction "and". But alternate solutions are possible.

Finally, syntactic constrain may be present related to gender and number in the differentia and must be solved. The use of articles must also be considered in some languages.

14.6.4 Implementing the definitions

The implementation of the definitions for the $\mathbf{T}_{logy}$is a huge task. There are more than 12'000 entities waiting for a definition and there are at least 4 languages as target languages of definitions. In 2023, nobody as ever fulfilled such a task, even a small part of it. Of course, these statements does not concern the editors of dictionaries and encyclopedia, that never consider to be computer readable.

In addition, we cannot pretend that our approach to the interchange of definitions between different computer applications is a final solution: we may have miss some important difficulties in this process of knowledge representation. But, whatever far we are from any workable solution, we are convinced that we are making significant progresses.

Hopefully, a large number of definitions - more than 80 percent - are either trivial or reproduce another definition with a small change. And the number of target languages is not a significant manpower problem, because the processing of languages is largely automatized in a similar fashion in all languages.

We propose today in 2024, an implementation in 3 steps:

Short term Realization of the formal definitions for a sample of 20 entities selected in different contexts.

Mid term Extension to an entire chapter of the $\mathbf{T}_{logy}$with 200 entities, with less than 10 percent of failures. Internal validation by selected experts.

Long term Planning and realizing the definitions for the whole terminology, involving a executive committee of anatomists for a validation process.

Bibliography

[Cimino, 1998] Cimino, J. (1998). Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med. 1998 November ; 37(4-5): 394–403*. PDF.

[Michael et al., 2001] Michael, J., Mejino, J., and Rosse, C. (2001). The role of definitions in biomedical concept representation. *In proceedings of AMIA 2001*. PDF.

[Mungall, 2004] Mungall, C. (2004). Obol: integrating language and meaning in bio-ontologies. *Comp Funct Genom 2004; 5: 509–520*. PDF.

[Rosse and Mejino, 2003] Rosse, C. and Mejino, J. (2003). A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics 36 (2003) 478–500*. PDF.

[Rosse and Mejino, 2007] Rosse, C. and Mejino, J. (2007). The foundational model of anatomy ontology. *Anatomy Ontologies for Bioinformatics: Principles and Practice, Springer, New York*. PDF.

14.7 Log of updates

23 Nov 2023 Adding the formal definitions.

04 Apr 2022 Creation of the file.

14.8 Credentials

This document is part of the book ”Universal Terminology” accompanying the website on Terminologia Anatomica. It expresses the vision of the authors of the $\mathbf{T}_{logy}$about the foundations of the science of ontology, supporting the here presented terminology. Despite it is as exact as possible, close to the reality of the database of the terminology and the surrounding software, approximations, errors and ambiguities are possible and should be considered as independent of their willingness and intents.

Identified comments about the content of the website and its presentation are welcome. An appropriate answer will be given when pertinent.

Authentic URL of this file: `https://ifaa.unifr.ch/Public/TNAEntryPage/help/Chap14.pdf`