

Universal Terminology

ANATOMY OF A TERM

This chapter examines the nature of an anatomical term. The question is: what is a term? It is certainly more than a string of characters in a list. An anatomical term has an internal structure which control insures of their regularity, as well as their appearance when visible in multiple languages.

An anatomical term is basically an ordered sequence of words usually separated by spaces. A term is necessarily related to an entry of the terminology and reciprocally any entry of the terminology has one or more term. Any term is commonly expressed in a vernacular language which syntax is documented elsewhere and which words belongs to a dictionary of that language. The sequence of words is governed by a set of structural rules issued from the core of the terminology, restricting the terms to a precise sublanguage and excluding the ambiguities generally attached to natural languages. The main goal of this chapter is to describe these structural rules and to introduce the basic principles governing the structure of the terms.

This document is the chapter 8 of the book Universal Terminology which presents a global documentation on the \mathbf{T}_{logy} .

Contents

8.1	Introduction	2
8.2	The quantitative dilemma	3
8.3	Network of terms	4
8.4	Multiple forms of a term	5
8.5	Overview of the syntax	8
8.6	Structure of terms	10
8.7	Vocabulary	11
8.8	Base part of a term	12
8.9	Expansions of a term	13
	8.9.1 Genitive expansion	14
	8.9.2 Adjective expansion	14
	8.9.3 Preposition expansion	15
	8.9.4 Optional expansion	15
8.10	Universal formulas	16
8.11	Epilogue	17
8.12	Log of updates	19
8.13	Credentials	19

8.1 Introduction

The reciprocal link between a term and an entry of the terminology must be made clear. The referred entry is a Unit of the terminology. It is a group of 1 to 5 anatomical entities sharing a unique term. We consider humerus, left humerus, right humerus and pair of humerus as 4 distinct anatomical entities. They altogether form a Unit and a single term is attributed to the Unit, not to the entities. Later, the variants of the term are computer generated.

This design decision to connect the terms to the units and not to the entities is unusual and new. Most actual terminologies, including TA98 attribute the terms directly to the entities. But this implies a multiplication of terms that is not compatible with a safe management of a modern terminology. The section below about the quantitative dilemma indicates why it is urgent to manage the terms without redundancies.

At large, a **terminology should be independant of any language**: this is a strong statement because most existing terminology are published in a single vernacular and are evidently depending on it. However, a terminology is the expression of some knowledge of its underlying domain and this knowledge is not language dependant! Let consider that the gap between the existing terminologies and a universal terminology can be fulfilled at the cost of additional developments, making finally all terminologies compatible with the single here described framework.

A language independant terminology should have a representation that is not based on any particular language. Even Latin is not neuter enough for an adequate representation. The terminology can be represented on the basis of specific classes of the taxonomy of the underlying domain. This results in additional entities call *vocabulary entities*. Each such entity potentially can be mapped to one or more words of any language where the terminology is intended to be expressed. On this basis, a term can be modeled to a universal formula using specific functions which arguments are identifiers pointing to the vocabulary entities defined by the taxonomy. This approach is the ground floor of a Universal Terminology of Anatomy, with the potential to resolve the multiple imprecisions and ambiguities observed in most terminologies currently available.

This chapter will consider the structure of an anatomical term, applicable to any regular term, weather it is an official term, a synonym or any related term. Those terms not in conformity to this structure will be call irregular terms, not part of the Universal Terminology. However, irregular terms may remains present in any vernacular presentation. The target ratio of regular terms in a given vernacular is expected to be above 95 percent. This compromise about irregular terms is necessary in order to accommodate numerous terms issued from the traditional language, as well as predominant usages whatever are their conformity to the rules.

It is clear that the present effort in order to promote regularity rules is necessary for the realization of a central reference model providing a solid ground on which to build the different vernacular versions. Nevertheless, each vernacular will keep its freedom to accept or reject any single term, and to replace a regular term by a traditional term or any variant derived from a local usage. It should be remembered here our prognosis that for most vernaculars 95 percent of regular terms will be directly accepted.

8.2 The quantitative dilemma

The question is : how many anatomical terms do exist ? This section documents this crucial aspect of an anatomical terminology. Starting from Terminologia Anatomica as published in 1998, we have 7500 entries. This figure has to be augmented for the following reasons:

More detailed terminology

The details of the entities are augmented giving new entities as subparts of the existing ones,

More explicit terminology

A number of entities were implicit and should become explicit. For example, there are 23 intervertebral syndesmoses and any single one must be identified as such. The imperative reason is that the terminology is used for the medical records and any single event should be potentially identified. The terminology is also used for statistics about patients.

Extension of the terminology

The terminology originally limited to gross anatomy is progressively extended to histology. Other specialties are also candidates for expansions. Embryology is currently a separate terminology, but a future junction is foreseeable. Therefore, the initial value of 7500 should be nearly doubled and we retain the value of 12500 entries.

Then, the taxonomy must be developed and by definition its size is equivalent to the number of initial entries. This makes the number of entries to be 25000.

Each entry of the terminology may have several terms attached to it : preferred term, official synonyms, eponyms, related terms. A mean value of 2 terms is a representative figure: many terms have no synonyms but a number of them have several synonyms. In addition, any term may be present in multiple form: short or expanded term, display or basic term, etc. The multiplication factor here is 1.5. This makes an overall factor of 3, giving the value of 75'000.

Most entries in the terminology denote in fact several anatomical entities: the left an right members, the pair entity or the set entity. In gross anatomy, we observe that 60 percent of entities are paired entities. This gives a factor of 3 to the number of terms, giving the value of 215'000, rounded to 200'000 terms.

This means that we may found 200'000 terms for the anatomical terminology presented in a single language. This figure is confirmed by the FMA that currently manages more than 100'000 terms and is also subject to multiplication factors as presented above.

Finally, the Universal Terminology should be delivered in any available vernacular, at least, in any language in which medicine is taught. This number is not really documented, but Europe alone has already 25 languages. Therefore, we arbitrarily decide that the number of candidate vernaculars is 50. This makes the number of terms to be 10'000'000 !

We will see in another chapter that the transcription of the Universal Terminology can be automated at the cost of acceptable manpower resources. Even if this perspective is yet far of being implemented at large, this makes this number of terms to be a reality at some future time. Actually, 4 languages have an automated translation process implemented: Latin, English, French and Spanish, already reaching half of the estimated target of 800'000 terms.

The evident conclusion of this section is that manual handling of terms is not manageable. It would require a not existing amount of manpower resources, and the overall liability of the resulting product would be insufficient compared to automated processes.

8.3 Network of terms

A term specifically documents a single unit of the terminology, but this is rarely as simple. For example the term *LA:sternum* is defined by a noun without further reference. But this is not a frequent situation.

For a large majority of all units, the **terms document a unit with references to other units**. In reality, the terminology is a network of interrelated terms. References are present from zero to four distinct references (see Figure 8.1). In addition, the referenced unit may itself present other references. Such inter-references are directly issued from the traditional languages where historically was anatomy developed. The links expressed by these references are not formal and poorly documented. They may express a part-of relation, a proximity relation, a functional relation, an analogy relation, etc. But even if the reader has some idea about a possible meaning, it cannot be systematically documented. For this reason, the inter-references are unusable for modeling the domain. But they are primordial for building the terms, defining their exact wording and insuring their exact translations in any vernacular.

If the mean number of inter-references per term is 1 (it is probably above), this means that half of the text of the terms is redundant. By simply making explicit the references, we divide by two the quantity of texts to be validated. This is of course true for all languages (currently five in 2022). We have diminished the burden of the maintenance by a factor 10! Is it possible to escape to this option?

A simple example is *LA:incisura clavicularis sterni*, where we have 2 references to other units. They are: *LA:clavicula* and *LA:sternum*, both being without further references. The reference to the *clavicula* is an adjectival reference and the reference to *sternum* is an optional reference. Each one is identified by the identifier of the pointed unit. The universal formula is N(12306),&A(919),&O(877) giving the following term:

What we say in this section is that the reality of a terminology is a network of inter-referenced terms. We must make explicit the two main benefits of this approach:

Coherence of translation between languages:

Because the terminology is intended to be translated in several languages, the common universal model insures a common structure in the specific vernaculars. The universal formulas are the source of universal terms which are expected to be accepted as official terms in more than 95 percent of all cases. All these terms are strictly equivalent and the users moving from one language to the other will never be surprised.

Each vernacular remains free not to accept all the universal terms and can readily presents its particular tradition and even scientific habits. This is not forbidden and even recommended within agreed limits. But even in

such situations, the universal term may be displayed on the side of an idiomatic term, and acts as a bridge towards other vernaculars.

Non redundancy of terms:

Each unit get a specific term that is carefully chosen and then recommended by the authors of the terminology. Now, when another term is expanded on this initial term, it is of prime importance to use the very same text. If not, we immediately create ambiguities which are one of the main weaknesses of some terminologies.

The universal formulas totally guarantee against the risk of redundancy: any term exists only once in the source database, and multiple and different occurrences result only of automated processes. Such processes, when fully integrated and tested, will be quite sure without errors. Syntactical text generation is rather simple for a computer!

8.4 Multiple forms of a term

There are two dimensions in the space of development of a universal term. As seen above, one dimension concerns the languages candidates to a translation. We must consider now a second dimension: the variants in the usage of a source universal term in the expression of an anatomical unit. Typically, terms must be expressed at nominative, genitive and possibly other cases; they appear as generic or specific terms, paired or not; they may be singular or plural; they may have a display form and a formal form in different contexts.

This variability of terms is evidently language dependent. It has to be adapted for each language, through language specific rules issued from the syntax of this language. Hopefully, this task is simplified in the domain of anatomy, because the terminology is using a real subset of any language at large. The result is that each *anatomical language* has a rather localized variability and vocabulary, that can be mastered.

For the implementation of the entire terminology, we have found a number of situations on need of a specific form derived from the universal formula. Table 8.1 shows all such situations. They will be reviewed thereafter, together with significant examples.

Universal terms:

A universal term is directly issued from a universal formula expressed in a specified language. It is dependent on the syntax of that language. By definition, a universal term is composed according to the formal grammar of terms, therefore it is regular.

Example: The following example has been selected with a unit where the 5 languages accept the universal formula.

Inherited terms:

An inherited term is a variant of a universal term, with a simple modification. It is regular.

Example: The term *LA:hemispherium cerebri* is a universal term, but in English it is an inherited term: *EN:cerebral hemisphere* because the genitive is specified for *cerebrum* and an adjective has been preferred in English. The genitive is chosen in the universal formula because *hemisphere part_of*

Term	Description
universal	the term directly issued from the universal formula
inherited	the term issued from the universal formula with simple modification
official	the term chosen to represent this unit, usually the universal term
synonym	a term chosen as official synonym for the unit
regular	a regular term specified for one language, to be used in place of a universal term
irregular	an irregular term specified for one language, to be used in place of a universal term
managed	an irregular term correctly managed by the terminology
discarded	a term refused for use as a synonym in a specific language
related	a non official term in use outside of the terminology for that unit
eponymous	a non official term acting as an eponym for that unit
formal	a formal presentation term, not for ordinary displays
short	any above term, without its optional expansion
full	any above term, with its optional expansion
set	any above term, at plural for a set entity
pair	any above term, paired for a pair or pset entity
lateral	any above term, for a lateral member of a pair

Table 8.1: All different forms of a term for different contexts of use. Multiple forms may be present simultaneously. All terms possibly exist in all languages.

cerebrum, with the genitive to be preferred in the presence of a partonomic link.

Official terms:

An official term is a term which has been chosen for the representation of the unit. It is the universal term in 95 percent of cases. Otherwise, it can be an inherited term, a regular term or an irregular term.

Example: The term EN:*sternal facet of clavicle (pair)* is an official term differing from the universal term which is EN:*sternal articular facet of clavicle (pair)*. In the English version, the adjective *articular* has been removed.

Synonym terms:

A synonym term is a variant of an official term acting for an official synonym. It can be regular or irregular like an official term. There are up to three levels of official synonyms.

Example: For the universal term LA:*pars thoracolumbalis* there is a second universal formula for an official synonym, giving LA:*pars sympathica*, EN:*sympathetic trunk* and FR:*tronc sympathique*. The official synonyms may also be issued from regular or irregular terms.

Regular terms:

A regular term is specified for one language only and may be selected as official term instead of the universal term. It is considered to be regular according to the formal grammar of terms.

Example: The term FR:*neurones en panier de l'hippocampe* does not use the proposed adjective *corbiforme*, but a prepositional locution is selected in place. A regular term must be specifically entered.

Example: The term FR:*corps amygdaloïde olfactif* does not correspond to the universal formula giving in Latin LA:*amygdala olfactoria* proposing two

words instead of three.

Irregular terms:

A irregular term is specified for one language only and may be selected as official term instead of the universal term. It is considered to be irregular.

Example: The term EN:*tuber cinereum* in English cannot be issued from the universal formula, that will propose an English text. But the authors of the terminology have preferred here to use the Latin term. It must be specified explicitly as an irregular term.

Managed terms:

A managed term is an irregular term which can be correctly handled by the automatic procedures generating the texts in multiple languages. They cause no damage (ambiguities, weaknesses, errors, etc) to the terminology but like irregular terms they are not welcome on the long term.

Example: The term LA:*ramus sinus cavernosi anterior* is an irregular term and it has been preferred to the proposed universal term LA:*ramus anterior sinus cavernosi*. The term is irregular because the adjective *anterior* is not related to the noun immediately preceding it, but to *ramus*. Such an irregularity is quite common in traditional Latin, but it brings some ambiguous texts! Here, the authors of the terminology have preferred the traditional term.

Discarded terms:

A term specified as synonym under the form of a universal formula and not accepted in a specific language.

Example:

Related terms:

A related term is any term not retained in the terminology, but thought to vehicle some interesting information on the present unit, useful for retrieval purpose. It will never be retained as official term.

Example: The term EN:*prevertebral ganglion* is also named *collateral ganglion* by FMA, but the authors of the terminology did not accept it as official synonym. For retrieval purpose, it is inserted as related term.

Eponymous terms:

A variant of the related terms, made from a proper name of some preponderant actor playing a role in the history of terminology in the past and still having a current usage in that language.

Example: The term EN:*vertebral nerve* is also known as *nerve of Cruveilhier*. In a KWIC index, the entity is also retrieved under the name *Cruveilhier*.

Formal terms:

A formal term is a term always at nominative singular, representing its unit in theoretically non ambiguous way. Such a term is not intended to be used for general presentation and in expected to remain in the background.

Example: The term EN:*rami viscerales thoracici (pair)* is also known by its formal name that is LA:*classis parium ramorum visceralium thoracici* and EN:*set of pairs of visceral branches of thorax*.

Short terms:

A short term is any term without its optional expansion.

Full terms:

A full term is any other term where an optional expansion is present. Such an expansion is generally presented in italics.

Example: The term LA:*rami laryngopharyngei* has the following full name, using its optional expansion LA:*rami laryngopharyngei ganglii cervicalis superioris*.

Set terms:

A set term is a term at plural for its nominative part. Such a term usually represents the specific or generic set entities. The plural form is automatically generated in the language of the unit.

Example: The term LA:*ganglion thoracicum* gives the following set term LA:*ganglia thoracica*. This is simply the same term at plural.

Pair terms:

A pair term represents a paired unit or a pset unit. It is created by the addition of the label (pair) on its right. The pair term is expected to represent the specific paired entities.

Example: The term FR:*lame spinale V (paire)* represents pair of two entities: FR:*lame spinale V gauche* and FR:*lame spinale V droite*. This French example is evidently valid in all languages.

Lateral terms:

The lateral term is one of the left or right member of a pair. The lateral form is automatically generated in the language of the unit by addition of the adjective left or right.

Example: The term EN:*posterior nucleus of lateral funiculus of spinal cord (pair)* gives for its left member EN:*posterior nucleus of left lateral funiculus of spinal cord*. There are three entities involved in this term and one of them should receive the lateral adjective. There is a strict rules without exception about the positioning of this adjective, valid in all languages.

8.5 Overview of the syntax

The syntax of terms is evidently dependent on the selected language. Considerable variations are observed between the existing vernaculars. It is hard to find general rules in Western languages, and that's nearly impossible if we consider also the Asian languages. However, the problems to be resolved by the syntax are relatively the same for all languages, even if the solutions are different. The problems at the level of the syntax will be examined thereafter, having in mind the fact that the syntactic solutions are left to each vernacular.

In this section, we will discuss about the path from the words being the building blocks up to the final term. This is to be considered as an initial step, preparing the ground for a formal description to come in the following sections.

A term is an ordered sequence of words separated by spaces. In principle, no punctuation is accepted in a term, but this statement does not fit with all languages (Russian as commas in the text). This is a rather intuitive description on need of further precision. At first, consider what is a term in a given language, whatever is this language. A term is a sequence of words of that language, the order being constrained by the syntax rules. Each vernacular has its own set

of syntax rules and consequently the order is different from one vernacular to another. For example, Latin, French and Spanish position a noun before the qualifying adjectives, and English and Russian make the opposite choice. We will see later that the universal formulas use the Latin order, but the generating procedures are able to rebuild any order from the specified Latin order.

Words are generally separated (at least in Western languages) one from the next by a space, except in the presence of compound words that are more or less frequent depending on the vernacular. The word like *EN:spinocerebelar* is a typical compound word in English and other vernaculars. Possibly in presence of prefixes, a specific vernacular may add a hyphen, but the general trend in modern languages is to suppress them. Some languages like German, multiplies the compound words. Anyway, for all considered vernaculars, on order of importance is significant especially when in presence of consecutive words of the same category, typically a sequence of adjectives.

Words are accepted or limited by their syntactical category. The main restriction is the non acceptance of verbs, except past and present participles that are assimilated to adjectives. Adverbs and conjunctions are also forbidden. Prepositions may be accepted in some conditions, but this aspect in dependent on further developments. Consequently, terms are essentially built with nouns, adjectives, the preposition required for genitive, a few invariants (numbers, acronyms, proper names, etc) and a few articles imposed by the syntax. This selection process is of major importance, because it defines a restricted sub language which is considerably easier to manage than any full size natural language. Moreover, this restriction on word categories is accompanied by a selection process on the syntax rules of the specified vernacular. This is a necessary step, to be repeated for each vernacular, in order to avoid either syntactically ambiguous terms or difficult to interpret terms.

Some languages do have syntactically ambiguous constructions, like the Latin term *LA:meatus nasi communis* (part of TA98) : to the syntax point of view, we do not know what is common, the meat or the nose. We can only interpret such a term when bringing some knowledge of the domain. If here the correct solution is straightforward, this is not always true. Another fuzzy situation is the inverted genitive in English, because apposition of two nouns is also possible, like *LA:musculus sphincter* : syntactically, there is no way to resolve such a situation. The examples are given for the English, because this text is written in English, but similar situations may be found in most languages.

Another unexpected but major difficulty comes from the insertion of Latin words mixed with vernacular words. This practice is relatively current in English and leads to ambiguous situations. There are several reasons for that, one them being to decide about the generation of plural terms. Let take the example of the *EN:renal plexus*. There are two of them, therefore the plural is pertinent : should we speak of the *EN:pair of renal plexus* or the *EN:pair of renal plexuses* ? Unless you prefer the astonishing solution found on the Internet *EN:pair of renal plexi* ! If *plexus* is a Latin word, like most Latin words intermixed in English terms, the plural is defined by the Latin rules and it is *plexus*. Consequently, *plexuses* and *plexi* are incorrect. But the usage in different English speaking regions may contradict such an affirmation. The reality is that we are in a no man's land where some rules for plural of nouns are missing, undocumented or, if they are, not shared by the entire community of English speakers.

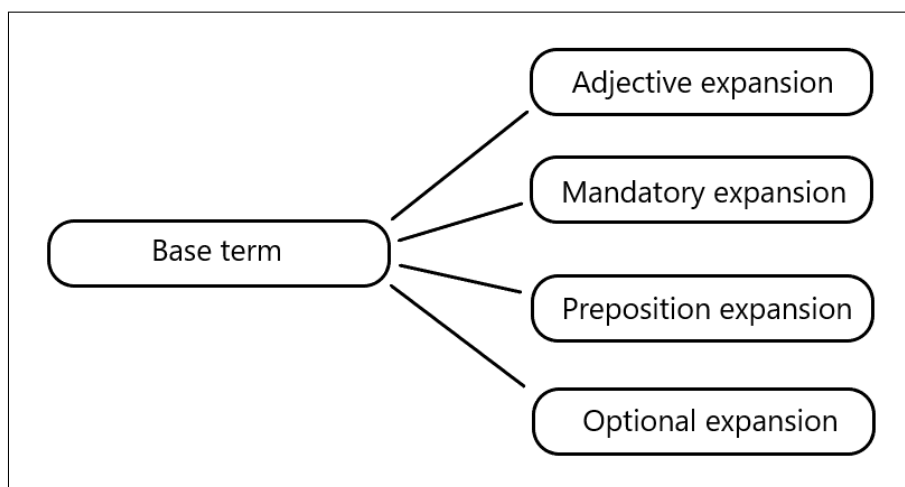


Figure 8.1: The structure of a term is illustrated here. The base term is mandatory and may be accompanied by zero to four expansions. Each expansion is specified by the identifier of the referenced unit. This structure is the basic building block of the network of terms.

8.6 Structure of terms

After fixing the syntactical context, we can now consider the structure of a term. It is made of a mandatory **base part**, that is a noun group, followed by zero to several **expansions**, which are links to other entities and are derived from their own terms. This simple organization is the skeleton of any term. Figure 8.1 illustrates this basic schema, to which all terms of the terminology must adhere.

Let present here a typical term with multiple expansions. We have selected a part of the *acoustic radiation*. It presents two expansions: one to the LA: *diencephalon*, and one to the LA: *acoustic radiation* itself.

The expansion to *diencephalon* allows to retrieve the adjective to be inserted in the term qualifying the main noun. The second expansion is an optional expansion necessary to qualify this entity, because other *diencephalic parts* do exist.

To figure out the importance of expansions, the approximate number of expansions has been estimated from the database of the terminology. It results that for the TNA, up to 40 percent of all terms have at least one expansion.

The structure of a base part followed by expansions is important and was not mandatory in previous terminologies. In particular, TA98 had multiple occurrences of a genitive inserted before the whole set of adjectives. The result is that the rightmost adjective is not related to the noun that precedes it. This tolerated exceptions leads to purely ambiguous terms, not acceptable in a modern terminology. For example consider the term LA: *fascia abdominis visceralis*. Here, the adjective *visceralis* is qualifying *fascia*, but the syntax is ambiguous and silent on that. The recommended solution is LA: *fascia visceralis abdominis* and would be implemented as the universal term. But the specific vernaculars remains free to adopt it or not!

The concept of multiple expansions implies that no anatomical entity can be

hidden within another entity: any related entity must be explicitly mentioned using its identifier and its name can never be redundantly reformulated. This is a fundamental aspect to guarantee the highest coherence of a modern terminology. Considering the past anatomical terminologies, there were developed without a database and major computer applications, not compatible with the resources available at that time. In particular, TA98 did not apply this basic principle.

Is it possible today to ignore the internal network of terms ? The evident answer is yes, because most of the anatomical terminologies are ignoring it. The most trivial reason is the fact that most terminologies are monolingual. A more technical reason is that the necessary linguistic techniques are not familiar to anatomists and they are only available since the beginning of the 21st century. But these techniques are relatively simple to implement and totally mature considering the sub language of anatomy. The domain of anatomy is evidently universal and we advocate a universal terminology.

In the present terminology, the expected number of base part is expected to be not greater than 25'000 items. Relatively to the expected number of terms of 200'000 terms, the ratio is 1 to 8, the rest being auto generated. This is a considerable gain in correspondence of the manual work of creation and maintenance of a traditional terminology. As already said, the benefit is first at the level of manpower, but predominantly at the level of the liability of the entire process : the less they are manual interventions, the less are the text errors.

8.7 Vocabulary

Before examining the structure of the base part, it is necessary to define what are the building blocks: what is the source of the words from which the terms are created ?

There are two approaches in order to make available the words of the domain of anatomy: either a dictionary is predefined and contains all the necessary information; or we have at disposal an open recipient ready to accept any new word when it appears. In the first solution, safety is insured by an initial acceptance and validation of any new word, but the missing words will block the system. In the second solution, any word is accepted whatever it is, with the risk of accepting non relevant items.

This last solution has been preferred in the implementation of the TNA, except for Latin where a dictionary has been created and a formal parser of this language is in action. The different solution adopted for Latin is explained by the fact that Latin as a dead language is relatively stable and well documented. The other modern languages are preferably viewed with the open approach because the task of building a dictionary for each new language is too important. Instead of a heavy initial set-up, we privileged an incremental acquisition of the syntactical knowledge.

The open solution is based on a **rule-based system**, having the potential to acquire the syntactical necessary information, at each presentation of a word. The rule-based system must be created again for each language, but it is a relatively simple task because a general model is available and only the specificity or the differences of a new language have to be reformulated. In other words, the formal theory-based approach is replaced by a try-correct strategy. The

proof of concept has been now given by the TNA implementation in 4 modern languages. The rule-based system is documented elsewhere.

In addition to this syntactical treatment, a corpus of significant words is created as a **concordance table** between all languages. This table includes all words entering the base part of any universal formula. For the whole TNA made of more than 5000 units, some 1200 words are used, giving the indication of one word for four terms as a mean value, each word being evidently used multiple times. The figure in the margin shows a small part of this table for the letter I.

The words of the concordance table are issued of the **vocabulary entities**. Such entities are authentic grains of the terminology as demonstrated on the chapter on granularity. An entity concerns a single word declined as a noun, an adjective and a prefix, any one present only when it exists in its language. Each entity is normally expressed in all languages, though a few missing words are possible. The words are always given at nominative singular masculine, whatever applies in a given language. As seen above, the rule-based system will take care of any variations to occur to any word in any language.

ilium 13957
impression 13851
incisive 13439
incus 13841
indusium 13095
infundibulum 12189
insula 12280
interneuron 11599
intestinum 12804
iris 12982
ischium 13959
isthmus 11718

Nouns with I

8.8 Base part of a term

The base part is essentially a noun group. A noun group is formed by a head noun, followed (or preceded) by a set of 0 to n adjectives or assimilated forms (like an ordinal adjective or a present participle). The head noun is most often mandatory and the rest is optional. The adjectives usually have a position from the more important to the less important, governing their display position in different vernaculars. A typical noun group is *LA:meatus acusticus externus cartilagineus* or *EN:cartilaginous external acoustic meatus*. In this example, the adjective *acousticus* is the more important.

But the reality is more complicated. The following situations are issued from the reality of an anatomical terminology:

Working with appositions:

An apposition is a sequence of two nouns at the same case, usually the second bringing a complement information to the first. Syntactically, both nouns always share the same case and the same number and they cannot be separated by other words. A typical apposition is *musculus sphincter*.

The need to accept appositions in the terminology is variable depending on languages and not well documented. In Latin, there are numerous apposition like *LA:musculus abductor ...* or *LA:musculus extensor* In English, the trend is to the suppression of these appositions, using a simplified language. In French and Spanish, the trend is probably to maintain all appositions related to the muscles, as for Latin. As a temporary conclusion, appositions must be maintained and will be part of universal terms.

An immediate effect of the existence of apposition is the exclusion of the inverted genitive in English, because there is no way to syntactically distinguish an inverted genitive from an apposition. But their meaning are quite different. The only way to have inverted genitive in English is by irregular terms or related terms. Anyway, the inverted genitive are always

ambiguous in presence of an adjective. What about the meaning of a ^{EN}*left sternum notch* that is syntactically perfectly correct !

Prefixed nouns and adjectives:

A consequent number of noun and adjectives accept a prefix as a modifier, resulting in a new noun or adjective. There are even doubly prefixed nouns or adjectives.

An adequate treatment of prefixes is necessary and adapted to the different languages.

Existence of invariants:

An invariant is an invariable word like a cardinal value (Arabic or Roman numerals) or an acronym. They are certainly not welcome in a terminology because frequently they are not universally accepted. The cardinal values are possibly misleading when one item in the list disappear: the Rexed lamina V is the fourth in the list today, because recent arguments conclude to the merge of laminae III and IV ! Anyway, cardinal are rarely significant. Roman numerals are not common in all world regions.

The syntax related to the invariants is not well defined. In our approach, we force the invariants to the last position of the base part in all languages. This is the most common situation, but is it acceptable for all languages?

Existence of noun complements:

The genitive in anatomical terms are generally due to the mandatory and optional expansions and they are numerous. But not all genitive are related to another anatomical entity. There are not frequently some general genitive like ^{LA}*fibrae associationis longae* but they do exist. Here, the word *association* is a noun complement.

The existence of a noun complement accompanied by an adjective has not yet been seen, and can be ignored. In Latin, the position of such genitive is immediately after the main noun that they modify. It may be different in other languages.

An adequate design for the acceptance of noun complement is necessary for all languages, though it is of minor priority.

Proper name complements:

The proper name known as eponyms are not officially part of the terminology, but it is well known that many anatomists are attached to proper names and they are expecting an adequate documentation on this topic. Is it necessary to remember that TA98 rejected the eponyms in its main lists, but published a full index of proper names with references to the concerned entities.

A modeling of eponyms necessitates the acceptance of proper names as noun complements.

8.9 Expansions of a term

The expansion of terms is performed with four sorts of expansions : with a genitive (mandatory and optional), with an adjective or with a preposition.

The main goal of expansion is to avoid to repeat the same term several times. A extensive review of all expansions is presented now.

8.9.1 Genitive expansion

The genitive (or mandatory) expansion is an expansion at genitive: the term referenced by the expansion is added at the right of the actual base term after having been transformed to the genitive singular. The referenced term is built with all its own expansions except a possible optional expansion. It has a nominative singular part that is the target of the transformation to genitive singular. The target is produced by the base part of the referenced term plus a possible adjective expansion.

The genitive expansion is intended at first to represent the *part_of* relation. This means that, if we have A as base term and B as expansion term, the relation A *part_of* B is expected to be true. This is due to the fact that the genitive expansion is expressed by a genitive term, and that the genitive is also call the possessive case. This only indicates that the genitive is the predominant expression of the partonomy. But, because of the versatility of the languages, this rule is only recommended and numerous exceptions will be accepted. Indeed, this rule is impossible to apply in all circumstances, and the tradition and the usage explain multiple exceptions.

8.9.2 Adjective expansion

The adjective expansion is a specialized form of the genitive expansion, when the referenced term may be represented by an adjective (or a prefix). This means that the terminology defines some vocabulary entities as representative of some anatomical entities. This is set of strict rules, linking one anatomical entity to one and only one vocabulary entity. Typically, the entity LA: *clavicula* is naturally represented by the adjective *clavicularis* in Latin.

On this basis, each time the adjective *clavicularis* is used in a term, it is not specified in the base term by a reference to the vocabulary entity, but it is specified in the adjective expansion by reference to the anatomical entity. During the processing of the term, the link from the anatomical entity to the vocabulary entity will be followed, the adjective will be retrieved and inserted immediatly after the noun in the base term. In this way this expansion is essentially a pointer to an entity of the terminology, as are all other expansions.

Sometimes, an adjective expansion is specified in a universal term and is not accepted in a given vernacular. Instead of that one prefer a genitive expansion. It is then possible to inherit from the universal formula, asking for the replacement of the adjective expansion by the genitive expansion. A typical example is with LA: *cortex cerebelli* or LA: *cerebellar cortex*: here the universal term ask for the genitive expansion, because this *cortex* is *part_of* the *cerebellum*, but in English one prefer the adjective expansion. This sort of interplay is necessary to accomodate the different languages, nevertheless issued of a universal terminology.

8.9.3 Preposition expansion

The preposition expansion is a project not yet developed. It consists of an expansion introduced by a preposition. Until it is implemented, the preposition expansion is replaced by the genitive expansion or by an irregular term, specifying explicitly the preposition, but this may imply that autogeneration of some terms like lateral members is not possible !

For example, the term EN:*branch to posterior limb of internal capsule* is realized with an irregular term, and the lateral member EN:*branch to posterior limb of left internal capsule* cannot be generated !

8.9.4 Optional expansion

An optional expansion is exactly the same that a genitive expansion in its definition, but is different in its usage. See the above definition under genitive expansion.

However, an optional expansion exist for the differentiation of terms which are similar after all expansions but the optional expansion. Some 20 LA:*tunica mucosa* do exist in the terminology. They will become different only by the adjunction of their optional expansions, giving LA:*tunica mucosa gastris*. The purpose of this expansion is that, when in a well defined context like the partonomy, the expansion part - here *gastris* - is not necessary, but when outside of any precise context, the expansion must be present. For this reasons, most partonomic lists are presented without the optional expansions, in order to simplify the presentation.

The terminology specifies that no term is allowed to have an homonym. In other words all anatomical entities are represented by a different term. This important statement is true only in the presence of the optional expansions.

This being said, it appears an important problem strictly related to the optional expansions, not really easy to resolve. The problems appears when generating the lateral members of units with an optional *left* or *right* adjective. The rule without exception defining this location is *the location is to the rightmost bilateral entity composing the term*. This universal rule is valid in all languages. This rule implies that the lateral adjective, once positioned on the term without optional expansion present, may have to be moved when adding the optional expansion.

An example of that paradox is LA:*equator* as a part of the *lens*. The short term of its left member is LA:*equator sinister* and the full term is LA:*equator lentis sinistri*. This gives in English EN:*left equator* and EN:*equator of left lens* and in French FR:*équateur gauche* and FR:*équateur de la lentille gauche*.

On the contrary of terminologies where each term is manually composed by a human, where the rule would be naturally applied (with some errors of course), our implementation of the terminology need to formally set the above rule and to apply it in all generating processes, what has been done. The solution is evidently to recompute the entire term, once without expansion and once with expansion. It cannot be built by simply adding the optional expansion to its right ! This is systematically applied in the terminology.

6017	N(11577)			6017	
6281	N(13882),A(13373)			6013	
6087	N(9526)			6013	6077
6278	N(13882),A(13372)			6013	
14221	N(13882),A(13389)		5991	6013	
11829	N(13882)		5991	6013	
7574	N(11695)		5769	6004	
11830	N(13882),A(13372)		6017	6004	
6086	N(9526)			5991	6077
5994	N(11835),A(13389)			5991	
5992	N(11925)			5991	5971
5999	N(11835),A(13398)			5991	
9511	N(13882),P(13052),A(11925)			5991	

Figure 8.2: An extraction of the source table of formulas in the database of the terminology. The columns are from left to right: UID, base part, prefix, adjective, genitive and optional expansions. The line with UID = 7574 corresponds to *LA: fibra thalamoparietalis*. The line with UID = 9511 corresponds to *LA: fasciculus frontalis orbitopolaris*.

8.10 Universal formulas

How to assemble a base part and a few expansions into a tangible formula to be stored in the database of the terminology ? What is the adequate representation of the core of the terms ? The problem is to figure out the drawing of figure 8.2 in a relational form.

The base part is made of a sequence of words controlled by the formal grammar of terms. Each word is of a specific type: noun, adjective, prefix and invariant and can be represented by a single corresponding letter, respectively: N, A, P and I. Each letter is a link to a vocabulary entity known by its identifier. By grouping a letter and the corresponding identifier in round parentheses and enumerating the sequence of words comma-separated, we obtain the formula for the base part.

For example the term *LA: fissura postclivalis* is made of a noun, a prefix and an adjective and gives the following formula: N(12774),P(11720),A(12828). The noun is pointing to the vocabulary entity 12774. If we look at the Spanish part of this entity we find the noun *fisura* or *fissure* for French.

The expansions are of four types each represented by a letter: genitive G, adjective Q and X, preposition R and optional F. Like the words, each expansion is pointing to a term that is represented by its identifier in round parentheses. The set of all expansions is again represented by a comma-separated sequence. Finally, the two sequences are appended separated by a comma. Figure 8.2 is extracted from the table of formulas in the database of the terminology.

For example the term *LA: glomerulus cerebelli* is made of a single word base part and an genitive expansion giving N(11604),G(5686).

For example the term *LA: fibra pontocerebellaris* is made of a single word base part and an adjective expansion doubled by a prefix giving N(11695),X(5266),Q(5686). The formula naturally gives the different language translations:

8.11 Epilogue

What did we learn in this important chapter? Indeed, we just received several lessons about the essence of a terminology, the terms. We must keep in mind all the mentioned aspects before any decision about future implementations. Here is a summary of this chapter:

Language independence

A terminology pretending to be universal cannot be based on a specific language, whatever it is: the terminology must be language independent. Though this statement is rarely applied in reality, it must act a guiding principle.

Latin has long been thought to be that support language, because it is neuter from the modern languages. It has a precise syntax and as a dead language it does not suffer from an uncontrolled usage. However, it presents a few ambiguities and it is not well accepted by countries where its teaching is low. The modern languages all suffer from bias and trends dictated by non scientific reasons, usually not under control. In addition, political arguments pros and cons are necessarily present. In this situation, the choice of a purely formal language representation, independent of any existing language, is an issue to this dilemma.

Quantitative aspect

It has been demonstrated from current data on anatomical terminologies that the actual volume of the task of building a modern terminology cannot be adequately handled by one or even several persons. In addition, the search for manpower for developing new terminologies is without budget. In other words, the authors of the terminology are working for free, but they have usually a University position or are retired.

Such a statement makes a computer based implementation a condition for success. The terminology must be implemented as a relational database and the interface from the database to a website or any other form of publication must be realized as programmed processes.

Multiple languages

The existence of the 5 languages version of TNA is the proof that multiple languages at exactly the same level of wording and the same precision is a feasible goal. The maintenance and current evolution of the terminology will be facilitated and immediately available in the all languages. Other languages may be added to the website at the cost of a reasonable effort.

Languages are built with words and they need to be recognized with their syntactical attributes. Nevertheless, the present treatment of languages is not dependent on predefined dictionaries. Indeed, it is preferable to work in an open environment where any new word is immediately accepted. Its syntactical properties are then defined by a rule-based approach specific to any vernacular. The result is highly positive - up to 99 percent - and the possible errors will be corrected later in a list of exceptions.

However, the corpus of vocabulary entities defines a central repository of words. A concordance table can be established in a two dimension table of words and languages.

Network of terms and expansions

The truth is that terms are made of terms and that there is an important inter dependency in the terminology under the form of a network of terms. Such a network must be made explicit and acts as quality insurance. Redundancies between terms must be avoided.

The edge of the network of terms are the expansions defined on a number of terms. There are four distinct types of expansions, serving all the needs of the actual terminology. A good knowledge of the expansions is necessary by the authors of the terminology. An adequate coordination should provide guidelines insuring a global coherence of developments.

Formal grammar

The fine structure of a term need a precise definition under the form of a formal grammar. Everything what is allowed or forbidden is written, acting as a guide for present and future authors of the terminology. Upgrades to the formal grammar are possible at any time, but they would need decision and agreement from the main actors.

The formal grammar is given in a technical document and its understanding is not necessary by the casual users of the terminology. The formal grammar is the chapter 9 of the general documentation.

8.12 Log of updates

31 Mar 2022 New formatting as standard chapter

4 Jan 2022 Creation of the file.

8.13 Credentials

This document is part of the General Documentation accompanying the website on Terminologia Anatomica. It expresses the vision of the authors of the terminology about its content and its form of presentation. 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/pdf/Chap08.pdf>