Universal Terminology

Anatomy of a term

This chapter examines the nature of an anatomical term. The question is: what is a term? It is much more than a simple string of characters in a list, edited in a word processor. An anatomical term has an internal structure that controls its regularity, appearance, form, and content in multiple languages.

An anatomical term is an ordered sequence of words, usually separated by spaces. A term is necessarily linked to a terminology entry, and conversely, every terminology entry contains one or more terms. Any term is commonly expressed in several vernacular languages whose syntax is documented elsewhere, and whose words belong to the chosen vocabulary of that language. The sequence of words is governed by a set of structural rules derived from the terminology, limiting terms to a specific sublanguage and excluding the ambiguities generally associated with natural languages. The main objective of this chapter is to describe these structural rules and introduce the fundamental principles governing term structure.

This document constitutes chapter 8 of the work *Universal Terminology*, which presents a comprehensive documentation of $\mathbf{T}_{logy}$.

Contents

8.1 Introduction . . . 2
8.2 The Quantitative Dilemma . . . 3
8.3 Network of terms . . . 4
8.4 Multiple forms of a term . . . 6
8.5 Overview of the syntax . . . 10
8.6 Structure of terms . . . 12
8.7 Vocabulary . . . 14
8.8 Base part of a term . . . 15
8.9 Term Expansion . . . 16
8.9.1 Mandatory Expansions . . . 16
8.9.2 Adjectival Expansion . . . 17
8.9.3 Preposition Expansion . . . 17
8.9.4 Optional Expansion . . . 18
8.10 Universal Formulas . . . 18
8.11 Epilogue . . . 20
8.12 Log of updates . . . 22
8.13 Credentials . . . 22

8.1 Introduction

The link between a term and a terminology entry must be made explicit. The referenced entry is a Def:**unit** of the terminology. It is a group of 1 to 5 Def:**anatomical entities** sharing a single term. For example, consider the humerus, the left humerus, the right humerus, and the pair of humeri as four distinct anatomical entities. Together, they form a paired unit to which a single term is assigned. The specific terms assigned to the entities are then generated programmatically. This aspect has been documented in *chapter 05*.

This design choice of relating terms to units and not to entities is unusual and novel. Most current terminologies assign terms directly to entities, or often do not distinguish between units and entities. This confusion leads to a proliferation of terms that is incompatible with the sound management of modern terminology. The section below, devoted to the quantitative dilemma, explains why it is urgent to manage terms economically and without unnecessary duplication.

Overall, a terminology must be independent of any language: this is a strong statement, because most existing terminologies are published in a single vernacular language and obviously depend on it. To reduce this dependence on a language, consideration has been given to publishing terminology based on a dead language, Latin, with the advantage that this language is no longer subject to evolution or any other dependence on the spoken language of a population. Although such a solution has undeniable advantages, the fact remains that Latin is not universally shared enough for it to be adopted by the scientific community.

The strong proposition remains that a terminology is the expression of knowledge of its underlying field, and that this knowledge does not depend on any language: anatomy is universal; it is neither English nor Chinese! The gap between existing terminologies and a universal terminology can be bridged with further development. This document shows how te $\mathbf{T}_{logy}$resolves this problem by giving a significant position to Latin, while remaining independent of any modern language.

A language-independent terminology must have an abstract representation independent of any language. Even Latin is not neutral enough for an adequate representation. Terminology can be represented using *vocabulary entities*, which are classes specific to the taxonomy of the domain in question. Each of these entities can potentially be associated with one or more words in the language in which the terminology is intended to be expressed. On this basis, a term can be modeled according to a universal formula using specific functions whose arguments are identifiers pointing to the defined vocabulary entities. This approach forms the basis of a universal terminology for anatomy, capable of resolving the multiple inaccuracies and ambiguities observed in most currently available terminologies.

This chapter examines the structure of an anatomical term, applicable to any term, whether it is an official term or a synonymous term. Terms that do not conform to this structure will be referred to as irregular terms. It is true that irregular terms may be admitted into any vernacular presentation according to specific needs, meeting the constraints of tradition and popular usage. These compromises regarding irregular terms are necessary to account for the many terms derived from the traditional language, as well as predominant usages, regardless of their conformity to the rules. However, the target ratio of

regular terms in a given vernacular expression should be greater than 95 %. The completion of the terminology attests that this objective has been achieved.

It is clear that the current effort to promote regularity rules is necessary to achieve a central reference model that provides a solid basis for constructing the various vernacular versions. Nevertheless, each vernacular language will retain its freedom to accept or reject a particular term, and to replace a regular term with a traditional term or any variant derived from local usage as often as necessary. Each language remains free to position itself more or less close to the universal reference. But at least one language that is far from the universal reference will not influence other languages.

In this regard, the position of the English language deserves comment. Some authors consider this language to be the modern scientific language, and it must be noted that the majority of scientific publications are published in English. However, the English language performs poorly in terms of its compliance with universal standards. This is due to at least two aspects of this language, which diminish its scientific precision.

Poorly controlled use of Latin

The English language, at least in the field of anatomy, uses Latin words included in the text to varying degrees. These Latin words retain part of the Latin syntax, such as the plural of nouns, and ignore another part, such as sometimes the genitive form. Added to this is the fact that many speakers have limited knowledge of Latin. A native speaker of English will have difficulty telling what the plural of plexus is!

Syntax simplification

English is easy to use thanks to a relatively simple syntax compared to other languages. The downside is that this simplification is a source of inaccuracies or ambiguities. For example, the inverted genitive quickly becomes problematic if one of the nouns receives an adjective.

Given advances in machine translation, the problem of scientific literature is set to disappear. In the future, there is no doubt that scientific articles will increasingly be written in the authors' own language. Publication will then take place in several languages.

8.2 The Quantitative Dilemma

The question is: how many anatomical terms are there? This section documents this crucial aspect of anatomical terminology.

Since the publication of Terminologia Anatomica in 1998, we have 7,500 entries. This number needs to be increased for the following reasons:

More detailed terminology

Entity details are being enriched, with new entities being added to existing ones.

More explicit terminology

A number of entities were implicit and must be made explicit. For example, there are 23 EN: *syndesmoses of vertebral column*, and each of them

must be identified as such. This was not the case in TA98. The imperative reason is that this terminology is used for medical records and any localized event must be precisely identifiable. Another reason is to enable various statistical approaches.

Extension of the terminology

The terminology, initially limited to gross anatomy, is gradually being extended to histology. Other specialties are also candidates for extensions. Embryology is currently a separate terminology, but a future merger is expected.

Therefore, the initial value of 7,500 should be almost doubled, and we will retain the value of 12,500 entries.

Then, the taxonomy must be added, and by definition, its size is equivalent to the number of initial entries in the partonomy, which doubles the total. The number of entries is therefore 25,000 terms.

Each entry in the terminology can be associated with several terms: universal term, official synonyms, eponymous terms, associated terms. An average value of 2 terms is a representative figure: many terms have no synonyms, but a number have several. Furthermore, any term can be present in several forms: short or expanded term, display term or formal term, etc. The multiplication factor here is 1.5. This gives an overall factor of 3, or a value of 75,000 terms.

Most terminology entries actually refer to multiple anatomical entities: the left and right limbs, a paired entity, or a set of entities. In general anatomy, we observe that 60 % of entities are paired entities. This multiplies the number of terms by 3, resulting in a value of 215,000, rounded to 200,000 terms.

This means that we could find 200,000 anatomical terms in a single language. This figure is confirmed by the FMA, which currently manages over 100,000 terms and is also subject to the multiplication factors presented above.

Finally, universal terminology should be taught in all available vernacular languages, at least in all languages used in medical education. This number is not actually documented, but Europe alone already has 25 languages. Therefore, we arbitrarily decide that the number of candidate vernacular languages is 50. This brings the total number of terms to 10 million!

We will see in another chapter that the transcription of *Universal Terminology* can be automated with adequate human resources. Although this prospect is still far from widespread, this number of terms could become a reality. Currently, four languages have a machine translation process: Latin, English, French, and Spanish, already reaching half of the estimated target of 800,000 terms.

The obvious conclusion from this section is that manual processing of terms is unrealistic. It would require human resources beyond what is convenient, and the overall reliability of the final product would be insufficient compared to automated processes.

8.3 Network of terms

A term specifically documents a single terminological unit, but it's rarely that simple. For example, the term EN:*sternum* is defined by a name without any other reference. But this case is rare.

In the vast majority of cases, terms describe a unit through references to other units. In reality, terminology is a network of interdependent terms. References are present in the form of zero to four distinct references (see figure 8.1). Moreover, the referenced unit may itself present other references. These references are directly derived from the traditional languages where anatomy historically developed. The relationships expressed by these references are informal and poorly documented: a part_of relationship, a proximity relationship, a functional relationship, an analogy relationship, etc. But even if the reader has an idea of a possible meaning, it cannot be systematically documented. For this reason, references are difficult to use for modeling the domain. However, they are essential for constructing terms, defining their exact formulation, and ensuring their accurate translation into any vernacular language.

If the average number of references per term is 1 (probably higher), this means that half of the term text is redundant. By simply making the references explicit, we halve the amount of text to validate. This is, of course, true for all languages. This is how the maintenance burden of the $\mathbf{T}_{logy}$has been reduced by a factor of about 10! Is it possible to avoid this option?

A simple example is EN:*clavicular notch of the sternum*, where we have two references to other units. These are: EN:*clavicle* and EN:*sternum*, both without additional references. The reference to the *clavicle* is an adjectival reference and the reference to the *sternum* is an optional reference. Each is identified by the identifier of the unit pointed to. The universal formula is N(12306),&X(919),&F(877) giving the following term:

→TAH879

LA:*incisura clavicularis sterni* EN:*clavicular notch of sternum* FR:*incisure claviculaire du sternum* ES:*incisura clavícular del esternón* RU:	Adjectival and optional expansion of a term. In fact, this term is built from three units, and this is repeated in each language.

What we are arguing in this section is that a terminology is actually a network of referenced terms. It is important to clarify the two main advantages of this approach:

Translation consistency across languages:

Since the terminology is intended to be translated into multiple languages, the universal model ensures a common structure in specific vernaculars. Universal formulas are the source of universal terms that should be accepted as official terms in more than 95 % of cases. All these terms are strictly equivalent, and users switching from one language to another will never be surprised.

Each vernacular language remains free to not accept all universal terms and can easily exhibit its own traditional terms, even its own scientific habits. This is neither prohibited nor even recommended within agreed limits. But even in this case, the universal term can be displayed alongside an idiomatic term and serve as a bridge to other vernacular languages.

Non-redundancy of terms:

Each unit has a specific term, carefully chosen and then recommended by the authors of the terminology. However, when another term is developed from this initial term, it is essential to use the same text. Otherwise, ambiguities are immediately created, which is one of the main weaknesses of some terminologies.

Universal formulas completely guarantee non-redundancy: a term generally exists only once in the source database, and multiple and different occurrences are the result of automated processes. Such processes, once fully integrated and tested, will be perfectly secure and error-free. Generating syntactic text is rather simple for a computer!

8.4 Multiple forms of a term

The space of expansions of a universal term has two dimensions. As previously indicated, the first concerns the languages that are candidates for translation. A second dimension should now be considered: the variant usages of a universal term source in the expression of an anatomical unit. In general, terms must be expressed in the nominative, genitive, and possibly other cases; they appear as generic or specific terms, pairs, composite or not; they can be singular or plural; they can have a display form and a formal form in different contexts.

This variability of terms obviously depends on the language. It must be adapted to each language, using specific rules derived from its syntax. Fortunately, this task is simplified in the field of anatomy, because the terminology uses a simplified subset of any language and its syntax. As a result, each *anatomical language* has a rather localized variability and vocabulary, which are easy to master.

To implement the entire terminology, we have identified several situations requiring a specific form derived from the universal formula. The table 8.2 lists all these situations. They will be examined below, along with relevant examples.

Universal term
`A universal term is (definition) a term directly derived from a universal formula expressed in a given language.`

A universal term depends on the syntax of that language. By definition, a universal term is composed according to the formal grammar as presented in *chapter 09*: it is therefore regular.

Example: The following example was selected with a unit where all five languages accept the universal formula.

Inherited term
`An inherited term is (definition) a variant of a universal term with a single modification in a specific language.`

An inherited term exists only in the language in which it is formulated. However, the same universal formula can be inherited multiple times in different languages. By definition, an inherited term 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 was preferred in English. The genitive

Term	Description
Universal	term derived directly from the universal formula
Inherited	term derived from the universal formula with a simple modification
Official	term chosen to represent this unit, usually the universal term
Synonym	term chosen as the official synonym for the unit
Irregular	irregular term specified for a language, to be used in place of a universal term
Deleted	term rejected as a synonym in a specific language
Associated	unofficial term used outside the terminology of this unit
Eponym	unofficial term serving as an eponym for this unit
Formal	formal presentation term, not intended for ordinary presentations
Short	any term above, without its optional extension
Full	any term above, with its optional extension
Set	any term above, in the plural for a set entity
Pair	any term above, paired for a pair or pset entity
Lateral	any term above, for a lateral member of a pair

Table 8.1: All forms of a term correspond to different contexts of use. Several forms can occur simultaneously. All terms probably exist in all languages.

→TAH13886

LA: *membrana intercostalis externa costae secundae* EN: *external intercostal membrane of second rib* FR: *membrane intercostale externe de la seconde côte* ES: *membrana intercostal externa de la segunda costilla* RU:	The universal formula is applied in all languages. A distinction is made between the use of a definite article in French and Spanish. Apart from this difference, we have the guarantee of the equivalence of terms in each language.

is chosen in the universal formula because *hemisphere part_of cerebrum*, the genitive being preferred in the presence of a partonomic link.

Official term

```
An official term is (definition) the predominant term
chosen to represent the unit.
```

It is always the official term that represents a unit when it is chosen for an expansion. In more than 95 % of cases, the official term is a universal term. Otherwise, it may be an inherited term or an irregular term.

Example: The term EN: *sternal facet of clavicula* is an official term different from the universal term giving in French FR: *facette articulaire sternale de la clavicule*. In the English version, the adjective *articular* has been removed.

Synonymous term
`A synonymous term is (definition) an alternative official term that can occasionally represent a unit.`

A synonymous term has all the qualities of an official term; in particular, it is just as often derived from a universal formula. But it remains a reserve for the official term for specific situations. These are, for example, cases where a language uses two different words, but both recognized. Or it may happen that doctors choose a different term than surgeons, both being valid.

In any case, whatever the representative value of a synonym, it is always the official term that prevails. Formally, synonyms are not welcome: when they are introduced for reasons of tradition, they are too often a source of ambiguity. It is better to focus on the official terms.

A synonymous term can be regular or irregular, like an official term. There can be zero to three official synonyms per unit. There are also unofficial synonyms; see the DEF:**associated terms** below.

Example: For the universal term EN:*thoracolumbar part*, there is a second universal formula for an official synonym, giving LA:*pars sympathica*, EN:*sympathetic trunk*, and FR:*tronc sympatique*.

Irregular term
`An irregular term is (definition) a term explicitly specified in a specific language.`

An irregular term is a term formulated explicitly in a language, completely independent of the syntactic rules prevailing for that language. In other words, an irregular term is a sequence of characters assumed to have meaning in the field of anatomy for the language used.

The $\mathbf{T}_{logy}$knows nothing about irregular terms, except that they are attached to a unit and are supposed to represent it, according to the authors of the terminology.

Irregular terms are the antithesis of universal terms. While the latter represent the rules of languages in general, irregular terms are in principle free from any rules and are thus completely unpredictable. Yet, created by the authors of the terminology, they are certainly valid and trustworthy. They are, of course, as valid as all the terms in a traditional terminology, which are by definition all irregular. Irregular terms are like a Mongolian text inserted into an English sentence: there is no reason to believe that it means nothing, but in any case it cannot be connected to anything in the terminology; it is a blind text. Yet they are indeed present in the $\mathbf{T}_{logy}$and entirely necessary.

Example: The term EN:*tuber cinereum* in English cannot be derived from the universal formula, which would suggest an English text. However, the authors of the terminology preferred to use the Latin term. It must be explicitly specified as an irregular term.

Example: The French term FR:*neurones en panier de l'hippocampe* does not use the proposed adjective *corbiforme*, but a prepositional phrase is selected instead.

Example: The French term FR:*corps amygdaloide olactif* does not correspond to the universal formula giving in Latin LA:*amygdala olfactoria*, suggesting two words instead of three.

Discarded term
`A discarded term is (definition) a term specified as a synonym in the form of a universal formula and rejected in a specific language.`

It is sometimes convenient to prepare a synonym in the form of a universal formula valid for all languages, but a specific language may well reject it.

Associated term
`An associated term is a term not included in the terminology, but considered to carry interesting information about the unit concerned, useful for research.`

It is clear that an associated term is a synonymous term that has been rejected as such. It is therefore an unofficial synonym. However, some relevant information may escape the terminology and prove useful for finding units that are difficult to identify. Associated terms do not engage the responsibility of the authors of the terminology. They are provided for convenience.

Example: The term FR:*prevertebral ganglion* is also called *collateral ganglion* by the FMA, but the authors of the $\mathbf{T}_{logy}$have not accepted it as an official synonym. For research purposes, it is included as an associated term.

Eponymous term
`An eponymous term is (definition) an associated term comprising a word derived from the proper name of a prominent figure who played a role in the field of anatomy.`

Eponymous terms are common in all terminologies and sometimes serve as important reference points. However, they are not strictly part of the $\mathbf{T}_{logy}$and do not directly engage the responsibility of their authors.

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

Formal term
`A formal term is (definition) a term representing a unit in a theoretically unambiguous manner.`

It is commonly accepted that set terms, and in particular pair terms, are represented using presentational artifices rather than pure text. Thus, set terms are given in the plural, and pair terms are followed by the word *(pair)*. In these cases, a formal term is a term equivalent in full text without the said presentational artifices, always nominative singular. Such a term is not intended for general presentation and is intended to remain in the background.

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

Short term
`An short term is (definition) a term defined with an optional expansion without exercising this function.`

Full term
`A full term is (definition) a term defined with an optional expansion with exercising this function.`

The optional expansion is generally presented in italics.
Example: The term EN:*laryngopharyngeal branch* has the following full name, using its optional expansion LA:*laryngopharyngeal branch of the superior cervical ganglion*.

Set term
`A set term is (definition) a term in the plural for its base part, representing a set unit.`

This term generally represents specific or generic set entities. The plural form is automatically generated in the language of the unit, if it is a universal term.
Example: The term LA:*ganglion thoracicum* yields the following set term LA:*ganglia thoracica*. This is simply the same term in the plural.

Pair term
`An pair term is (definition) a term representing an pair unit.`

It is created by adding the label *(pair)* to its right. The pair term is intended to represent specific pair entities.
Example: The term EN:*spinal lamina V (pair)* represents a pair of two entities: EN:*left spinal lamina V* and EN:*right spinal lamina V*. Note that *spinal lamina V* is in fact synonymous with *neck of posterior horn*.

Lateral term
`A lateral term is (definition) a term representing a left or right member of a pair unit.`

The lateral form is automatically generated in the language of the unit by adding the adjective left or right.
Example: The term EN:*posterior nucleus of the lateral funiculus of the spinal cord (pair)* gives for its left member EN:*posterior nucleus of the left lateral funiculus of the spinal cord*. Three entities are involved in this term, and one of them must receive the lateral adjective. There is a strict rule regarding the positioning of this adjective, valid in all languages: the first rightmost entity that is bilateral receives the adjective.

8.5 Overview of the syntax

The syntax of terms obviously depends on the chosen language. Considerable variation is observed among existing vernacular languages. It is difficult to find general rules for Western languages, and almost impossible if we also consider Asian and African languages. However, the syntactic problems to be solved are relatively the same for all languages, even if the solutions differ. The syntactic problems will be examined later, with the syntactic solutions being left to each vernacular language.

In this section, we will discuss the path from the words, constituting the building blocks, to the final term. This is a first step, preparing the ground for a formal description to come in *chapter 09*.

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 apply to all languages (Russian uses commas in the text). This is a rather intuitive description that requires clarification. Let's first consider what a term is in any given language. A term is a sequence of words in that language, the order of which is constrained by syntactic rules. Each vernacular language has its own syntactic rules, and therefore the order differs from one vernacular language to another. For example, Latin, French, and Spanish place a noun before qualifying adjectives, while English and Russian do the opposite. We will see later that universal formulas use the Latin order, but generator procedures are capable of reconstructing any order from the specified Latin order.

Words are generally separated (at least in Western languages) by a space, except in the presence of compound words, which are more or less frequent depending on the vernacular. The word EN:*spinocerebelar* is a typical compound word in English and other vernacular languages, such as FR:*spinocérébelleux* in French. In the presence of prefixes, a specific vernacular may add a hyphen, but the general tendency in modern languages is to remove them. Some languages, such as German, have a large number of compound words. In any case, for all the vernacular languages considered, word order is important, especially in the presence of consecutive words of the same category, usually a sequece of adjectives.

In the field of anatomy, words are accepted or restricted according to their syntactic category. The main restriction is the exclusion of verbs, with the exception of past and present participles, which are treated like adjectives. Adverbs and conjunctions are also prohibited. Prepositions may be accepted under certain conditions, but this aspect depends on further development. Consequently, terms are essentially constructed with nouns, adjectives, the preposition required for the genitive case, a few invariants (numbers, acronyms, proper names, etc.), and a few articles imposed by syntax. This selection process is of major importance because it defines a restricted sublanguage, considerably easier to manage than any complete natural language. Furthermore, this restriction of word categories is accompanied by a process of selection of the syntactic rules of the specified vernacular. This is a necessary step, to be repeated for each vernacular, in order to avoid terms that are syntactically ambiguous or difficult to interpret.

Some languages present syntactically ambiguous constructions, such as the Latin term LA:*meatus nasi communis* (belonging to TA98): From a syntactical point of view, it is unclear which is common, the meatus or the nose. Such a term can only be interpreted with knowledge of the subject area; Latin syntax is not sufficient. While the correct solution is simple here, this is not always the case. Another unclear situation is the inverted genitive in English, since the apposition of two nouns is also possible, such as EN:*sphincter muscle*: syntactically, there is no way to resolve such a situation. Similar situations can occur in most languages.

Another unexpected but major difficulty arises from the insertion of Latin words mixed with vernacular words. This practice, relatively common in English, creates ambiguous situations. There are several reasons for this, including

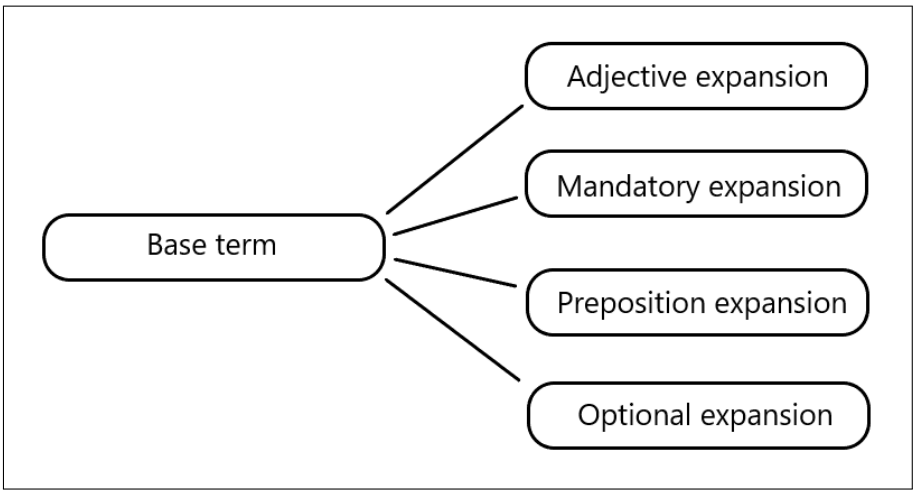

Figure 8.1: The structure of a term is illustrated here. The base term is mandatory and can be accompanied by zero to four expansions. Each expansion is specified by the identifier of the referenced unit. This structure constitutes the basic element of the terminological network.

the need to determine the generation of plural terms. Let's take the example of en:*plexus renal*. There are two of them, and the plural is therefore relevant: should we speak of en:*pair of renal plexus* or en:*pair of renal plexuses*? Unless you prefer the surprising solution found on the internet: en:*pair of renal plexi*! If plexus is a Latin word, like most Latin words mixed into English terms, the plural is defined by Latin rules and is *plexus* invariable. Therefore, *plexuses* and *plexi* are incorrect. However, usage in different English-speaking regions may contradict such a claim. The reality is that we find ourselves in a no man's land where certain rules relating to the plural of nouns are absent, undocumented, or, if they are, not shared by the entire English-speaking community.

8.6 Structure of terms

Having defined the syntactic context, we can now examine the structure of a term. It consists of a mandatory base part, a noun phrase, followed by zero to several expansions, which are links to other entities and are derived from the terms of those entities. This simple organization constitutes the skeleton of any term. Figure 8.1 illustrates this basic schema, to which all terms in the terminology must conform.

Here we present a typical term with multiple expansions. We have selected a portion of the *acoustic radiation*. It has two expansions: one toward the en:*diencephalon*, the other toward the en:*acoustic radiation* itself. The expansion towards the diencephalon allows the adjective to be inserted into the term qualifying the main noun. The second expansion is optional and necessary to qualify this entity, since other diencephalic parts exist.

To determine the importance of expansions, their approximate number was estimated from the $\mathbf{T}_{logy}$database. As a result, for the TAH, up to 40 % of

→TAH5873

LA: *pars diencephalica radiationis acusticae* EN: *diencephalic part of acoustic radiation* FR: *partie diencéphalique de la radiation acoustique* ES: *porción diencéfalica de la radiación acústica* RU:	This example shows a universal formula with two expansions.

terms contain at least one expansion.

The structure of a base part followed by expansions is important and was not made explicit in previous terminologies. In particular, TA98 contained several occurrences of a genitive inserted before all the adjectives. As a result, the rightmost adjective is not linked to the noun preceding it. This tolerance of exceptions leads to purely ambiguous terms, unacceptable in modern terminology. Take, for example, the term LA: *fascia abdominis visceralis*. Here, the adjective *visceralis* qualifies *fascia*, but the syntax is ambiguous and silent on this point. A recommended solution would be LA: *fascia visceralis abdominis*, but it has not been accepted by defenders of traditional Latin.

The concept of multiple expansions implies that no anatomical entity can be hidden within another entity: any related entity must be explicitly mentioned by its identifier, and its name must never be reformulated redundantly. Any unit must be named once and only once; this is the rule for $\mathbf{T}_{logy}$. This is a fundamental aspect to ensure the greatest consistency in modern terminology. Considering past anatomical terminologies, they were developed without databases or major computer applications, incompatible with the resources available at the time. In particular, the TA98 standard did not apply this fundamental principle.

Is it possible today to ignore the internal network of terms? The answer is obvious, because most anatomical terminologies ignore it. The most trivial reason is that most terminologies are monolingual. A more technical reason is that the necessary linguistic techniques are unfamiliar to anatomists and have only been available since the beginning of the 21st century. However, these techniques are relatively simple to implement and perfectly mature given the sub-language of anatomy. The field of anatomy is obviously universal, and we advocate a universal terminology.

In the $\mathbf{T}_{logy}$, the expected number of base parts does not exceed 20,000 elements. Compared to the expected 200,000 terms, the ratio is 1 to 10, with the remainder being generated automatically. This represents a considerable gain compared to the manual work of creating and maintaining traditional terminology. As previously stated, the advantage lies primarily in labor, but above all in the reliability of the entire process: the fewer manual interventions, the fewer textual errors.

8.7 Vocabulary

Before examining the structure of the base part, it is necessary to define its constituent elements: What is the source of the words from which the terms are created?

Two approaches are possible for making words in the field of anatomy accessible: either a predefined dictionary containing all the necessary information, or an open container ready to accept any new word as soon as it appears. In the first solution, security is ensured by initial acceptance and validation of any new word, but missing words will block the system. In the second solution, any word is accepted, regardless of its type, with the risk of accepting irrelevant elements.

The latter solution was preferred for implementing syntactic analysis in $\mathbf{T}_{logy}$, except for Latin, where a dictionary was created and a formal parser is operational. The different solution adopted for Latin is explained by the fact that Latin, a dead language, is relatively stable and well-documented. The other modern languages are considered using the open approach, because building a dictionary for each new language is too labor-intensive. Rather than a tedious initial setup, we favored a gradual acquisition of syntactic knowledge.

The open solution is based on a rule-and-exception-based system capable of acquiring the necessary syntactic information for each presentation of a word. This rule-based system must be recreated for each language, but the task is relatively simple because a general model is available and only the specificities or differences of a new language need to be reformulated. In other words, the formal theoretical approach is replaced by a test-and-correct strategy. Confirmation of this concept was provided by the implementation of the TAH algorithm in four modern languages. The rule-and-exception-based system is documented in *chapter 23*.

In addition to this syntactic processing, a corpus of meaningful words is created in the form of a concordance table between all languages. This table includes all the words included in the universal formulas. For the entire TAH, approximately 2,000 entries are used, including more than 5,000 nouns, adjectives, and prefixes.

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

The figure in the margin shows a small portion of this table for the letter I. An alphabetically sorted concordance table for English can be viewed here: https://ifaa.unifr.ch/Public/TNAEntryPage/auto/test/EN/TestConcordance%20EN.htm. The same table can be sorted alphabetically for any language. These variants are available on the $\mathbf{T}_{logy}$website.

The words in the concordance table are derived from vocabulary entities. These entities constitute true fragments of terminology, as shown in the chapter on fundamental objects. An entity designates a single word declined into a noun, adjective, and prefix, present only when it exists in its language. Each entity is normally expressed in all languages, although some missing words are possible. Words are always given in the nominative singular masculine form. As previously stated, the rule-based system takes into account all possible variations of gender, number, and case for a word in any language.

8.8 Base part of a term

The basic part is essentially a noun phrase. A noun phrase is formed by a head noun, followed (or preceded) by a set of 0 to n adjectives or related forms (such as an ordinal adjective or a present participle). The head noun is generally obligatory, the other words being optional. Adjectives are generally listed from most to least important, which determines their position in different vernacular languages. This notion of importance is vague, but in any case, terminology authors determine a precise order that must be respected in all languages. A typical noun phrase is LA:*meatus acusticus externus cartilagineus* or EN:*cartilaginous exteral acoustic meatus*. In this example, the adjective *acoustic* is the most important.

But the reality is more complex. The following situations arise from anatomical terminology:

Working with Appositions:

An apposition is a sequence of two nouns in the same case, the second generally providing complementary information to the first. Syntactically, the two nouns always share the same case and number, and they cannot be separated by other words. A typical apposition in English is *sphincter muscle*. This is not a inverted genitive, sphincter being not a noun complement of muscle!

The need to accept appositions in terminology varies from language to language and is poorly documented. In Latin, there are many appositions such as LA:*musculus abductor ...* or LA:*musculus extensor ....* In English, the trend is to eliminate these appositions, using a simplified language. In French and Spanish, the trend is probably to maintain all appositions related to muscles, as in Latin. In conclusion, appositions must be maintained and will become part of universal terms.

The existence of apposition has the immediate consequence of excluding the inverted genitive in English, because it is syntactically impossible to distinguish an inverted genitive from an apposition in this language. Yet their meanings are very different. The only way to obtain an inverted genitive in English is to use irregular terms. In any case, the inverted genitive is always ambiguous in the presence of an adjective, because we cannot know which noun is involved.

Prefixed Nouns and Adjectives:

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

Adequate treatment of prefixes is necessary and adapted to different languages.

Existence of Invariants:

An invariant is a word that cannot change, such as a cardinal value (Arabic or Roman numerals) or an acronym. They are certainly not welcome in terminology, because they are often not universally accepted. Cardinal values can be misleading when an element of the list disappears: EN:*spinal*

lamina V is currently the fourth on the list, as recent arguments conclude that laminae III and IV merged, as can be seen in in this list. In any case, cardinal values are rarely meaningful. Roman numerals are not common in all regions of the world and should be avoided in universal terminology.

The syntax for invariants is not well defined. In our approach, we force 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:

In anatomy, the genitive is generally due to obligatory and optional expansions, and there are many of them. However, not all genitives are linked to another anatomical entity. It is rare to have a general genitive as in the term EN:*long association fibers*, but they do exist. Here, the word *association* is such a noun complement and cannot be treated as an expansion.

The existence of a noun complement accompanied by an adjective has not yet been observed and can be ignored. In Latin, this genitive occurs immediately after the main noun it modifies. This may be different in other languages.

A proper design for the treatment of noun complements is necessary in all languages, although it is a low priority.

Proper noun complements:

Proper nouns known as eponyms are not officially part of the terminology, but it is well known that many anatomists are attached to proper nouns and await adequate documentation on this subject. It is worth recalling that TA98 rejected eponyms from its main lists, but published a complete index of proper nouns with references to the entities concerned.

The modeling of eponyms requires the acceptance of proper nouns as noun complements.

8.9 Term Expansion

Term expansion is carried out using four types of expansion: with a genitive (mandatory and optional), with an adjective, or with a preposition. The main purpose of an expansion is to avoid repeating the same term multiple times. Let's take the example of LA:*nomen sternum*: this word appears more than 70 times in the $\mathbf{T}_{logy}$, but it is only defined once in its proper unit EN:*sternum*. A detailed review of all expansions is presented below.

8.9.1 Mandatory Expansions

Mandatory expansion is a genitive expansion: the term referenced by the expansion is added to the right of the base term after being transformed into the genitive singular. The referenced term is constructed with all its own expansions, with the possible exception of an optional expansion. It has a singular base part that is the starting point for the transformation into the genitive singular.

Mandatory expansion is initially intended to represent the relation *part_of*. This means that, if A is the base term and B is the expansion term, the relation A *part_of* B is assumed to be true. This is because the mandatory expansion is expressed by a genitive term, and the genitive is also called the possessive. This simply indicates that the genitive is the predominant expression of partonomy. However, due to the versatility of languages, this rule is only recommended, and many exceptions will be accepted. Indeed, there are many examples where the genitive represents another relationship, such as proximity or resemblance, in all languages. Over time, the original meaning of the genitive has been expanded.

8.9.2 Adjectival Expansion

Adjectival expansion is a specialized form of genitive expansion, where the referenced term can be represented by an adjective (and a prefix). This means that the terminology defines certain vocabulary entities as natural representatives of certain anatomical entities. This is a set of strict rules linking an anatomical entity to a single lexical entity. For example, the entity FR:*clavicle* is generally represented by the adjective *clavicular*, as can be seen with the word LA:*nomen clavicula*.

Thus, whenever the adjective *clavicular* is used in a term, it is not specified in the base term by a reference to the lexical entity, but it is specified in the adjectival expansion by a reference to the anatomical entity. When processing the term, the link between the anatomical entity and the lexical entity is enforced: the adjective is retrieved and inserted immediately after the noun in the base term. An insertion somewhere other than after the noun can be specified. In this way, this expansion is essentially a pointer to a terminological entity, as are all other expansions.

Sometimes an adjectival expansion is specified in a universal term and is not accepted in a given vernacular language. In this case, a mandatory expansion is preferred. It is then possible to inherit the universal formula, requiring the replacement of the adjectival expansion with the mandatory expansion. A typical example is that of the *cerebellar cortex*: the universal term here requires the mandatory expansion, because this cortex is part of the cerebellum and the genitive is indicated, but in English, adjectival expansion is preferred. This type of interaction is necessary to adapt to different languages, while still using universal terminology.

8.9.3 Preposition Expansion

Preposition expansion is a partially developed project. It consists of an expansion introduced by a preposition. Until its implementation in a given language, preposition expansion is replaced by mandatory genitive expansion or by an irregular term explicitly specifying the preposition.

The difficulty with preposition expansion is that in declension-based languages prepositions require different cases that are not yet implemented, such as in Latin and Russian. In English, French, and Spanish, however, implementation is relatively easy.

For example, the term EN:*branch to posterior limb of internal capsule* is a preposition expansion.

8.9.4 Optional Expansion

An optional expansion is identical to a mandatory expansion in its definition, but its use is different.

However, an optional expansion allows for the differentiation of similar terms after all expansions except the optional expansion. About twenty LA:*tunica mucosa* are present in the terminology. They are differentiated only by the addition of their optional expansions, which gives LA:*tunica mucosa gastris*. The purpose of this expansion is that, in a well-defined context such as partonomy, the expansion part – here *gastris* – is not necessary, but outside of any specific context, it must be present. For this reason, most partonomic lists are presented without optional expansions, in order to simplify presentation. But these expansions are generally present in taxonomic lists and elsewhere.

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

That said, it seems that a significant problem, strictly related to optional expansions, is difficult to resolve. This problem arises when generating the lateral members of units with a left or right adjective. The rule without exception defining this location is it is the rightmost bilateral entity composing the term. This universal rule is valid in all languages. It implies that the lateral adjective, once positioned on the term without the optional expansion, may have to be moved when the optional expansion is added.

An example of this paradox is EN:*equator* as part of the *lens*. The short term on its left-hand side is EN:*left equator* and the full term is EN:*equator of left lens*. This gives in English EN:*left equator* and EN:*equator of left lens*.

Unlike terminologies where each term is composed manually, where the rule would apply naturally (with some errors, of course), our implementation of the terminology must formalize the above rule and apply it to all generation processes, which was done. The solution obviously consists of recalculating the entire term, once without expansion and once with expansion. It cannot be constructed by simply adding the optional expansion to its right! This is systematically applied in the terminology.

8.10 Universal Formulas

How can we assemble a basic part and some expansions into a tangible formula to be stored in the terminology database? What is the appropriate representation of the universal core of terms? Universal formulas represent universal terms.

Universal Formula
`A universal formula is (definition) a representation of a universal term by a sequence of functional pointers to the identifiers of the entities that make up its composition.`

Several examples of universal formulas are given in the table 8.2.

The sequence of a universal formula is constituted as a universal term, namely a sequence of the base part followed by the sequence of existing expansions.

UID	Universal formula with expansions
EN:*6017*	N(11577),&X(6017)
EN:*6281*	N(13882),A(13373),&X(6013)
EN:*6087*	N(9526),&X(6013),&F(6077)
EN:*6278*	N(13882),A(13372),&X(6013)
EN:*14221*	N(13882),A(13389),&Q(5991),&X(6013)
EN:*7574*	N(11695),&Q(5769),&X(6004)
EN:*9511*	N(13882),P(13052),A(11925),&X(5991)

Table 8.2: Excerpt from the source table of formulas in the terminology database. The corresponding English term for each formula can be reached following the links in the UID column.
Given this partial table, it is appropriate to imagine the same table with more than 12,000 elements representing all the terms in the $\mathbf{T}_{logy}$, regardless of language.

The base part consists of a sequence of words governed by the formal grammar of terms. Each lexeme is of a specific type: noun, adjective, prefix, and invariant, and can be represented by a corresponding single 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 its identifier in parentheses and listing the sequence of words separated by commas, we obtain the formula of the base part.

For example, the term EN:*postclival fissure* consists of a noun, a prefix, and an adjective and yields the following formula: N(12774),P(11720),A(12828). The noun points to the vocabulary item 12774. If we examine the Spanish part of this item, we find the noun *fisura* or *fissure* for French.

There are four types of expansions, each represented by a letter, each letter must be preceded by &. The mandatory expansion G, the adjectival expansion Q and X, the prepositional expansion R, and the optional expansion F. Each letter must be preceded by &. As with words, each expansion points to a term represented by its identifier in parentheses. The set of expansions is represented by a comma-separated sequence. Finally, the two sequences are added, separated by a comma. The figure 8.2 is extracted from the formula table of the terminology database.

For example, the term EN:*glomerulus of the cerebellum* is composed of a single base part and a mandatory expansion, yielding:

N(11604),&G(5686).

For example, the term EN:*pontocerebellar fibers* is composed of a single base and a prefixed adjectival expansion, yielding:

N(11695),&X(5266),&Q(5686).

The formula naturally yields the different linguistic translations:

→TAH11532

LA:*fibra pontocerebellaris*
EN:*pontocerebellar fiber*
FR:*fibre pontocérébelleuse*
ES:*fibra pontocerebelosa*
RU:

This example shows a universal formula with a prefixed adjective accepted in all languages.

8.11 Epilogue

What have we learned in this important chapter? We have just learned several lessons about the very essence of terminology: terms. It is important to keep all these aspects in mind before making any decisions regarding its future implementation. Here is a summary of this chapter:

Linguistic Independence

A terminology claiming to be universal cannot be based on any specific language: it must be language-independent. Although this statement is rarely applied in reality, it must serve as a guiding principle.

Latin has long been considered the supporting language, because it is neutral with respect to modern languages. It has a precise syntax and, as a dead language, does not suffer from uncontrolled usage. However, it presents some ambiguities and is poorly accepted in countries where its teaching is weak. As for modern languages, they all suffer from biases and tendencies dictated by non-scientific reasons, generally uncontrolled. Furthermore, political arguments for and against are inevitable. In this situation, the choice of a purely formal linguistic representation, independent of any existing language, constitutes a challenge to this dilemma.

Human Resources Aspect

Current data on anatomical terminologies demonstrate that the true scope of the task of creating a modern terminology cannot be handled by a single person, or even several. Furthermore, the search for labor for the development of new terminologies is hampered by major budgetary problems. In practice, the authors of terminologies work on a voluntary basis and generally hold academic positions or are retired.

This assertion makes computer implementation a prerequisite for success. The terminology must be implemented as a relational database, and the interface between the database and a website or other form of publication must be achieved through programmed processes.

Multilingualism

The availability of the five-language version of the TNA by 2022 proves that it is possible to have multiple languages with exactly the same level of wording and precision. Maintenance and evolution of terminology will be facilitated and immediately available in all languages. Other languages can be added to the website with reasonable effort.

Languages are made up of words and must be recognized using their syntactic attributes. However, current language processing does not depend on predefined dictionaries. Indeed, it is preferable to work in an open environment where any new word is immediately accepted without having to worry about updating a dictionary. Its syntactic properties are then defined by an approach based on rules specific to each vernacular language. The results are very positive – up to 99 % success rate – and any errors will be corrected later using an exception list.

However, the corpus of lexical entities defines a central repository of words. A concordance table was created in a two-dimensional array of words and languages.

Term Network and Expansions

In reality, terms are composed of several units, each with its own terms, and there is significant interdependence within the terminology in the form of a term network. Such a network must be made explicit and constitutes a quality assurance measure. Redundancies between terms must be avoided.

Expansions defined across multiple terms constitute the boundary of the term network. There are four distinct types of expansions, meeting all the needs of current terminology. A good understanding of these expansions is essential for terminology authors. Adequate coordination should provide guidelines to ensure overall consistency of expansions.

Formal Grammar

The fine structure of a term requires a precise definition in the form of a formal grammar. Everything that is permitted or prohibited is written down, serving as a guide for current and future authors of the terminology. Updates to the formal grammar are possible at any time, but they require a decision and agreement from the key stakeholders.

The formal grammar is presented in a technical document and its understanding is not necessary for casual users of the terminology. It constitutes the general documentation.

8.12 Log of updates

1 Aug 2025 Complete revision of this chapter, now translated from the source French version.

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