Build a thesaurus

Subject	**Terminological procedures**
Purpose of the procedure	"Link the terms together and publish the thesaurus"
Key words	intentional aspect, terminology, relation, thesaurus, generation, Praxeme, method, procedure

Reference	PxPCD-14f	*State*	Validated
Version	1.2.2	*Date*	12 April 2016

Authors, contributors	Loïc DEPECKER[1], Dominique VAUQUIER Contribution by the consulting firm [2] CONIX
Translator	Joanne TOWARD

Table of contents

Creative Commons - By, Sa

[1] General Delegate for the French language and other languages of France; author of numerous linguistic works.

[2] See http://www.conix.fr/.

Methodological reminders

In the context of the Praxeme method, a *procedure* is "a way of doing something, an operating mode for executing a task"[3]. It is therefore a stipulation on an individual level, in contrast to a *process,* which is a methodological response on a collective level.

The procedure sheets do not refer to possible processes in which these procedures may play a role, in order to facilitate their reuse in several contexts.

To follow the developments of the open method

- Mailing list
- LinkedIn Group
- Twitter
- the wiki

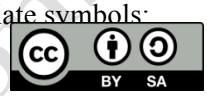
To participate in the work of the Praxeme Institute

- Become a member of the Praxeme Institute

http://www.praxeme.org/communaute/

Document protection

The initiative for an open method rests on voluntary work and the pooling of investments between contributors. It aims to develop and disseminate an open, royalty-free method. Its dynamics only works if this spirit is maintained in the way the documents, which have been made available to the public, are used. This is why the documents are protected with a "creative commons"[4] license, which authorizes the use or reuse of all or part of a document from the Praxeme corpus, the only condition being that the source is quoted. The same conditions should also apply to any documents likely to be derived from Praxeme content. They must refer to the "creative commons" and feature the appropriate symbols:



Updates to this document

To obtain the latest version of this document, please go to the Praxeme Institute website, to the page: http://www.praxeme.org/telechargements/catalogue/.

Document history

Index	Date	Author	Content
0.0.0	12/10/2015	DVAU	Creation
0.1.0	18/10/2015		Review
1.2	2015	Conix	Contribution in the context of the PxData project
1.2.1	08/03/2016	DVAU	Minor corrections
1.2.2	12/04/2016	J. TOWARD	Translation into English
1.2.2	12/04/2016		Current version of the document

[3] Cf. Thesaurus section on the *Praxeme Institute* website: http://wiki.praxeme.org/index.php?n=Thesaurus.Procedure.

[4] See the philosophy and license detail at: http://creativecommons.org/.

1. Application context of the procedure

1.1 Purpose

This procedure "Build a thesaurus" aims to:

"Link the terms together and publish the thesaurus".

It applies, therefore, to sets of terms and seeks to reestablish coherence among them.

The notion of thesaurus differs from that of dictionary in that it enables an access by meaning and not just by term. This notion is defined in more detail in the following chapter. We can guess, straightaway, the value that this may bring us: there are many situations in which we look for something that we are not yet able to name. On an enterprise scale, the efforts relating to these situations lead to costs that are certainly difficult to evaluate, but are, without doubt, considerable. These costs are not limited to time lost – so to amounts spent –; they also include any incomprehension generated, missed opportunities, any redundancy in the expression, which in turn leads to redundancy in practices and systems…

As well as facilitating communication and the sharing of a common vocabulary, the thesaurus fulfills another function. One has to remember that, in the enterprise transformation methodology, the enterprise description repository plays a key role. It gathers and organizes everything that is said about the enterprise. This mass of information and decisions is organized by aspects and architecture decisions. However, these structures are not sufficiently able, from a question or a concern, to allow us to find the place or places where the matter is processed. The thesaurus acts as a "security airlock", enabling us to access the content of the enterprise description repository intelligently and quickly. In the same way as the other content of the intentional aspect[5], it bridges the gap between, on the one hand, the initial and intuitive perception of the domain studied and, on the other hand, the formal and demonstrable model.

This is how a designer will be able to reuse elements that are already available, how an auditor will quickly spot the measures able to satisfy the regulatory requirements, or how a manager will have a precise idea of the implications of a decision.

Of course, to reach these objectives, we have to respect certain rules and impose a certain discipline on ourselves. This procedure aims to set them.

1.2 Usage situations

As soon as work on the enterprise terminology begins, the method recommends publishing one's results as a thesaurus. Ideally, the decision to build a thesaurus should be taken at the very beginning of the effort; the tool should be implemented before the first definition is produced.

The terminological effort can be deployed in a multi-organizational context (synergy between management teams, cooperation between departments, merger or acquisition…). In such cases, the thesaurus's value increases, due to the volume and multiplicity of the sources.

Another situation is that of modeling work. As soon as the volume of such work becomes considerable, even more so if it is carried out on a domain or system scale, it is important to anticipate, very early on, any future difficulties in exploiting or maintaining these models. The method recommends that we require the modeler to record the general or technical terms that will point towards the elements of the model (function of "security airlock"). Thus, in the future, when the detail of the model has been forgotten, the thesaurus will make it easier to find the elements that answer a set of issues.

Finally, the enterprise description repository (EDR) necessarily starts with the enterprise terminology. The thesaurus provides it with the most complete and the most exploitable form. Furthermore, out of all EDR content, the thesaurus concerns the highest number of actors. They all easily recognize its value. What is more, developing a thesaurus is often given as a quick-win in a business architecture approach and as a precondition to any transformation program.

[5] Cf. PxMDS-01, the general guide, for a definition of the aspects, and PxPRD-10, "Approach of the intentional aspect".

Creative Commons - By, SA

1.3 Positioning in the method

a. Place in the reference framework

The thesaurus shapes the "Vocabulary" facet in the intentional aspect[6].

b. Relations with other procedures

This procedure reuses the products generated by the previous procedures in the group on terminology. It completes them by adding relations between the terms, and delivers this material in an easily exploitable form.

In a complete approach, the sequence between the procedures of the group should be carried out in the following way:

- The decision to build a thesaurus, the implementation of the tool (step 1 described below) and the structuring of the "Vocabulary" facet (step 2) are preconditions to the terminological work. The choice of building a reference dictionary (PxPCD-14d) has an influence on how the thesaurus is structured.
- The content of the thesaurus can be initialized first by automatically assimilating a documentary corpus (procedure PxPCD-14b, "Harvesting the terminology" to be used in step 3). The terms are supplied one by one, using the definition procedure (PxPCD-14a), possibly completed by the usage analysis, for sensitive cases (PxPCD-14c).
- The "Build a reference dictionary" (PxPCD-14d) and "Project a term…" (PxPCD-14e) procedures can be done at the same time or when each term is dealt with.
- When the thesaurus has become rich enough, even if it is not finished (will it ever be?), we can move on to generating the sub-products, described in step 6 below.
- The last step of the procedure sees the thesaurus becoming part of the long term. It is promoted and developed through other activities and consolidated, so that it remains a reference and that the approved investment fructifies.

It is important to note that it is always possible to begin stocking and defining the terms, even if the structure of the thesaurus is not definitive, and even if the tooling is not yet in place. The sequence described above is not compulsory. We are in the *procedure* dimension and not in the *process* one[7]. In this dimension, the practitioner has a lot of freedom to assemble the procedures as circumstances dictate.

c. Posture

The procedure does not change, whether the effort is limited to analysis or whether it authorizes design decisions. The posture effect is seen in the choice of terms and the boldness of the definitions compared to current usages[8].

1.4 Conditions to be respected

Implementing this procedure relies on ad hoc tooling. The investment is considerable, as it must cover:

- the selection of the tool, which provides a natural interface to the models[9];
- its purchase and installation;
- almost certainly, personalization, notably regarding generation.

It should be reserved therefore for situations where these questions have already been answered or where the volume and stakes justify the effort.

[6] See the General guide, ref. "PxMDS-01" (Praxeme version 2).

[7] Cf. the Pro3 schema (described in PxMDS-03).

[8] See the discussion in the procedure sheet PxPCD-14.

[9] This is the question of hardware support of the enterprise description repository.

Creative Commons - By sa

2. Terminology used in this document

The terms used in common in this group of procedures are defined in the introduction PxPCD-14. We are concerned here with the terms directly linked to the notion of the thesaurus.

2.1 Methodology

"Aspect", "intention" and "facet" are general terms of the method that we have to understand in order to carry out this work, in particular the decisions to be made on structuring in step 2 of the operating mode. A reminder of their definitions is given below[10].

Aspect: "part of reality, which has been isolated for the sake of study, in accordance with its inner logic".

Facet: "subset of an aspect, characterized by the nature of its elements" [11].

Intentional aspect: "aspect that gathers together the elements of intent, setting the purposes and constraints of a system".

Element of intent: "formulation of an intention".

Intention: "content of will".

2.2 Thesaurus

Thesaurus: "alphabetic directory of normalized terms for content analysis and classification of information documents" (source: *le Grand Robert* French dictionary).

In fact, if we retain, in the expression "alphabetic directory", the idea of a classification in alphabetic order, this definition applies more to a thesaurus-index. On the one hand, this feature does not make the thesaurus different from the dictionary; on the other hand, there are some thesauri that do not respect this definition and group the terms in sets based on a criterion of signification[12].

The thesaurus sets itself apart from the dictionary:

- by the fact that it does not necessarily put the terms into alphabetical order or, if it does, it provides another access to go to the *meaning of the term*, by regrouping the terms by subject;
- by the importance that it attributes to the relations between the terms.

In our French documentation, we use the English spelling of "thesaurus" via Latin from Greek (*thēsauros* 'storehouse, treasure'). The plural should be "thesauri".

Thesaurus-index: "thesaurus whose lexemes are classified in alphabetical order".

The other means of discovering the terms are made available, inside some of the thesaurus entries (list of related words, representation) [13].

2.3 Relations within a thesaurus

So, what characterize the thesaurus are the relations established between the terms. They allow the reader to explore or to discover a knowledge domain, around a term, by going towards the other ones, little by little. This form is therefore far more open than that of a dictionary and authorizes more creative uses.

a. Types of relations

There are recognizable types of relations. The table below indicates the most commonly used types[14].

[10] These reference definitions in the Praxeme method are taken from the Thesaurus (see the bibliography).

[11] For further details, see the discussion at: http://wiki.praxeme.org/index.php?n=Thesaurus.Facet.

[12] This is the case with the first thesaurus of the modern era, that of Mark Roget, as well as the popular Webster's, widely used in the United Kingdom.

[13] A good example of a thesaurus-index is provided by Encyclopaedia Universalis's four volumes (out of thirty).

Creative Commons - By, Sa

Figure PxPCD-14f_1. The terminological relations

Type of relation	Definition	Original role	Destination role	Comment
Specialization	Relation of a general concept to a specific concept	Superordinate concept, type (or supertype)	Subordinate concept, subtype	Hierarchical relation.
Generalization (genericty)	Relation of a specific concept to a general concept[15]	Subordinate concept, subtype	Superordinate concept, type	Symmetric to the previous one, we usually represent the relation in this way.
Partition (aggregation, composition)	Relation between the whole and its parts	Superordinate concept, compound, aggregate	Subordinate concept, compound	There is a nuance between partition (in the logical sense), aggregation and composition (see below).
Instantiation	Relation between a concept and one of its occurrences (or appearances)[16]	Concept, type	Instance (occurrence, example…)	Relation is non-commutative but transitive[17].
Homonymy	Relation between two identical terms representing different concepts	Homonym	Homonym	Pure terminological relation. Symmetric.
Synonymy	Relation between two terms representing the same concept	Synonym	Synonym	Symmetric relation. We could give more importance to one of the synonyms.
Antonymy	Relation between two terms which designate opposite concepts	Antonym	Antonym	Symmetric relation. It combines with synonymy and generalization[18].
Metonymy	Relation between two logically linked concepts	Examples: cause/effect, part/whole, packaging/content, place/activity…		Dissymmetric relation[19].

[14] We can refer to the ISO 704 standard, which is packed with examples (see the bibliography at the end of this document). In the columns concerning the role names, the terms used in the standard are indicated first. The table only adopts the most commonly used terms to designate these relations.

[15] To simplify, the content of this table merges the conceptual relations and the terminological relations. We can observe that the former are often designated by names ending in "-ation" and the latter by names ending in "-nymy".

[16] The reciprocal relation is subsumption. The specific object is subsumed under the concept. Not to be confused with generalization which is a relation between two concepts.

[17] The instance of a substype is considered as the instance of its supertypes.

[18] The synonyms of an antonym for a term are also its antonyms. The same is true, to a certain degree, of antonymy, which combines with generalization and specialization.

[19] Example: in "have a glass", "glass" is subjected to a metonymic usage: we drink the *content* of a glass. The metonymy affects, in particular, specialized languages and not always in a conscious and controlled manner. A classic example is the confusion between the file and the central object (for example an insurance claim or a contract). The term designating the object replaces the term "file" in the mind and in the specialist's language, which can give rise to some difficulties in communication. Terminologists are of great help when they reveal these phenomena.

Creative Commons - By - Sa

Type of relation	Definition	Original role	Destination role	Comment
Connection, Association (general relation)	Significant relation between two terms	(roles named in an ad hoc way)		Connections between terms, associations between concepts.
Contextualization	Relation between a term and a concept, from which its meaning is specified	Term to be defined	Context	The context specifies the domain in which the term is positioned.

The operating mode (step described in § 4.4) specifies how these relations are used.

Modeling makes a big thing of the nuances that can exist between the notions presented here. For example, we distinguish between aggregation and composition. This distinction, formulated in the UML[20] notation, is of interest to the terminologist[21]. In interpreting "partition" in the sense that this term takes on in logic, we will understand that the sum of the subordinates provides the superordinate in its whole.

There is no need, necessarily, to weigh down the notation in order to represent these types of relations and their nuances. The terminologist satisfies him or herself with a small number of typical relations, adding comments to them to keep what he or she has learned about the studied domain. For example, a vehicle has wheels (composition), 4 for a car, 2 for a bike: this information will be the subject of a comment attached to the connection between "vehicle" and "wheel". This example allows us to introduce a very useful notion for building concept trees: the criterion that explains the subdivisions[22].

b. The case of associations

The association is a separate case. It plays out between concepts. The equivalent for terms is the connection. A term's connectivity is its capacity to be linked to others. In thesauri, we can find, in one form or another, a category for related terms[23]. In terminology, the usage is to link the terms, without any other detail. Graphically, we can simply draw a line or, to make it different from other forms of relations, we can specify "relating to" or "related". By limiting ourselves to this means of expression, the relation is symmetric and, almost always, transitive[24].

Creative Commons - By, Sa

[20] Unified Modeling Language, an OMG (Object Management Group) standard.

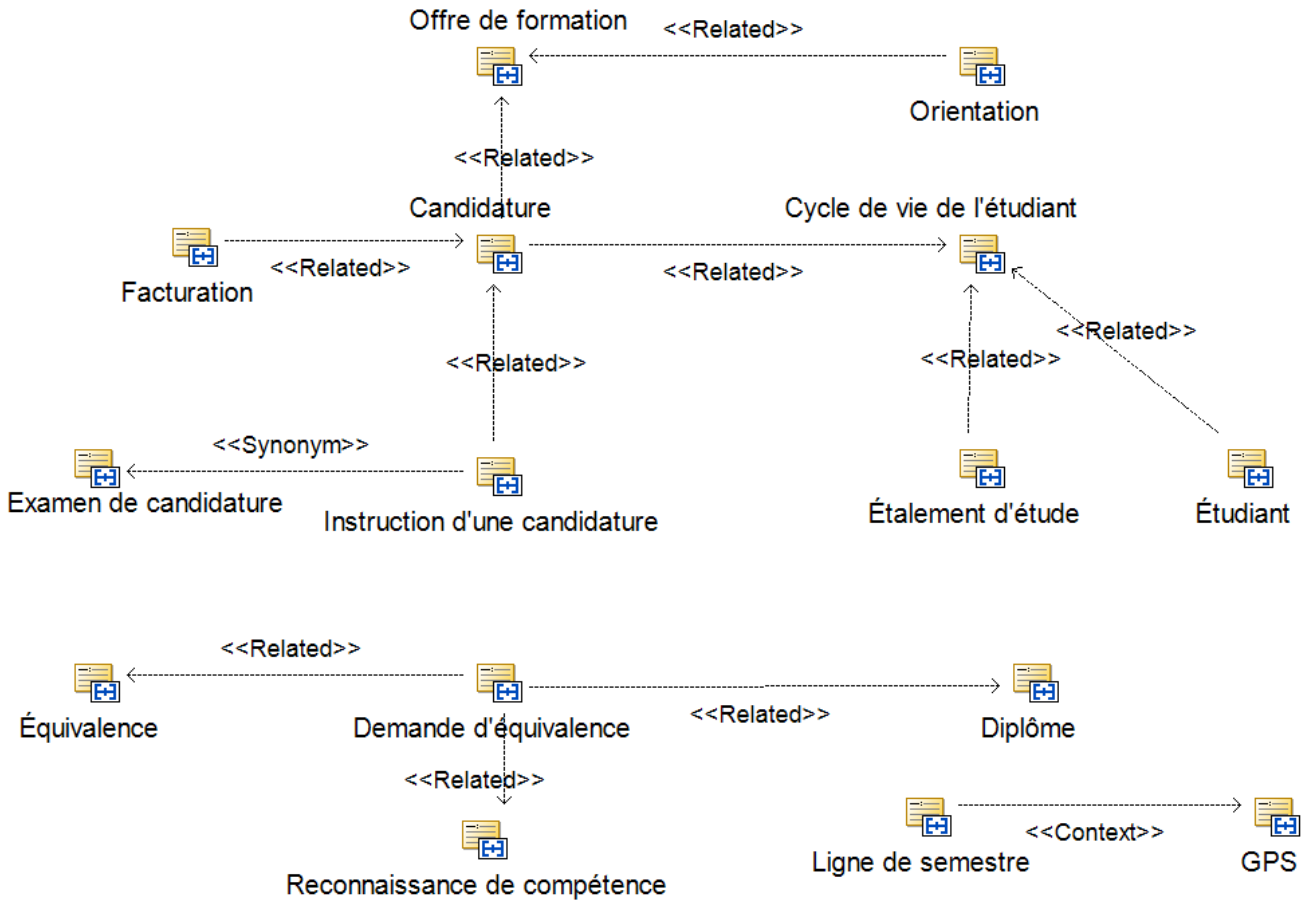
[21] See the ISO 24156 standard.

[22] UML calls it "powertype" and offers to represent it on an inheritance tree. This idea is taken up in the ISO 24156 standard. In a terminology (more easily than in a model), we may find ourselves faced with cases where, from a single concept, several criteria are applied to define its species/kinds. This situation is common: we encounter it every time we think about the classification of our information, files, books… On this point, semantic modeling has very different answers to those of terminology.

[23] See the example of the Praxeme Thesaurus (http://wiki.praxeme.org/index.php?n=Thesaurus.Thesaurus).

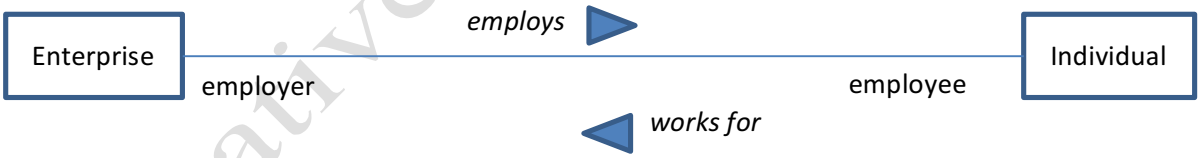
[24] The navigation of its relations may quickly lead to a "grapevine" phenomenon.

Figure PxPCD-14f_2. Illustration of types of terminological relations



If we take an interest now in the association between concepts, we will soon feel the need to clarify things. We are spontaneously led to reproduce the language and re-form the sentences. For example, rather than simply stating that the terms “employer” and “employee” are related, we can easily move to a representation such as the one shown below.

Figure PxPCD-14f_3. An example of a fully commented association (almost)



The representation is that of a class diagram, following the UML notation. The only thing missing are the cardinalities[25].

But we have gone from terminological work to another, that of modeling. This diagram is the result of several decisions that go beyond the scope of terminology[26] and involve other rules. This shift from terminology to

[25] In modeling, cardinalities indicate the number of elements, targets of the association that can be attached to an element at the start of the association. They must be studied at both ends of the association (in the case of a binary association, as in the example).

[26] Among these decisions, we can mention: the reuse of more general notions (enterprise, individual) and the assignation of the terms “employer” and “employee” in relation to the association between both notions (in UML terms, these are the “roles” of the association).

modeling presents dangers, not least of which is the impression – false – of having produced a "business objects"[27] model.

2.4 Relations with elements outside of the thesaurus

The Praxeme methodology places terms in the intentional aspect, where they mix with other elements of intent: values, objectives, requirements, rules, and indicators. These elements also raise the need for clarification. The simplest is to link them with the terms in the thesaurus, by selecting preferred terms, to which a definition is added according to the rule book. These relations are simple cross-references ("trace "). They are only directed one way, from these elements of intent towards the terms.

Traceability: "ability to reconstruct a determination chain".

In the same way, the modeling elements refer back to the terms in the thesaurus. The relation is the same[28].

Between the terms and the modeling elements, a relationship is established that is very important for the value that an enterprise description repository brings: projection.

Projection: "action that consists in giving an element of intent a more formal representation, one that is better integrated into the system, in another aspect".

3. Required competence

It goes without saying that terminology skills are required for this work to be carried out as best as it can be (see the description of skills in the procedure sheet PxPCD-14).

In addition, implementing this procedure relies heavily on the tooling. Mastering the tooling is therefore a prerequisite. On this plane, we distinguish between three levels of competence:

1. the capacity to use the tool to carry out the terminological work;
2. the capacity to personalize the tool or to specify the necessary adaptations to make it comply with the method;
3. the capacity to develop more advanced adaptations, such as generators and macro-commands to manipulate the sets of terms and their documentation.

Apart from the tool used for terminology, it may be necessary to develop programs to deal with a documentary corpus and to incorporate its content in the tool used[29]. These tooling skills are used in the same way for all types of elements of intent.

Finally, another skill mobilized in this procedure, more specifically in its second step, is that of business architecture. At the very least, we have to obtain the structural constraints from business architecture that we will have to respect when building the thesaurus.

From here on, we will call the person who is tasked with building the thesaurus the "terminologist". The terminologist has, above all, skills in terminology and knows how to use the chosen tool. The task of adapting the tool falls to the tools programmer. The tools programmer does not necessarily have any terminological skills. The terminologist has to be able to specify how the tool should be adapted so that it conforms to his or her practices.

4. Operating mode

4.1 Putting the tool in place

The investment in a thesaurus is only justified:

---

[27] A lot more effort will be required to reach the formal representation given by a semantic model. The latter will not be able to limit itself to the static representation given in this example. Cf. the semantic modeling procedures, group PxPCD-20.

[28] It will be illustrated in the operating mode, further on.

[29] For example, from documents or tables.

Creative Commons - By, Sa

- if the effort is part of the general project to put together an enterprise description repository;
- if the content is destined for publication and sharing.

This is why the tooling is particularly important for this procedure, contrary to most of the other ones in the terminological procedures group. Indeed, the volume of information handled and the links created are such that it is impossible to manage efficiently without the minimum of facilities.

The first step, therefore, consists in selecting the tool, then making it available and, possibly, adapting it. The requirements on the tool are deduced in the following steps. Chapter 6 summarizes them.

4.2 Structuring the thesaurus

Regardless of the tool, it is necessary to fix the architecture of the intentional aspect and the rules to be respected for controlling the information in the long term. This is the work of business architecture. First, it must decide on how the "vocabulary" facet is managed, within the intentional aspect.

If we let ourselves be guided by the tools, the facet becomes the decomposition unit of the intentional aspect: we gather the terms among themselves, the objectives among themselves, etc. (see opposite figure). It is not necessarily the best solution. One other solution consists in opting for the origin criterion: each "domain" of the intentional aspect corresponds to a source of information, and we can find all kinds of elements of intent jumbled up together. This option is well suited to analysis activities. For the design, the business architect may prefer a criterion that helps project oneself towards the future of the Enterprise System. For example, he or she may decide to organize the subject matter according to the main aims of the enterprise.

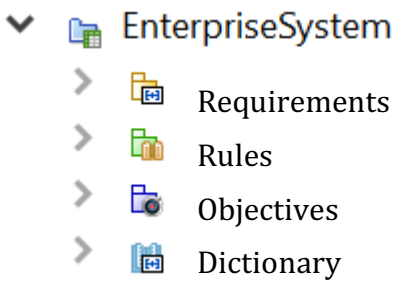


Figure PxPCD-14f_4. An example of structure in place in the tool (Modelio)

In the solution where the dictionary forms a separate whole, we still have to decide on creating sub-dictionaries and the relations between them.

Beyond the main choices of structure, there remain several questions to resolve:

- the separation between the material gathered and the new products, that is to say: on the one hand, the glossaries and texts collected (during the analysis); on the other hand, the reference dictionary (a design product);
- the handling of multilingualism (how should the terms be arranged according to language? how should the connections between one language and another be established? how should the terms be selected according to language in order to generate the sub-products? how can the coherence from one language to another be maintained in the event of any development?);
- the indication of sources (if the thesaurus is not structured by sources, how should we indicate the origin of a definition? how should we reconstruct all the elements which come from the same source?

The choice of structure results in rules to be respected in establishing relations between the elements of intent. They concern step 4. To avoid things becoming complicated, the architecture prohibits mutual relations between sets. For example, if the terms in sub-dictionary A refer to those in sub-dictionary B, then the terms in B cannot refer back to those in A.

4.3 Supplying the thesaurus

There are two ways of recording the terms in the thesaurus:

- by incorporating them automatically,
- by adding them manually.

Certain tools offer the possibility of delving into the texts that we have marked or organized beforehand. This is a very good way of laying the foundations of the thesaurus, by rapidly exploiting the existing material and, possibly, by keeping the links between this material and the elements in the enterprise description repository.

Creative Commons - By Sa

The thesaurus can be added to manually, term by term. In analysis actions (interviews with business experts, questioning workshops, exploiting texts…), each time we come across a term that seems to hold particular importance, we add it. In this way, we can find extracts from reports. In these cases, the quality of the definition is probably not optimum, but what is important is to show that we attach a value to these elements and that they are exploited. We have to distinguish, of course, between the elements burdened with approximate definitions and comments, disparate, noncompliant with the rule book, and the definitions selected or considered as "canonical".

In design, for example when we decide to build a reference dictionary, things happen by making additions, term after term. This is the time for applying the definition procedure[30].

4.4 Linking the terms between themselves

What distinguish the thesaurus from the dictionary are the relations between the terms. It becomes apparent therefore that the linking of terms is an essential task. From the terminological relations, depend several usages that increase the effects of this investment: exploration dynamics, training, reconciling points of view (linked to "convergence"), function of "security airlock[31]"…

In this step, we impose several rules on ourselves, with a view to reducing the number of connections to what is strictly necessary:

- The relations between terms belonging to different sub-dictionaries are oriented in the same direction as the link established between these sub-dictionaries[32].
- Cross-references are made towards the terms of the reference dictionary or that which serves as one, and not the other way round[33].
- The generalization/specialization connections are established from the subordinate concept to the one that is immediately superior to it, and not towards the more generic concepts[34].
- The composition links are established from the compound to the components, the latter being able to exist outside their relation with the compound[35].
- Two synonyms are linked by a single connection. If one of the terms is preferred, the connection leads there[36].
- The same is true with antonyms, except that they are likely to have the same status. We can set ourselves the rule of drawing the link from the negative term to the positive term[37].
- In instantiation, the relation connects the instance to the concept, and not the other way round[38].

---

[30] See procedure sheet PxPCD-14a.

[31] This function makes us go from vocabulary to models, whereas, in this step, we are speaking of the relations between terms. The richness of the terminological connections increases the power of this function. The future user of the enterprise description repository will enter with a term that he or she is familiar with; if this term has been linked, directly or indirectly, to a preferred term, then the chances of finding the appropriate modeling element increase.

[32] That is a basic architecture rule. See above, regarding the structure (§ 4.2).

[33] This rule aims to ensure the stability of the reference dictionary. At any time, a new sub-dictionary may be added, corresponding for example to a glossary proposed by a newly involved entity. The terms will be linked in the new sub-dictionary to the terms in the reference dictionary. Proceeding in this way makes the distribution of responsibilities easier, the reference dictionary is subjected to specific governance associated with transformation.

[34] Example: "human being" towards "animal" and "animal" towards "living being", but not "human being" towards "living being". Indeed, the transitivity of this relation allows use to deduce the last connection, from the first two.

[35] For example: "car" towards "wheel"; also "bike" towards "wheel".

[36] Imagine that "learner" has been preferred to "student" for its greater genericity or by referring to didactics. So, the synonymy link goes from "student" towards "learner". Other links, from other synonyms, lead to this term. Thus we obtain a representation in which the preferred term jumps out at us. This rule reduces any hesitations the terminologist may have. Of course, all this rests on the assumption that the tool will know how to interpret the synonymy relation as being symmetric.

[37] For example, "demotivation" (linked to "frustration", "absenteeism", "obstruction"…) towards "motivation".

Creative Commons - By, Sa

- The association presents a difficulty. First, we have to decide if we want to keep it as a means of expression within the thesaurus. If affirmative, the risk is to overburden the terminological work with modeling concerns. To limit this risk, the terminologist makes it a rule to reflect the language in usage, without adding anything to improve it. Even so, the simplest is to link the terms by a general relation, without seeking to capture all the vocabulary that accompanies this relation (verbs, role names… as shown in the example on page 8).
- Within the intentional aspect, the elements belonging to other facets refer back to the terms, not the contrary[39].
- In contextualization, the link is made towards the context.

These rules presuppose that the tool is sufficiently intelligent to allow us to "go backwards" to the relations established. From a single term, the tool must instantly help us to find all the terms that refer back to it.

The terminologist will appreciate having the possibility to comment the links, as the modeler does. In the example of the vehicles, that a bike has two wheels and a car four, are pieces of information that the connection links can carry towards the term "wheel". These comments will still have to appear in the sub-products generated by the tool.

This step "link the term between themselves" does not necessarily take place in one go, on the whole thesaurus. It can be carried out over time, as and when new entries are created.

4.5 Ensuring traceability towards the models

This step prepares the function of "security airlock". It consists in linking the term to the modeling element that formally recreates its meaning. The link is of a "trace" type[40]. The link goes from the modeling element towards the element of intent. It could not be otherwise in the representation framework provided by the Enterprise System Topology.

In Praxeme, this operation is called "projection". It is the subject of a specific data sheet as it involves other skills than those described here (p. 9).

From the point of view of how the work is organized, this step depends, of course, on the existence of models and their state of progress. It is simpler to give this work to modelers. The terminologist can, at regular intervals, check that the main terms have been "traced".

Note that the information that traceability brings is not of interest to everyone. Indeed, reading these links presupposes a basic understanding of the types of modeling elements. It is therefore wise to plan on displaying the thesaurus in such a way as to hide the traceability links.

The figure below gives an example of the projection of a term. It should be read from right to left:

- On the right appears the term "student".
- Two "trace" links end there: they translate the fact that the (semantic) model does not use a class to represent the notion of student, but that it reuses the generic class "Individual". Moreover, student is defined in the relation between an Individual and a Training course. This is what the link between "learner" (role of the association) and the term indicates.
- The diagram also shows part of the associations involving the class Individual. This is not essential if the focus is limited to the term. Ideally, the tool should allow us to adjust the level of depth desired.

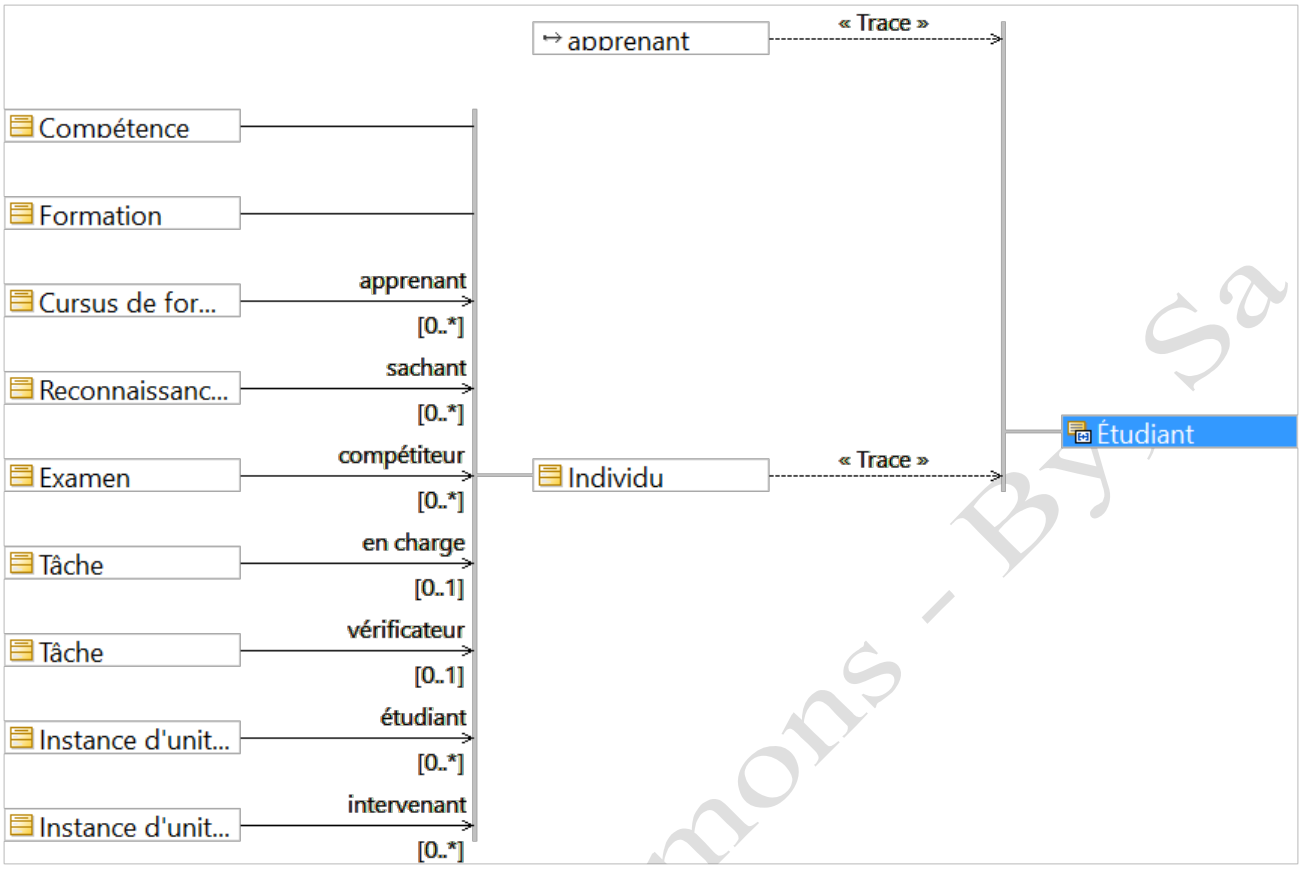
Creative Commons - By, Sa

[38] "Eiffel Tower" towards "tower". This rule is obvious as many examples can illustrate a single concept.

[39] It is in this way that the analyst will dispel any ambiguities in the way a requirement is expressed, will ensure the comprehensibility of a rule, specify an objective or define an indicator.

[40] This is a reserved word in UML.

Figure PxPCD-14f_5. Representation of the traceability links (link editor in the Modelio tool)

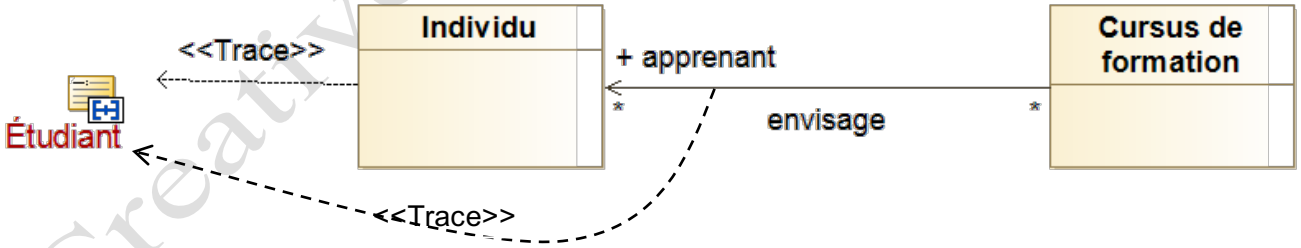


4.6 Generating and publishing communicable forms

This step is applied to the whole thesaurus or to one part of it, a dictionary or a sub-dictionary. It comes into play when all the terms and their documentation are at a sufficient level of quality and stability.

First, the person in charge checks the presence of terminological diagrams or requests that they are produced automatically from the links established in steps 4 and 5.

Figure PxPCD-14f_6. Example of a traceability diagram



Comment: This diagram was generated automatically from the two traceability links pointing towards the term "student".

Once these diagrams have been produced and the content of the terminological whole checked, it is possible to generate the documentary sub-products:

- a single document;
- HTML pages to be published on an intranet or on the Internet (depending on the content).

Generation exploits – or not – the traceability links, depending on the destination.

In hypertext form, the user will appreciate the navigation on the diagrams.

Creative Commons - By Sa

4.7 Promoting and consolidating the thesaurus

This final step completes the life cycle of the thesaurus by projecting it in the long term. As soon as the thesaurus exists, even in an embryonic state, we have to get the most out of it and continue efforts to clarify the terminology. This action goes through accompanying other work (projects, design, thinking on topics…) to make the thesaurus available for use. This work linked to an enterprise's transformation and its investments will always benefit from the perfect command of terminology.

Other opportunities allow us to promote and develop the thesaurus: initiatives such as "clear communication ", focused on the client, homogenization of forms in government administration (with a view to administrative simplification), convergence between partner enterprises or between subsidiaries, mergers and acquisitions…

Through these uses, the thesaurus will evolve naturally. Not only will it be added to, but also, its content will sometimes be reworked, corrected or amended. It will become enriched, particularly through a return effect from the models[41].

5. Results produced

The thesaurus is a deliverable in the form of a document or a set of HTML pages, available for consultation via an intranet or on the Internet. A thesaurus can contain thousands of entries, even – in the scientific environment – hundreds of thousands.

Among the sought-after features, we can mention:

- exhaustiveness, at least the coverage rate of the domain studied;
- tolerance, that is to say taking usages into account;
- the quality of the definitions, in view of respecting the rule book;
- pedagogy (examples, illustrations, terminological discussions to justify the definition, uses of the term…);
- the pertinence of the search results (from a term, a combination of terms, partial terms, a source, etc.);
- navigability, ability to browse the thesaurus by going from one term to another, following the terminological links;
- the level of illustration (using graphics to present the networks of terms).

Inside the enterprise, we might want the thesaurus, in its published form, to communicate with the published form of the models. This function of "security airlock", transposed to the published sub-products, adds to the complexity of the solution.

6. Tooling the procedure

Everything that has gone before has shown the importance of tooling in implementing this procedure.

Many thesaurus management tools are available on the market. They target, above all, the profession of information officer.

In the context of enterprise transformation methodology, the first criterion for selecting the tool remains the assimilation in the enterprise description repository. This leads us to favor modeling tools, so long as they have the required functionalities for building and publishing a thesaurus. The basic functionality is that of a dictionary. The table below summarizes the requirements.

[41] In the previous example, the definition of the term "student" will not be the same if it draws its inspiration from the semantic model.

Creative Commons - By, Sa

Tableau PxPCD-14f_7. Requirements for a thesaurus management tool

Requirement	Definition	Comment	Level of requirement
Incorporation of external material	Capacity to automatically assimilate parts of texts coming from documents from outside the thesaurus	These documents (texts, tables…) can be processed beforehand to identify which sections of text are to be taken from them.	Useful (allows us to quickly lay the foundations of the thesaurus)
Freedom in structuring the thesaurus	Capacity to structure the "Vocabulary" facet without being constrained by the tool	This is rarely the case, except if we throw something together outside the dictionary function.	Desirable (architecture point of view)
Commenting the elements and links	Capacity to attach comments to the terms as well as to the links between terms	Check that the comments entered into the tool reappear correctly when published.	Essential for the terms (definition and description), desirable for the links
Management of multilingualism	Facility to match the terms or sets of terms in different languages	We not only have to distinguish between the language of each term, but also to separate the subsets and manage the correspondences properly (translation tables).	Depending on the needs (by anticipating the future).
List of standard relations	Provision, by the tool, of the typology of relations	If the tool does not provide all the types wanted (see pp. 5 & sq), it is desirable either to be able to add to them, or to reinterpret those that are provided.	Important (linked to the method that we give ourselves)
Interpretation of relations	Capacity of the tool to interpret the types of relations and to adapt its behavior to them.	See the discussions on the features of the relations (symmetry, transitivity…). Exploiting these features enables automatic checks.	Interesting (not respecting this requirement means we have to create more links)[42]
Representation of networks of terms	Capacity to draw links between terms as terminological diagrams	The diagram attaches itself to the term it considers central, or to the whole, providing an overall view of it.	Desirable (a thesaurus without representation is too sad!)
Automatic production of diagrams	The tool's capacity to automatically produce diagrams and update them, according to the relations established between the terms	The user must be able to specify the types of relations wanted (in particular, traceability) and the scope (level of depth).	Very useful (allows us to work faster)
Commenting the representations	Capacity to comment the diagrams	Point of caution: when a comment is attached to a dynamic diagram, we want to keep it through the update.	Essential (it is the means of discussing a lexicographical field)

[42] The case arises systematically with synonyms. If the tool does not exploit the symmetry of the relation of synonymy, it will produce incomplete terminological diagrams. The terminologist will find him or herself obliged to double the links: A synonym of B, B synonym of A. This inflation is dangerous.

ive Commons - By Sa

Requirement	Definition	Comment	Level of requirement
Navigability	Capacity to go from one term to another, by following the terminological links	Navigability also happens from the terms towards the modeling elements.	Desirable (to facilitate the terminologists' and modelers' work)
Generation of the sub-products	Capacity of the tool to publish the content of the thesaurus as texts or hypertexts	Production of documents or publishing in HTML	Essential (it is inconceivable to give all interested actors access to the tool)
Navigability through the sub-products	Navigability inside the products generated from the thesaurus	See, above, the definition of navigability.	Desirable (to better exploit the thesaurus)
Synchronization thesaurus/corpus	Capacity to transfer the modifications between the thesaurus and the source documents, in one direction or another (corollary of the "incorporation" and "generation" requirements)	Once the texts have been integrated ("incorporation…" requirement), it is useful to update them from their state in the thesaurus. In the same way, corrections in the sub-products generated should be able to be transferred back into the thesaurus.	Useful (1st point: to give a new life to these documents; 2nd point: to be able to use spell checkers on the documents to improve the content of the thesaurus)
Version management	Capacity to keep and link the different states of a definition	Always a difficult problem in repositories.	Possibly (important in the event of controversy)
Traceability	Capacity to link modeling elements to terms and to exploit these links	See definition p. 9.	Essential from the standpoint of the enterprise description repository
Authorization management	Capacity to define and manage the responsibilities on parts of the thesaurus or on certain actions	Collective dimension of the terminological work. At least, separate the responsibilities of the terminologist from those of the modeler.	All the more useful as the work involves a large number of contributors
Capacity to personalize	Facility to adapt the tool, symbols and generation to the method and the context	For example, enrich the typology of relations.	Almost always necessary to generate the sub-products

Creative Commons - BY SA

7. Extending one's knowledge

7.1 Correspondences with other frameworks

We found no mention of enterprise terminology in TOGAF[43], neither in the deliverables, nor in the "Business Architecture" phase. None either in the BABoK Guide[44], which will come as a surprise to no one: how can we formulate unambiguous requirements until we have clarified the vocabulary?

The OMG is more prolix:

- On the one hand, the notion of traceability is detailed in the UML standard (http://www.omg.org/spec/UML/).
- On the other hand, the "Semantics of Business Vocabulary and Business Rules" (SBVR) standard proposes a formalization of the expressions that concern the business. It clarifies the notions required for a knowledge approach (expression, concept, designation, representation…), introduces modal logic and quantifiers as a means of better formulating the expressions and contains a typology of propositions. Cf. http://www.omg.org/spec/SBVR/.

7.2 Practical bibliography

a. Standards

The two main standards that we refer to in this procedure are:

- ISO 704 (Third edition, 2009), "Terminology work — Principles and methods"
- ISO 24156-1 (2014), "Graphic notations for concept modeling in terminology work and its relationship with UML -- Part 1: Guidelines for using UML notation in terminology work"

b. Praxeme resources

To find the latest published version of the terminological procedures, the reader can go to the Praxeme Institute website, to the catalog page:

http://www.praxeme.org/telechargements/catalogue/.

The Praxeme Thesaurus is of double interest with regard to this procedure:

- on the one hand, it provides the definitions of the terms of the method;
- on the other hand, it can act as an example as to the form.

It is based on the PmWiki open source technology by Patrick Michaud:

http://www.pmwiki.org/.

Each entry (lexical unit, lexeme) gives rise to an HTML page containing fields. A summary table is built dynamically. It can be found at:

http://wiki.praxeme.org/index.php?n=Thesaurus.Thesaurus.

Let us note in passing that "Thesaurus" is one of the sections of the Organum, the broadest structure proposed by the Praxeme method[45].

c. Examples

Thesaurus of governmental activity on the "Quebec portal": http://www.thesaurus.gouv.qc.ca/tag/accueil.do

[43] The Open Group Architecture Framework.

[44] Business Analysis Book of Knowledge from the IIBA (International Institute of Business Analysis) association.

[45] The Organum is the topic covered in the procedure sheet PxPCD-02.

Creative Commons - By Sa

Index

Creative Commons - By, Sa

Table of illustrations

Creative Commons - By, Sa

Analytical table

Creative Commons - By, Sa