

# BFO-Based Ontological Analysis Framework

## Definition Annotation Manual (Version 1)

By Selja Seppälä

Last modified: October 8, 2015

### Introduction

This definition annotation manual provides guidelines for:

- Segmenting natural language definitions into their *genus* and *differentia* parts.
- Tagging each segment with an ontological label provided by a *relational model* that is based on any of the categories of the Basic Formal Ontology (BFO), a realist upper-level ontology (Arp, 2015; Smith et al., 2015; Spear, 2006).

This manual specifies the basic principles for structuring a corpus of terminological definitions and for implementing the proposed BFO-based ontological annotation schema in XML.

### Note

These specifications correspond to the first version of the annotation schema. The schema is currently being updated to allow, among other things, for an easier integration with OWL.

### Background

This definition annotation schema is part of the 'BFO-Based Ontological Analysis Framework' put forward in Seppälä (2012, 2015a) for carrying out content analyses in terminology. It is an adaptation of the annotation schema defined in Seppälä (2004, chapter IV); both are inspired by the conceptual analysis schema put forward by Sager and L'Homme (1994).

The ontological analysis framework is in many respects similar to the formalisms adopted by García de Quesada et al. (2002), Montero Martínez (2002), or Montero Martínez et al. (2002) in their treatment of the terminology of oncology. The guiding principle in all these works is that definitions are segmented in genus and differentia elements, and then tagged with conceptual 'class' and 'relation' labels. The main

difference with these terminological annotation frameworks is that the set of descriptors proposed here does not represent concepts but types of entities in the world, their properties, and relations between them. This difference brings about distinct choices regarding the annotation schema, especially regarding the processing of the genus element. Another major difference is that the categories and relations used in the ontological analysis framework are very general, as the Basic Formal Ontology represents domain-independent entity types and relations between them.

## Definitions

What is meant here by **definition** is the short text that can be found in the 'definition' field of a (specialized) dictionary entry or in an annotation field of an ontology. Roughly speaking, it generally corresponds to a single sentence that gives a description of the meaning of a term and that does not include the term itself. The following examples illustrate terminological definitions from the domains of *chemical weapons* (Aureille, 2003) and *oil spill cleanup* (Graf, 2003).

**distilled mustard**

**sulfur mustard**

**bis (2-chloroethyl) sulfide**

*A blister agent which is an amber brown liquid with an odor similar to that of burning garlic.*

**submunition**

**bomblet**

*A chemical munition of small size contained in a main one and designed to disperse non-persistent agents.*

**net boom**

*A boom that is made of netting to facilitate the retention of viscous oils.*

**pump system**

*The part of the skimmer that transfers the recovered oil, oil and water and/or emulsions from the skimmer head to a storage tank.*

## Relational Models

What is meant here by **relational model** is a template that specifies the 'ENTITY TYPE+relation+ENTITY TYPE' triples, called **relational configurations** (RCs), that characterize each category of BFO<sup>1</sup>. Each model consists of the RCs that characterize

each category, as well as the RCs that these categories inherit from their parent categories in the ontology – since BFO categories are hierarchically organized through 'is\_a' relations.

The following example shows the relational model for the category OBJECT in BFO 2.0, the first version of the ontology<sup>2</sup>.

## **OBJECT**

### **Relational configurations characterizing the entity type OBJECT**

<has\_part object>

<has\_part object aggregate>

### **Relational configurations inherited from higher levels**

<bearer\_of disposition>

<bearer\_of quality>

<contains process>

<contains process boundary>

<has\_history process>

<has\_part immaterial entity>

<has\_part material entity>

<located\_in independent continuant>

<material\_basis\_of disposition>

<occupies three-dimensional spatial region>

<part\_of immaterial entity>

<part\_of material entity>

<participates\_in process>

These relational models are intended to describe and predict the contents of the definitions according to the category of the thing to which the defined term – and thus the definition itself – refers.

## **Internal Structure of Definitions**

This ontological annotation schema is intended to reveal the internal structure of (terminological) definitions in terms of ontological categories, and relations between them.

The following examples show a possible output of the above definitions annotated using this annotation schema and the model for the category OBJECT.

**distilled mustard**

**sulfur mustard**

**bis (2-chloroethyl) sulfide**

REF	<is_a object>	
GEN	<bearer_of role>	<i>A blister agent</i>
SPE	<bearer_of quality>	<i>which is an amber brown liquid with an odor similar to that of burning garlic.</i>

**submunition****bomblet**

GEN	<is_a object>	<i>A chemical munition</i>
SPE	<bearer_of quality>	<i>of small size</i>
SPE	<located_in site>	<i>contained in a main one</i>
SPE	<bearer_of function>	<i>and designed to disperse non-persistent agents.</i>

**net boom**

GEN	<is_a object>	<i>A boom</i>
SPE	<has_part object>	<i>that is made of netting</i>
SPE	<bearer_of function>	<i>to facilitate the retention of viscous oils.</i>

**pump system**

REF	<is_a object>	
GEN	<part_of object>	<i>The part of the skimmer</i>
SPE	<bearer_of function>	<i>that transfers the recovered oil, oil and water and/or emulsions from the skimmer head to a storage tank.</i>

**Usefulness of the BFO-based Ontological Annotation Framework**

Corpora of existing definitions annotated with this schema can, for example, be analyzed in order to test the hypothesis according to which the contents of definitions are at least in part influenced by the type of entity to which the referent of the defined term belongs (Seppälä, 2012, 2015a). If this hypothesis is verified, then it is possible to use the annotated definitions to see to what extent the entity types influence the defining contents. The annotated corpus can also yield statistical information on the kind of RCs that are relevant for defining terms, thus revealing typical RCs for each BFO category. The statistical results can also be used to assign weights to the RCs in the models.

# Specifications of the Annotation Schema

The annotation of definitions comprises two sub-tasks that can be realized separately or simultaneously:

1. Segmenting the definition into two or more parts.
2. Tagging the parts with an ontological label provided by the 'relational models' that are based on the upper-level realist Basic Formal Ontology (BFO).

The annotations of the definitions are done in XML. The annotation schema contains general tags that identify the dictionary entries constituting the corpus to be analyzed, the domain and sub-domain(s) to which they pertain, and their different fields: definition, terms, notes, phraseological information, and references to other entries. The ontological analysis as such only applies to the content of the definition field, that is, the defining sentence in between the <DF> tags.

The following example shows an excerpt of a corpus of multi-domain definitions structured with these XML tags<sup>3</sup>.

```
<DEFINITIONS_TERMINOLOGIQUES>
  <FICHE langue="en">
    <NI>EN_MA_111</NI>
    <CM>
      <DOMAINE>nettoyage des déversements d'hydrocarbures</DOMAINE>
      <SS-DOM1>lutte en mer</SS-DOM1>
      <SS-DOM2>récupération</SS-DOM2>
    </CM>
    <VE>pump system</VE>
    <DF>
      <GEN typeREF="is_a OBJECT" REFrelGEN="part_of OBJECT">The part of the
skimmer </GEN>
      <SPE REFrelSPE="bearer_of FUNCTION">that transfers the recovered oil, oil
and water and/or emulsions from the skimmer head to a storage tank. </SPE>
    </DF>
    <PH>Based on past experience, the chief cause for an unsuccessful oil skimming
operation has been due to failure in the pump system.</PH>
  </FICHE>
  <FICHE langue="en">
    <NI>EN_SS_151</NI>
    <CM>
      <DOMAINE>natation synchronisée</DOMAINE>
      <SS-DOM1>mouvement</SS-DOM1>
      <SS-DOM2>sortie</SS-DOM2>
    </CM>
    <VE>walkout</VE>
    <DF>
      <GEN REFrelGEN="is_a PROCESS">Movement </GEN>
      <SPE REFrelSPE="has_part PROCESS">which starts in a split position, unless
```

```

otherwise specified in the figure description, in which the hips remain stationary
as one leg is lifted in an arc over the surface to meet the opposite leg.</SPE>
    </DF>
    <NT/>
    <PH/>
    </FICHE>
    ..
</DEFINITIONS_TERMINOLOGIQUES>

```

These structuring tags are explained in the following sections.

## Structuring the Corpus

This section presents the XML elements and, if applicable, their attributes for structuring the corpus of definitions to be analyzed.

**<DEFINITIONS\_TERMINOLOGIQUES>**: 'terminological\_definitions' element that sets the beginning and end of the corpus of definitions.

**<FICHE langue="fr">**: 'record' element that sets the beginning and end of a terminological entry or record. This element has an attribute 'langue' for specifying the language of the record with the [ISO 639-1 two-letter language codes](#) (e.g., 'en' for English; 'fr' for French).

**<NI>**: 'id' element that sets the beginning and end of a unique identifier for each FICHE element<sup>4</sup>.

**<CM>**: 'subject-matter' element that sets the beginning and end of the domain and sub-domain specifications<sup>5</sup>.

**<DOMAINE>**: 'domain' element that sets the beginning and end of the main domain to which the terminological record pertains.

**<SS-DOM#>**: 'sub-domain' element that sets the beginning and end of the sub-domain(s) to which the terminological record pertains<sup>6</sup>. Since there can be several hierarchically organized sub-domains, this element includes a number ('#') specifying the level in the hierarchy (<SS-DOM1> to <SS-DOM3>).

**<VE>**: 'term' element that sets the beginning and end of a term<sup>7</sup>.

**<DF>**: 'definition' element that sets the beginning and end of the definition. The <DF> element includes sub-elements that set the beginning and end of the different parts of the definition (<GEN> and <SPE> elements), thus marking the internal structure of the definitions.

**<GEN>**: 'genus' element that sets the beginning and end of the genus element<sup>8</sup>.

**<SPE>**: 'specifier' (or 'differentia') element that sets the beginning and end of a specifier (or differentia) element<sup>9</sup>.

**<NT>**: 'note' element that sets the beginning and end of a linguistic or encyclopedic note.

**<PH>**: 'phraseology' element that sets the beginning and end of some phraseological (linguistic) information or example.

**<REL>**: 'related term' element that sets the beginning and end of a cross-reference to one or more other terms in the domain.

The <GEN> and <SPE> elements include attributes for the ontological tagging of each segment, respectively, with the categories of the relational models, and their RCs. These elements are presented in more detail in the following sections.

## Segmenting the Definitions

Definitions are segmented in parts that fulfill two different roles within the definition: the *genus*, and one or more *specifier* (or *differentia*) elements.

### The Genus Element

A definition has only one genus element (GEN) that serves to categorize the thing to which the definition refers – which is also the referent of the defined term.

The genus of a definition is enclosed within the following opening and closing tags:

```
<GEN>Genus element</GEN>
```

The genus part of the definition normally relates the defined term to a more general one. The genus can be a term of the same domain as the defined one or a common word. When it is a term of the domain, it can be an immediate parent of the defined term (the *genus proximus*) or a more general parent term. When the term is the *genus proximus*, it is marked with the following attribute<sup>10</sup>:

```
<GEN relationVE="GENRE_PROCH">Genus element</GEN>
```

### The Genus and the Defined Term Have the Same Referent

Normally, the genus part of the definition expresses the categorizing 'is\_a' relation (although the relation is only implicit). In this case, the genus refers to the same category as the referent of the defined term. This is reflected in the 'is\_a CATEGORY' value of the attribute `REFrelGEN`<sup>11</sup>:

```
<GEN REFrelGEN="is_a CATEGORY" relationVE="GENRE_PROCH">Genus
element</GEN>
```

### The Defined Term is an Instance of the Referent's Category

Sometimes, the genus part of the definition expresses an 'instance\_of' relation, meaning that the defined term is a particular instance of the more general category expressed in the GEN. For example, *the Large Hadron Collider* or *LHC* is an instance of a *particle accelerator*. In this case, the value of the attribute `REFrelGEN` is as follows:

```
<GEN REFrelGEN="instance_of CATEGORY">Genus element</GEN>
```

### Particular cases

When segmenting and tagging definitions, the annotator might however find different cases that depart from the ideal case where the GEN tag states the category of the defined term's referent.

### The Referent of the Genus and of the Defined Term Differ

Sometimes, the genus part of the definition refers to a category that is different from the category to which the defined term refers. In these cases, the referent of the defined term bears some other relation to the category expressed in the genus, such as 'part\_of'. The ontological analysis framework addresses these cases explicitly by specifying both relations in separate attributes:

- The 'is\_a' or 'instance\_of' relation between the referent of the defined term and the category to which it belongs (see the `typeREF` attribute).
- Another relation between the referent of the defined term and the category expressed by the genus (see the `REFrelGEN` attribute).

```
<GEN typeREF="is_a CATEGORY" REFrelGEN="other_relation
CATEGORY">Genus element</GEN>
```

### The Definition Has No Part Fulfilling the Genus Role

Sometimes, the definition has no genus part at all (for example in definitions of adjectival terms). In these cases, an empty GEN tag with the `typeREF` attribute is added before the (first) SPE tag:



```
<GEN typeREF="is_a CATEGORY"/><SPE>Some specific element</SPE>
```

### Attributes of the <GEN> element

As we have seen, the ontological analysis framework provides a set of XML attributes to the <GEN> element that address each of these cases while yielding an explicit, consistent, and uniform annotation of the definitions.

The annotation schema also provides attributes for cases where the genus element is cut by a SPE element (`voir_GEN`) and for indicating categorization issues (`statutTypeREF` and `statutREFrelGEN`).

The <GEN> element with all its attributes looks like this:

```
<GEN typeREF="is_a|instance_of CATEGORY" REFrelGEN="some_relation  
CATEGORY" relationVE="GENRE_PROCH" voir_GEN="DEBUT|FIN"  
statutTypeREF="categ" statutREFrelGEN="categ">
```

**typeREF**: attribute that indicates the category to which the defined term refers when it is different from the category to which the GEN refers or when the GEN tag is empty. Its value is always composed of an 'is\_a' or 'instance\_of' relation followed by a category. **REFrelGEN**: attribute that indicates the relation between the referent of the defined term and the category expressed by the genus element (in that order). Its value is always composed of some relation followed by a category.

**relationVE**: attribute that indicates that the category in the GEN is the *genus proximus*, i.e., the immediate parent category of the defined term. This attribute is included only when applicable and its value is **GENRE\_PROCH**.

**voir\_GEN**: attribute that indicates that the GEN is split in two parts by a SPE element. It has the values **DEBUT** for the beginning of the GEN and **FIN** for the last part of the GEN.

**statutTypeREF**: attribute that indicates that the categorization of the RC specified in the `typeREF` tag is uncertain. This attribute is included only when applicable and its value is **categ**.

**statutREFrelGEN**: attribute that indicates that the categorization of the RC specified in the `REFrelGEN` tag is uncertain. This attribute is included only when applicable and its value is **categ**.

### The Specifier Elements (Differentiae)

A specifier (or differentia) element (SPE) serves to specify the information provided by the genus thus narrowing down the meaning of the defined term. In the ideal case, a specifier differentiates the defined term from the genus and from its neighboring terms –

this is why it is classically called *differentia* element.

A definition can contain one or more specifier (or *differentia*) elements. If the specifiers are necessary for distinguishing the term from the neighboring terms, and if they are jointly sufficient to distinguish the defined term from the neighboring terms, then the definition has the classical Aristotelian form of *definition by necessary and sufficient conditions*.

A specifier element in a definition also expresses a relationship between the referent of the defined term and the category of the referent of the SPE, such as 'bearer\_of function'.

A specifier element can occur before the genus element. Generally, this is the case when:

- The SPE element indicates the domain to which the defined term belongs.
- An adjective or any other pre-modifier (that is not part of the domain term that might appear as a genus) precedes the GEN. This occurs in languages, such as English.

### Attributes of a <SPE> element

The ontological analysis framework provides a set of XML attributes to the <SPE> element. The <SPE> element with all its attributes looks like this:

```
<SPE REFrelSPE="some_relation CATEGORY" voir_SPE="DEBUT|FIN"
statutREFrelSPE="categ">
```

**REFrelSPE**: attribute that indicates the relation between the referent of the defined term and the category expressed by the specifier element (in that order). Its value is always composed of some relation followed by a category.

**voir\_SPE**: attribute that indicates that the SPE is split in two parts by another SPE element. It has the values **DEBUT** for the beginning of the SPE and **FIN** for the last part of the SPE.

**statutREFrelSPE**: attribute that indicates that the categorization of the referent specified in the **REFrelSPE** tag is uncertain. This attribute is included only when applicable and its value is **categ**.

## Tagging the Definitions

This section contains the specifications regarding the tagging of the different parts (GEN and SPE) of a definition. The tagging process can be subdivided into two

subtasks: segmenting definitions into GEN and SPE parts, and annotating the parts with <relation CATEGORY> tags.

### Segmentation Criterion

The segmentation of the definition yields a set of autonomous parts (features) that are directly related to the defined term. To determine whether a unit of information (defining feature) is autonomous, it must be possible to formulate an independent sentence separately for each part, and to assign it a relational configuration that describes the type of content that it expresses (in terms of the relation between the referent of the defined term and that of the GEN or SPE segment).

This type of segmentation may differ from the syntactic segmentation, which corresponds to a linguistic realization of the features. For instance, a SPE element can express a particular relation to the referent of the defined term, while linguistically speaking, this information is attached to the preceding SPE, as is the case in the following example (in French).

#### **barrage en filet**

*Barrage flottant constitué de filets qui facilitent la rétention des hydrocarbures visqueux et des déchets solides, mazoutés ou non.*

GEN	<is_a object>	<i>Barrage flottant</i>
SPE	<has_part object>	<i>constitué de filets</i>
SPE	<bearer_of function>	<i>qui facilitent la rétention des hydrocarbures visqueux et des déchets solides, mazoutés ou non.</i>

In this example, the second SPE is semantically and syntactically linked to 'filets' in the first SPE. However, the segmentation was carried out with respect to the referent of the defined term: the content of the second SPE refers to a function realized by the 'object' that is the referent of defined term.

### Annotation Procedure

Annotating a segment involves two subtasks:

1. Identifying the category of the referent of the defined term to assign to it the adequate 'is\_a' or 'instance\_of' categorizing relation and get the corresponding relational model.
2. Identifying the relational configurations of the corresponding relational

model that describe the contents expressed by each segment. If none of the RCs apply, then a new RC can be created using the BFO vocabulary. If the relatum (the second category that follows the relation) is not a leaf BFO category, it is recommended to assign a child BFO category of that more general category when possible. For example, if the RC is <has\_part material entity>, try to see if the referent of the annotated segment can be categorized under the more specific category 'object'. If that is the case, the segment is instead annotated with the tag <has\_part object>.

In both cases, the relational tag is assigned following ontological criteria, that is, according to the type of referent of the defined term.

### Testing the Segmentation and Annotation

To test the adequateness of the segmentation and annotation of the definition, it should be possible to rephrase each segment in an independent sentence. The sentence begins with the defined term followed by the relational configuration assigned to the rephrased segment and by the contents of the segment, as shown in the following example.

#### **submunition**

#### **bomblet**

GEN	<is_a object>	<i>A chemical munition</i>
SPE	<bearer_of quality>	<i>of small size</i>
SPE	<located_in site>	<i>contained in a main one</i>
SPE	<bearer_of function>	<i>and designed to disperse non-persistent agents.</i>

This definition from the domain of chemical weapons is composed of four segments. The autonomy of each segment can be tested with the following reformulations that include the relation that links the segment to the referent (in bold) of the defined term.

- A submunition **is a object** of the type chemical munition.
- A submunition **bears the quality** of being of small size.
- A submunition is **located in a site**, which is a main munition.
- A submunition was designed to and thus **bears the function** of dispersing non-persistent agents.

### Preprocessing

To save time and increase reliability, the segmentation of the GEN can be systematized by automating the process (Seppälä, 2007, 2012, section 6.4.2).

This preprocessing task leverages the hierarchical organization of the terms in a domain. A term appearing at the beginning of a definition (in some cases, with pre-

modifiers) can generally be considered to be the genus of the definition. The automation of the GEN tagging thus consists in automatically searching for the terms of the considered domain at the beginning of the definition. If a term is found, the system tags it with the <GEN> opening and closing tags and the attribute `relationVE="GENRE_PROCH"`, since we can assume that it is a 'genus proximus'.

When no term of the same domain is found at the beginning of the definition, language-specific lexico-syntactic rules may be applied to identify the beginning and end of the GEN.

### **Future Work**

Future work should allow for the automatic pre-tagging of the GEN category, the segmentation of the SPEs, and the pre-tagging of the SPEs with relational configurations. A future mapping from WordNet synsets to BFO will be available as an electronic resource for the automation of these annotation tasks (Seppälä, 2015b).

## References

Robert Arp, Barry Smith, and Andrew D. Spear. *Building Ontologies with Basic Formal Ontology*. MIT Press, Cambridge, MA, 2015.

Marie-Cécile Aureille. *Terminologie du nettoyage des déversements d'hydrocarbures = Oil spill cleanup terminology : anglais-français, English-French*. Master's thesis, École de traduction et d'interprétation, Université de Genève, 2003.

Mercedes García de Quesada, Pedro A. Fuertes Olivera, and Silvia Montero. Propuesta de estructura definicional terminográfica en OntoTerm. *Terminology*, 8(1):57–90, 2002.

Julie Graf. *Glossaire sur les armes chimiques = Glossary on chemical weapons : français-anglais, French-English*. Master's thesis, Université de Genève, École de traduction et d'interprétation, 2003.

Silvia Montero Martínez. *Estructuración conceptual y formalización terminográfica de frasemas en el subdominio de la oncología*. PhD thesis, Universidad de Valladolid, 2002.

Silvia Montero Martínez, Mercedes Garcia de Quesada, and Pedro A. Fuertes-Olivera. Terminological phrasemes in OntoTerm : A new theoretical and practical approach. *Terminology*, 8(2):177–206, 2002.

Juan Carlos Sager and Marie-Claude L'Homme. A model for the definition of concepts : Rules for analytical definitions in terminological databases. *Terminology*, 1(2):351–373, 1994.

Selja Seppälä. *Composition et formalisation conceptuelles de la définition terminographique*. Mémoire de DEA en traitement informatique multilingue, Université de Genève, École de traduction et d'interprétation, 2004.

Selja Seppälä. Repérage automatique de génériques dans les définitions terminographiques. In Nabil Hathout and Philippe Muller, editors, *Actes des 9e Rencontres des étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, pages 447–456, Toulouse, France, June 2007. Association pour le Traitement Automatique des Langues.

Selja Seppälä. *Contraintes sur la sélection des informations dans les définitions terminographiques: vers des modèles relationnels génériques pertinents*. PhD thesis, Département de traitement informatique multilingue (TIM), Faculté de traduction et d'interprétation, Université de Genève, 2012.

Selja Seppälä, Barry Smith, and Werner Ceusters. Applying the Realism-based Ontology-versioning Method for Tracking Changes in the Basic Formal Ontology. In *8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*, Rio de Janeiro, Brazil, 2014.

Selja Seppälä. An ontological framework for modeling the contents of definitions. *Terminology*, 21(1):23–50, 2015a.

Selja Seppälä. Mapping WordNet to the Basic Formal Ontology using the KYOTO ontology. In *Proceedings of ICBO 2015*, 2015b.

Barry Smith. Ontology. In L. Floridi, editor, *Blackwell Guide to the Philosophy of Computing and Information*, pages 155–166. Oxford, Blackwell, 2003.

Barry Smith. On Classifying Material Entities in Basic Formal Ontology. In *Interdisciplinary Ontology. Proceedings of the Third Interdisciplinary Ontology Meeting*, pages 1–13, Tokyo, 2012. Keio University Press.

Barry Smith, Mauricio Almeida, Jonathan Bona, Mathias Brochhausen, Werner Ceusters, Melanie Courtot, Randall Dipert, Albert Goldfain, Pierre Grenon, Janna Hastings, William Hogan, Leonard Jacuzzo, Ingvar Johansson, Chris Mungall, Darren Natale, Fabian Neuhaus, James Overton, Anthony Petosa, Robert Rovetto, Alan Ruttenberg, Mark Ressler, Ron Rudnicki, Selja Seppälä, Stefan Schulz, and Jie Zheng. *BASIC FORMAL ONTOLOGY 2.0: SPECIFICATION AND USER'S GUIDE*, June 26, 2015.

Andrew D. Spear. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany, 2006.

---

<sup>1</sup>Both *relata* correspond to entity types – i.e., categories – of BFO.

<sup>2</sup>To enhance readability, the first entity type of the RC is omitted.

<sup>3</sup>The names of the XML elements are in French or correspond to abbreviations of French terms for describing the fields of a terminological dictionary entry. Each tag will be defined in this manual.

<sup>4</sup>'NI' stands for 'numéro d'identification' in French, that is 'identification number' in English.

<sup>5</sup>'CM' stands for 'code matière' in French, the equivalent of 'subject code' in English, although the latter is not standard terminology.

<sup>6</sup>'SS-DOM' stands for 'sous-domaine' in French.

<sup>7</sup>'VE' stands for 'vedette' in French, that is 'headword' or 'entry term' in English.

<sup>8</sup>The genus element is further defined in the corresponding section (see page <sup>7</sup>).

<sup>9</sup>The specifier (or differentia) element is further defined in the corresponding section (see page <sup>9</sup>).

<sup>10</sup>'VE' stands for 'vedette', the French term for 'entry' in a dictionary. Thus, 'relationVE' is to be understood as the relation between the genus term of the definition and the defined term, as in: 'The genus term is the *genus proximus* of the defined term.'

<sup>11</sup>As it is reflected in the name of the attribute tag, all the relations are unidirectional and they are to be understood as going from the referent of the defined term to the category to which the segment refers. This applies both for GENs and SPEs.