# Semiotic Triangle Revisited for the Purposes of Ontology-based Terminology Management

Kudashev, Igor

Institut Porphyre

2010

Kudashev , I & Kudasheva , I 2010 , Semiotic Triangle Revisited for the Purposes of Ontology-based Terminology Management . in TOTh 2010. Actes de la quatrième þÿconférence TOTh - Annecy 3 & 4 juin 2010 . Institut Porphyre , Anne Terminologie & Ontologie : Théories et applications. Quatrième conférence TOTh , Annecy , France , 03/06/2010 . < http://www.porphyre.org/toth/files/actes/TOTh-2010_04_Kudashev.pdf >

http://hdl.handle.net/10138/33870

acceptedVersion

**Terminologie & Ontologie : Théories et applications**

# Actes de la conférence

# TOTh 2010

Annecy – 3 & 4 juin 2010

avec le soutien de :

- Ministère de la Culture et de la Communication, Délégation Générale à la Langue Française et aux Langues de France
- Association Européenne de Terminologie
- Société française de terminologie
- Ecole d'ingénieurs Polytech'Savoie – Université de Savoie
- Université de Sorbonne nouvelle
- Association EGC (Extraction et Gestion des Connaissances)
- ISKO (International Society for Knowledge Organization) France

Institut Porphyre
*Savoir et Connaissance*

`http://www.porphyre.org`

# Comité scientifique

**Président du Comité Scientifique :** Christophe Roche

**Comité de pilotage**

| | |
|---|---|
| Loïc Depecker | Professeur, Université de Sorbonne nouvelle |
| André Manificat | Directeur, GRETh |
| Christophe Roche | Professeur, Université de Savoie |
| Philippe Thoiron | Professeur émérite, Université de Lyon II |

**Comité de programme**

| | |
|---|---|
| Bruno de Bessé | Professeur, Université de Genève |
| Franco Bertaccini | Professeur, Université de Bologne |
| Gerhard Budin | Professeur, Université de Vienne |
| Marc van Campenhoudt | Professeur, Termisti, ISTI, Bruxelles |
| Danielle Candel | CNRS, Université Paris Diderot |
| Stéphane Chaudiron | Professeur, Université de Lille 3 |
| Rute Costa | Professeur, Universidade Nova de Lisboa |
| Luc Damas | MCF, Université de Savoie |
| Sylvie Després | Professeur, Université Paris 13 |
| François Gaudin | Professeur, Université de Rouen |
| Anne-Marie Gendron | Chancellerie fédérale suisse, Section terminologie |
| Jean-Yves Gresser | Ancien Directeur à la Banque de France |
| Ollivier Haemmerlé | Professeur, Université de Toulouse |
| Michèle Hudon | Professeur, Université de Montréal |
| John Humbley | Professeur, Université Paris 7 |
| Michel Ida | Directeur MINATEC, CEA |
| Hendrik Kockaert | Professeur, Lessius Hogeschool (Anvers) |
| Michel Léonard | Professeur, Université de Genève |
| Pierre Lerat | Professeur honoraire, Equipe Condillac |
| Widad Mustafa | Professeur, Université de Lille 3 |
| Fidelma Ní Ghallchobhair | Foras na Gaeilge (The Irish-Language Body) |
| Henrik Nilsson | Terminologicentrum TNC, Suède |
| Jean Quirion | Professeur, Université d'Ottawa |
| Renato Reinau | Suva, Lucerne |
| François Rousselot | MCF, Université de Strasbourg |
| Gérard Sabah | CNRS, Orsay |
| Michel Simonet | CNRS, Grenoble |
| Marcus Spies | Professeur, Université de Munich |
| Dardo de Vecchi | Professeur associé, Euromed-Management |

**Comité d'organisation** :

Responsable : Luc Damas

Samia Chouder, Joëlle Pellet

# Avant propos

Cette année la conférence a été précédée d'une journée de formation consacrée à la terminologie et l'ontologie, à leurs liens et leurs apports mutuels. L'intérêt qu'a suscité cette journée nous amènera certainement à réitérer l'opération les années suivantes.

Le succès de la conférence d'ouverture de notre collègue Frédéric Nef, portant sur l'ontologie prise dans sa dimension philosophique, a montré, s'il en était encore besoin, la richesse d'une approche pluridisciplinaire.

Animées par différents présidents, les sessions ont alterné présentations théoriques et démonstrations de systèmes, offrant ainsi l'opportunité à plusieurs industriels de nous parler de leurs projets. L'éventail des sujets abordés, à travers les quatorze présentations retenues (incluant la conférence d'ouverture) réparties sur deux jours, illustre la richesse mais aussi la vitalité de notre communauté : aide à la traduction, thésaurus multilingue, phraséologie, entité nommé, recherche d'information, etc. L' « actualité » n'a pas été oubliée à travers une ontologie des risques financiers.

Enfin, les Conférences TOTh sont devenues internationales à partir de cette année avec le français et l'anglais comme langues officielles. Le comité de programme s'est ouvert à de nouveaux membres portant à dix le nombre de pays représentés et à plus de 40% le nombre de personnalités étrangères. Gageons que cette ouverture sera prometteuse.

Christophe Roche

Président du Comité Scientifique

# Table des matières

## CONFERENCE INVITEE



## ARTICLES

## DEMONSTRATIONS

# Semiotic Triangle Revisited for the Purposes of Ontology-based Terminology Management

**Igor Kudashev, Irina Kudasheva**

**Abstract:** In this paper, we examine the limitations of the traditional semiotic triangle from the point of view of ontology-based, multipurpose terminology management and suggest an alternative model based on the concept of terminological lexeme. The new model is being tested in the *TermFactory* project aimed at creating a platform and a workflow for distributed collaborative ontology-based terminology work.

**Keywords:** Ontology-based terminology management, semiotic triangle, terminological lexeme.

## 1. Introduction

In this paper, we describe the principles of modelling of the core schema for an ontology-based terminology management platform called TermFactory (TF). In ontology work the focus is usually on the relations between concepts and their instances while the designations of concepts are given less attention. Often they are just listed as alternative string labels attributed to concepts. TermFactory, in contrast, represents designations, too, as named ontology resources, globally identified by URIs.

TermFactory has been designed not only as a terminology management system for human users but also as a source of terminological and linguistic information for different applications and processes, such as text parsing and generation, speech recognition, machine translation, etc. Due to this multipurpose character of the TermFactory platform, traditional data models used in terminology management have not been readily applicable while designing its schema. In particular, we felt the need to modify and extend the semiotic triangle which had been the starting point in the data modelling of the TF.

At the time of writing, TermFactory has reached the stage of a working demo but there is still a lot of testing and evaluation ahead, so the data model presented in this paper may still undergo some changes. Besides, some features described in this paper, in particular those related to the layout of the

headwords, are supposed to be implemented later. Terminology used in the naming of classes and data types may still be revised at the later stages.

In the sections to follow we discuss the benefits of ontologization of terminological and linguistic data, describe the TermFactory project and platform in more detail, review the traditional semiotic triangle and its interpretations in the terminology theory, examine some limitations of the semiotic triangle from the point of view of multipurpose terminology management, and propose an alternative model.

## 2. Benefits of ontologization of terminological data

Ontologization of terminological data has several benefits. Ontologies have proved a powerful instrument of creating and sharing common understanding about different domains. Defining a term as a concept instance requires precise thinking and negotiations between the parties which are going to commonly use it. Thanks to this, the parties can in the future be certain that they are referring to the same, globally identified object.

Furthermore, information about terms, just like information about the domain, becomes machine-readable. Ontologies can be automatically checked for logical errors. This helps finding mistakes and inconsistencies in existing and newly created terminology.

In ontologies, a lot of data can be inferred automatically using ontology reasoning. Concept and property inheritance lets developers of terminological collections build their work on top of other terminologies. Information about transitive and symmetric relations too gets propagated automatically. A typical TF terminology project does not start from scratch, but imports bridge concepts and properties from more general terminologies and ontologies.

When inferences are inductive (for instance, in the case of partial synonyms and cross-language equivalence relations), a reasoner can generate educated guesses to be verified by a human user. Both types of propagation speed up the input of data in a terminology management system and assists automatic management of links.

An interesting feature of the TF term ontology is that it is not bound to the notion of an entry as the mandatory rigid "container" that keeps individual data elements together. Instead, each object of the description and each element of the description are represented as OWL statements which spell out their relations explicitly and unambiguously. The same data can then be

presented to the end users in many different ways. Static entries become dynamic orientations and views that result from different traversals and serialisations of the term ontology graph.

# 3. TermFactory Project and Platform

## 3.1. TermFactory project

*TermFactory* (TF) is a part of a larger project *ContentFactory* (2008–2010) carried out at several departments of the University of Helsinki and Helsinki University of Technology. The project is financed by the Finnish Funding Agency for Technologies and Innovation (Tekes) and a number of enterprises specialising in language technologies. The initiator and responsible leader of the project is professor Lauri Carlson (University of Helsinki). The *TermFactory* work package of the project aims at creating a platform and a workflow for distributed collaborative ontology-based terminology work.

TermFactory's mission is to allow companies, organizations and individual contributors to collaboratively produce multi-domain special language vocabularies and ontologies (see *Figure 1*). TermFactory can be used to organize content and standardize communication in global multilingual organisations as well as to boost the exchange of ideas and innovations and support education across language barriers.

TermFactory can also serve as a source of terminological and linguistic information for different applications and processes, such as text parsing and generation, speech recognition, machine translation, etc.

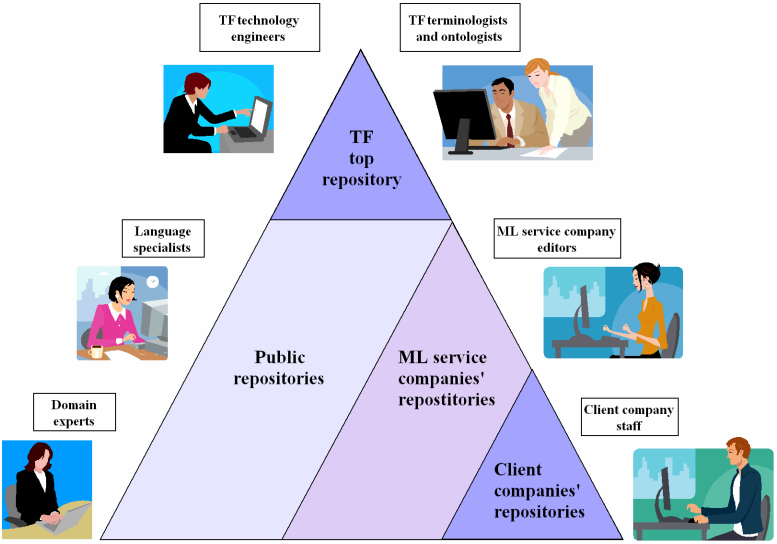


*FIG. 1 – TermFactory's layers and users.*

## 3.2. Collaborativeness

TermFactory is designed to support collaborative, Wikipedia-like terminology work by communities. The tools for collaborative terminology work are based on the content management systems like *Mediawiki*. They can also be implemented as plug-ins for the wiki platforms already used by companies, organisations or web communities.

## 3.3. TermFactory architecture

TermFactory is designed as a distributed resource. TermFactory network consists of OWL repository servers and collaborative wiki / forum platforms connected by a common directory (see *Fig. 2*). The nodes communicate in a peer-to-peer fashion on the web service layer. Collaborative platform servers are loosely coupled to the TermFactory repositories.

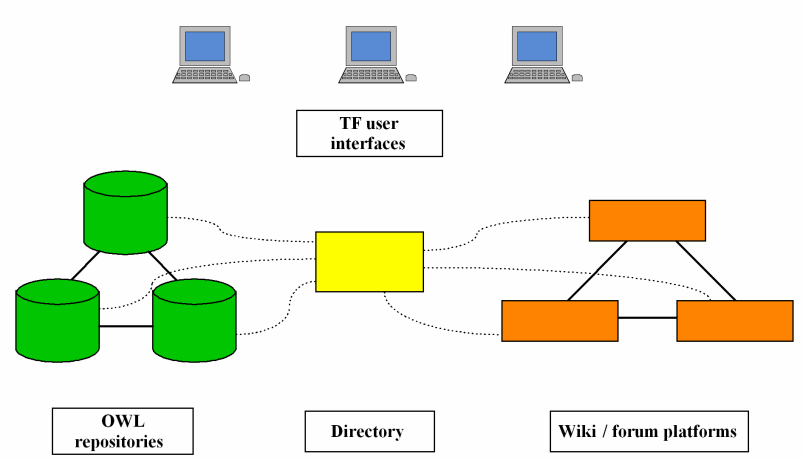


*FIG. 2 – TermFactory architecture.*

## 3.4. Multipurpose and federated character

TermFactory has been designed as a source of terminological and linguistic information for both human users and computer agents. TF accumulates and imports data from multiple sources, including wiki platforms. The multipurpose and federated character of the TermFactory has several important implications for the data modelling.

TermFactory has to be tolerant of a wide range of objects of description and different ways of data presentation. The range of described objects is not restricted to classic terms but includes other LSP designations as well. There is no notion of "minimum entry" in TF. Any set of triples conformant to the TF schema can constitute a TF document.

TF does not require that a particular set of data categories is used. Instead, it tries to federate heterogeneous content with the help of the list of upper-level "metacategories" (see Kudashev, 2009). For example, such class as "information about meaning" is general enough to accommodate all kinds of data related to meaning, including verbal and non-verbal definitions, descriptions, notes, etc. At the same time it is more precise for the purposes of search than "text fields" or "full text of the term record" used in many federated terminological resources.

In many cases computer agents need very specific and well-structured information about LSP designations. This is one of the reasons why TermFactory allows providing separate descriptions for concepts and words. For example, it is possible to describe a concept which has not been attributed a term yet or a word which is not a part of any term presented in the term bank.

Among other things this approach considerably reduces the duplication of data. For example, as the same word can be a component of dozens and hundreds of terms, it is reasonable to describe its linguistic form once and refer to this description as many times as needed.

In spite of its multipurpose nature, TermFactory is first of all a reference resource which contains more or less generalized, context-independent linguistic data. This means, for example, that TF mostly deals with stable, generalized meanings of LSP designations, so there is no need to take into account occasional, contextual or idiolect meanings.

## 4. Semiotic triangle and its interpretations in terminology theory

The starting point in the modelling of the TF schema was the so-called "semiotic triangle" (*Fig. 3*). The triangle is often referred to as the "Ogden and Richards' triangle" as it was famously depicted by these authors in their book "The meaning of meaning" (Ogden and Richards, 1923: 11). However, the idea of the triangle has its roots already in the works by Aristotle (Seuren, 2006: 469).

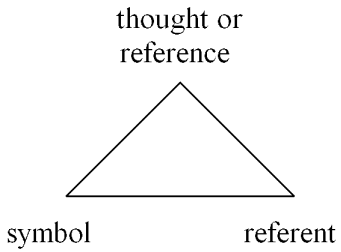


*FIG. 3 – The original semiotic triangle (Ogden & Richards, 1923).*

The triangle has undergone a long history of modifications and interpretations in different theories and by different authors. Terminology

theory has adapted the semiotic triangle to explain the relationship between objects, concepts and terms (e.g. Schmitz, 2006: 579). The triangle is sometimes also represented as a tetrahedron with definition as the fourth vertex (e.g. Suonuuti, 2001: 13; Sanastotyön käsikirja, 1989: 24).

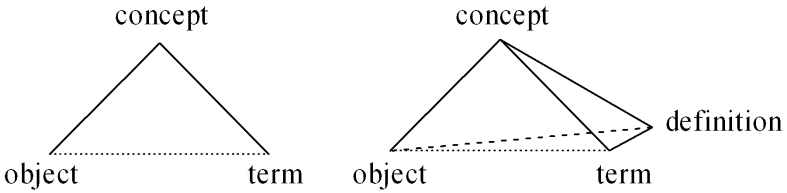


*FIG. 4 – Interpretations of the semiotic triangle in terminology theory.*

When we tried to apply the semiotic triangle to the TF schema, we realized that from the point of view of multipurpose terminology management the triangle had some limitations and required modifications and extensions.

# 5. Limitations of the semiotic triangle from the point of view of multipurpose terminology management

## 5.1. Range of designations

Terms are probably the most important object of description in terminology management systems but they are not the only object. At least the following LSP units should be taken into consideration as potential objects of a terminology management system in addition to terms (see Kudashev, 2010, forthcoming):

**Appellations**, i.e. designations of individual objects/concepts.

**Nomenclature**. This class has multiple interpretations in different terminology schools and theories. One interpretation is that nomenclature is designations of "primitive classes" which are formed by listing necessary characteristics without "closing the class" by specifying sufficient characteristics. A typical example of nomenclature in such interpretation is objects of mass production. It is hard to define them with a classical genus-species definition but it is possible to describe them precisely enough with the help of specifications. However, original specifications may become insufficient for singling out the product as new similar objects come to the market.

Names of objects of mass production, such as "Dreamland Soft" mattress, "Delux Beauty Relax" pillow or "Ecomoods Fabia" lamp, are often neglected in terminology theory and terminology management but in many companies there is a need to deal with them in a multilingual environment (for example, to translate and localize names of company's products). For more information about the presentation of different classes of nomenclature in LSP reference resources see (Kudashev, 2005).

Names of goods should not be mixed with alphanumeric designations assigned to objects in manufacturing and inventory control systems. Stock-keeping units usually refer to *consignments* of products rather than products themselves. Linking between a material management system and a terminology management system is possible but material management categories are unlikely to become objects of terminological description in a term bank.

**Formal notations**. Special concepts and objects may be referred to with the help of means of formal notation, such as

- special symbols: §, €, °, ∑, ∞;
- formulae: $H_2O$, $[As@Ni_{12}As_{20}]^{3\square}$;
- international scientific names: "*Salix starkeana* subsp. *cinerascens*";
- codes: "ABHD12", "C20orf22", "DKFZP434P106", "dJ965G21.2", "BEM46L2", "ABHD12A" are code names of a gene with the official name "abhydrolase domain containing 12";
- catalogue names: "Messier 31", "M31", "NGC 224" are catalogue names of the Andromeda Galaxy.

Formal notations are often used in LSP texts interchangeably with the corresponding appellations and terms and sometimes they are the only existing designation of a special object or concept.

**Lexicalized LSP expressions**, i.e. single-word or multi-word expressions which have a relatively stable form and function in a particular LSP or special area of application. Here are a few examples of lexicalized LSP expressions:

- Instructions: "handle with care", "this end up" (ISO 12620:1999: 10);
- Military commands: "Stand at ease!", "Eyes right/left!", "Double march!";
- Set phrases used in radio and signalling: "More to follow", "How copy?", "Solid copy!"

**Term elements**, i.e. components of LSP designations which have a relatively stable specific meaning in a given LSP. Classical examples of term elements can be found in domains of medicine and chemistry. Here are examples from the domain of medicine:

- Prefixes "a-", "an-" mean an absence of something (apathy, analgia);
- Suffix "-ac" means "pertaining to something" (cardiac);
- Root "aur(i)-" means "pertaining to the ear" (aural).

In addition to morphemes adjectives, participles and components of complex words often function as term elements. When described as objects in their own right, term elements can be provided with a more comprehensive description of their meaning, etymology, usage as well as term formation models.

In TF, difference is made between *term elements* and *term components* because not all words which make up a term have stable specific meaning in a given LSP and deserve a separate terminological description. However, all term components can be provided with a general linguistic description.

## 5.2. Relations within the triangle

For some LSP designations, in particular for appellations, lexicalized expressions and term elements, the model of semiotic triangle is applicable only with certain assumptions or not applicable at all.

For appellations, it can be argued that they designate objects directly and not via concepts. Indeed, if conceptualization is used to group various objects together on the basis of their essential characteristics, with individual objects there is nothing to group.

Term elements denote properties rather than objects so they have "morphological" rather than conceptual meaning.

Lexicalized expressions, such as instructions, are usually statements about specific situations, so they neither denote special concepts nor refer to objects, at least directly.

## 5.3. Additional components of meaning

While working on the Finnish-Russian Forestry Dictionary (Suomalais-venäläinen metsäsanakirja, 2008, awarded by EAFT in 2008) we noticed that in

many cases the meaning of an LSP designation was broader than the intention of the corresponding concept. In addition to the *conceptual meaning* an LSP designation may contain other components of meaning, such as:

- Different connotations, i.e. evaluative components of meaning;
- Inner form of expression (its "literal", morpheme-by-morpheme meaning);
- Components of meaning induced by other LSP or LGP meanings of the same designation;
- Components of meaning resulting from antonymous, synonymous, paronymous and other systematic relations of the designation;
- Different kinds of associations;
- Components of meaning resulting from consonance, rhymes, etc.

These components of meaning can be called *induced meaning* because they result from the attitude of language users towards the form and/or referent of the given LSP expression and/or units related to it by systematic relations or association.

Induced components are welcome and even cultivated on purpose when they create positive associations or allow users to express their attitude to the subject in informal communication. However, in most cases they only distract the users' attention from the logical meaning which is supposed to lie at the core of LSP communication.

This is probably one of the reasons why components of induced meaning have to a great extent been neglected in terminology theory. However, taking them into consideration is an important prerequisite for successful terminological nomination and LSP communication. In our opinion, this topic deserves deeper investigation; both in LGP and LSP (cf. Rigotti & Rocci, 2006, 444). For more information about the presentation of induced meaning in LSP reference resources see (Kudashev, 2006).

## 5.4. Synonymy, polysemy and homonymy

The traditional semiotic triangle abstracts from the fact that the same designation can have several meanings (polysemy, homonymy) and the same meaning can be denoted by several designations (synonymy). However, in terminology management it is very important to anchor the object of terminological description to exactly one form-meaning pair as meaning and domain of application may influence the formal and pragmatic characteristics of an LSP unit (e.g. pronunciation, inflection and combinatory power) and vice

versa, the inner form and relations of a unit may influence its meaning and usage.

### 5.5. Inflected forms

An LSP designation usually has multiple inflected forms. In languages with rich morphology the number of such forms can be quite substantial.

The traditional way of dealing with inflected forms in reference resources is to choose the so-called "canonical form" which represents the whole paradigm. Non-canonical forms are usually ignored but irregular forms may be provided as reference articles.

Division of forms into those included in the reference resource and ignored is important from the point of view of terminology management.

## 6. Alternative data model

Below are suggested several modifications to the original semiotic triangle and a few extensions to the vertexes of the triangle which help overcoming the problems mentioned above.

### 6.1. Removing the object

The *object* vertex of the semiotic triangle is of little significance for terminology management systems because the focus in them is mostly on the designations and concepts. Elements of encyclopaedic information may be considered supplementary information about the meaning of LSP units.

### 6.2. Introducing modified lexeme

Most of the problems described above (the need to deal with a wide range of LSP designations, term components, lexicalized LSP expressions, synonymy, polysemy, homonymy and inflected forms) can be solved if we introduce a class which is general enough to embrace a wide range of objects of description and which represents a union of a set of canonical and non-canonical forms with exactly one meaning.

A very close concept can be found in general lexicography, namely that of a *lexeme*. ISO 24613 defines lexeme as an "abstract unit generally associated with a set of forms sharing a common meaning" (2008: 4). This definition would be completely suitable for our needs unless lexemes were associated in general lexicography only with words and word-like units. However, we also

need to take into account LSP designations (designations of special objects and concepts), lexicalized LSP expressions and term elements.

We suggest that the lexicographical interpretation of lexeme (set of forms of a word or a word-like unit sharing a common meaning) should be called *lexicographical lexeme* while the terminological extension (set of forms of an LSP designation, lexicalized LSP expression or a term element sharing a common meaning) should be called *terminological lexeme*. The concept of lexeme can thus be broadened to embrace both lexicographical and terminological lexemes (*Fig. 5*).

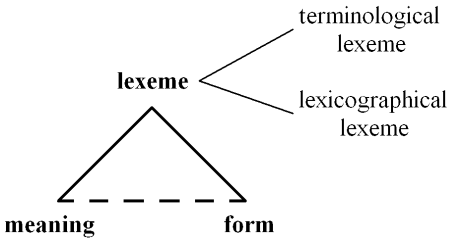


*FIG. 5 – Introducing modified lexeme.*

Lexeme is a tight union between meaning and form, so lines between lexeme and form as well as lexeme and meaning are solid. Meaning and form are often interconnected, too, but their ties are looser, which is depicted with a dashed line. As was mentioned before, meaning and form can even be described independently from lexeme and each other in the TermFactory platform. These descriptions serve as "shared resources" from which descriptions of lexemes sharing the same or similar form or meaning may be built up.

## 6.3. Extending the meaning

Meaning is divided in our model into denotative meaning and induced meaning. Denotative meaning covers the whole range of objects, concepts, properties, situations, conditions, processes, etc., which are referred to by the sign directly and not via associations. Induced meaning includes components of the meaning which are imposed on the sign via associations and relations of different kinds.

Denotative meaning is a broader class than *concept* as it also covers specific types of meaning characteristic of appellations, nomenclature, term elements and lexicalized LSP expressions.

Denotative meaning may be further divided into core denotative meaning and supplementary denotative meaning. This division roughly corresponds to the division made in terminology work between essential characteristics and non-essential but pragmatically important characteristics (see also Kudashev, 2006).

We would like to illustrate the interconnection between form and meaning and the need to differentiate between different types of meaning with the acronym *PIGS* which became popular in economics during the financial and economic crisis 2008–2010. First, a few quotations from the Wikipedia (*http://en.wikipedia.org/wiki/PIGS_(economics)*; a mix of versions from 19.5.2010 and 30.6.2010; references and hyperlinks removed):

"*PIGS*, the original acronym referred to in 1997 to Portugal, Italy, Greece and Spain. […] The acronym has long been used by bank analysts, bond and currency traders […] and is used by some analysts, academics and commentators as a concise way to refer to the Eurozone countries of southern Europe noted for similar economic environments. […] Similar terms, such as "the Olive Belt" or "Club Med", have also been used for the same or similar groupings of countries in southern Europe. […] Ireland, previously known as the Celtic Tiger, became associated with the acronym in 2007 […]. The acronym thus became *PIIGS*, or remained *PIGS* with Ireland replacing Italy. […] The acronym is understood by many to be pejorative, but has been used as a term of art by some. It was denounced as racist in 2008 by the then Portuguese finance minister, Manuel Pinho, following a headline in the Financial Times that read, "Pigs in muck". *http://en.wikipedia.org/wiki/PIGS_(economics) - cite_note-24* Some variants of the acronym have been criticised because economies with similar financial problems, often notably the United Kingdom, are arbitrarily excluded. This has raised some doubts about a possible hidden agenda behind the acronym that would in reality correspond, according to some interpretations, to a wish to deviate the world's attention away from the delicate financial and budgetary situation after 2008's crisis in the UK and the US. GIPSY … refers to the same group as PIIGS, and has the same derogatory sense. It was adopted after protests against the PIGS acronym. It hasn't arrived to substitute the term, though, since it incorporates clear racist connotations

*http://en.wikipedia.org/wiki/PIGS_(economics) - cite_note-23#cite_note-23".*

The case of *PIGS* is interesting in several respects. First, it has undisputable pejorative connotations resulting from its inner form which relates it to an LGP meaning of the word that has the same negative connotations. Second, the meaning of the acronym is highly dependant on its referent and form: new countries can not be added to the group or dropped unless the acronym is changed. As the form is so colourful, it is more tempting to change the meaning, which becomes blurry.

In fact, unlike the 1990s when the acronym was first introduced and had a relatively strict special meaning (four Eurozone countries of Southern Europe with particular economic problems like high or rising government debt levels and a high government deficits relative to annual GDP), in the 2000s the acronym seems to be used more often just to point to the "weak link" in Europe's economy without referring to particular region or economic problems (cf. inclusion of Ireland and arbitrary exclusion of the UK from the list). The core denotative meaning is shrinking and gets more and more shaded by the induced meaning while the supplementary denotative meaning (background information needed for correct understanding of the acronym) is growing.

Synonymous terms and acronyms (the Olive Belt, Club Med, GIPSY, etc.) demonstrate other possible shades of induced meaning for the same referent – from neutral and even slightly positive to ironic and racist.

## 6.4. Extending the form

As was mentioned above, in terminology management and dictionary work it is important to divide the set of forms into the canonical form and non-canonical forms. Non-canonical forms may also be divided into those included in the reference resource and ignored. Extensions to the meaning and form vertexes of the triangle are depicted in *Figure 6*.

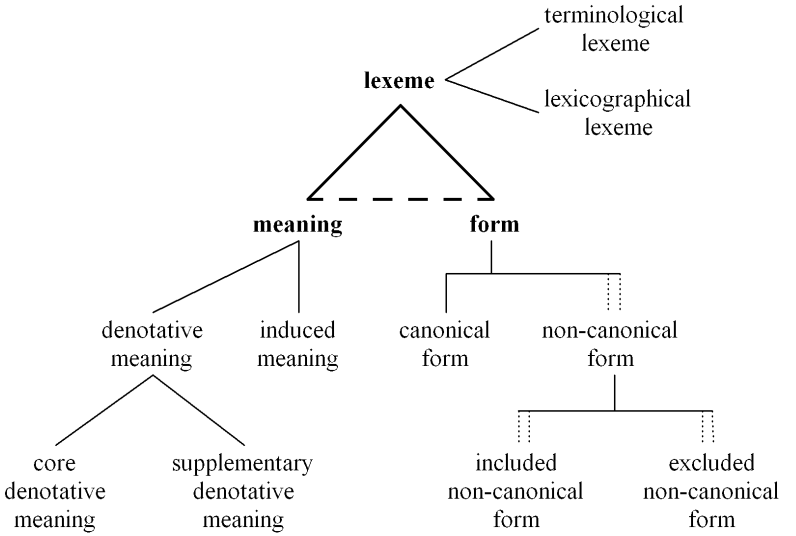


*FIG. 6 – Extensions to the meaning and form vertexes of the triangle.*

## 6.5. Additional extensions to the form

Headwords in a terminological database may also be divided into several classes according to their layout. This can speed up querying and displaying the records in a terminology management system. Such division has not yet been implemented in the TermFactory platform but it was used in MyTerMS terminographic processor in which the Finnish-Russian Forestry Dictionary and several other dictionaries were prepared (see Kudashev & Kudasheva 2006).

Headwords in the entry may contain elements of inline formatting, some additional comments and even other data categories, for example, hyphenation. In printed dictionaries a tilde sign or other means of compression can be used to substitute some portions of the headword. "Entry layout" takes into consideration possible elements of inline formatting as well as "foreign" elements.

In different lists, for example in term lists, hitlists and indexes, additional components of headwords are usually no longer needed but inline formatting has to be preserved.

For the purposes of search headwords usually have to be cleared from all the extra elements and inline formatting. Besides, headwords can be further optimized for the search by bringing them to an upper or lower case. *Figure 7* shows the additional subclasses for canonical and non-canonical forms included in the reference resource as headwords and index words.

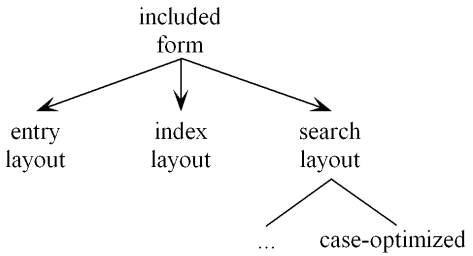


*FIG. 7 – Additional extensions to the form vertex of the triangle.*

Let us illustrate possible differences between the layouts with the help of a term *$CO_2$ laser*. In the entry layout it will have an element of inline formatting – a lower index. Besides, let us suppose that the term also has its hyphenation marked ($CO_2$ la-ser). In the index layout we preserve the lower index but remove the hyphenation ($CO_2$ laser). In the search layout we remove the lower index, too (CO2 laser). If we also bring the term to the lower case, it will be best optimized for the search (co2 laser).

## 7. Conclusion

We have suggested a data model for the purposes of ontology-based multipurpose terminology management which helps overcoming some limitations of the traditional semiotic triangle. In particular, the model takes into account a wider range of LSP designations which can be objects of terminology management, additional types and components of meaning of LSP designations as well as issues which are pragmatically important in terminology management, such as lemmatization and resolving of homonymy, polysemy and synonymy.

Besides, our model partially breaches the gap between reference resources created within terminolographic and lexicographic frameworks and facilitates data exchange between them.

# Bibliography

ISO 12620:1999(E). *Computer Applications in Terminology – Data Categories*. Geneva: ISO.

ISO 24613:2008(E). *Language Resource Management – Lexical Markup Framework (LMF)*. Geneva: ISO.

Kudashev, Igor (2005). *On the Treatment of Nomenclature in Specialised Dictionaries*. In Mård-Miettinen, Karita & Niemelä, Nina (eds) Erikoiskielet ja käännösteoria. Vakki-symposiumi XXV. Vöyri 12.–13.2.2005. Vaasa: Vaasan yliopisto, 176–183.

Kudashev, Igor (2006). *Additional Meaning Components of Terms and their Treatment in LSP Dictionaries*. In Lehtinen, Esa & Niemelä, Nina (eds) Erikoiskielet ja käännösteoria. VAKKI-symposiumi XXVI. Vaasa 11.–12.2.2006. Vaasa: Vaasan yliopisto, 143–149.

Kudashev, Igor (2009). *Improving Compatibility of Terminological Collections with a Bridging Classification of Data Categories*. Terminologija. Vol. 16. Vilnius: Lietuvi kalbos institutas, Terminologijos centras, 36–55.

Kudashev, Igor (2010) *What Can be an Object of Terminological Description in a Terminology Management System?* In Proceedings of the XXX Vakki symposium, Vaasa, Finland, 12-13.2.2010 (forthcoming).

Kudashev, Igor & Kudasheva, Irina. *Software Demo: The Terminographic Processor MyTerMS*. In Schryver, G.-M. de (ed.) DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems, Tuesday 5th September 2006, Turin, Italy (Pre-EURALEX 2006). Pretoria: (SF)² Press, 35–40.

Ogden, C.K. & Richards, I.A. (1923). *The Meaning of Meaning. A study of the influence of language upon thought and of the science of symbolism*. London: Routledge & Kegan Paul.

Rigotti, E. & Rocci, A. (2006). *Denotation versus Connotation*. In Keith Brown (ed.) Encyclopedia of Language and Linguistics. 2nd ed. Vol. 1. Oxford: Elsevier, 436–444.

*Sanastotyön käsikirja* (1989). Soveltavan terminologian periaatteet ja työmenetelmät / Toimittanut Tekniikan sanastokeskus. Helsinki: Suomen Standardoimisliitto.

Schmitz, Klaus-Dirk (2006). *Terminology and Terminological Databases*. In Keith Brown (ed.) Encyclopedia of Language and Linguistics. 2nd ed. Vol. 1. Oxford: Elsevier, 578–587.

Seuren, Pieter A.M. (2006). *Aristotle and Linguistics*. In Keith Brown (ed.) Encyclopedia of Language and Linguistics. 2nd ed. Vol. 1. Oxford: Elsevier, 469–471.

*Suomalais-venäläinen metsäsanakirja* (2008) / Kudasheva, I. & Kudashev, I. (authors); Vehmas-Lehto, I. & Gerd, A. (eds). Helsinki: Metsäkustannus.

Suonuuti, Heidi (2001). *Guide to Terminology*. Helsinki: Tekniikan Sanastokeskus.

## About the authors

**Kudashev Igor**

Affiliations : University of Helsinki

Adresse postale : P.O. Box 239 (Paraatikenttä 6), FI-45100 Kouvola, Finland

Adresse électronique : igor.kudashev@helsinki.fi

Site web : http://www.helsinki.fi/palmenia/kouvola/index.htm

**Kudasheva Irina**

Affiliations : University of Helsinki

Adresse postale : P.O. Box 239 (Paraatikenttä 6), FI-45100 Kouvola, Finland

Adresse électronique : irina.kudasheva@helsinki.fi

Site web : http://www.helsinki.fi/palmenia/kouvola/index.htm

TOTh 2010. *Actes de la quatrième conférence TOTh - Annecy – 3 & 4 juin 2010*
Editeur : Institut Porphyre, *Savoir et Connaissance*
http://www.porphyre.org
Annecy, 2010
ISBN 978-2-9536168-1-1
EAN 9782953616811