Terminology and the Semantic Web

Paradigms, challenges and problems

by Carolina Dunaevsky

A thesis

submitted for the degree of

Master of Arts (M.A.)

in Terminology and Language Technology

at the Technical University of Cologne

2015

Terminology and the Semantic Web

Paradigms, challenges and problems

Examiner 1	Prof. Dr. Klaus-Dirk Schmitz
Examiner 2	Prof. Dr. Lisa Link
by	Carolina Dunaevsky
Matriculation nr.	11093784

TECHNICAL UNIVERSITY OF COLOGNE
Claudiusstraße 1, D-50678
Cologne, Germany

Abstract

Units of knowledge are today most commonly shared across the globe on the Web and the new semantic technologies offer terminologists a challenge and an opportunity at the same time. Ontologies in the sense of knowledge engineering are a highly appreciated resource in the Semantic Web as they provide the descriptive metadata of concepts, which allow identifying terms and associating them to their corresponding domain knowledge hierarchies.

This thesis aims at explaining the role that terminology plays as the interface between knowledge and language in this globalized era, as well as determining whether the (SW) knowledge models available today meet the requirements of terminologists and norms on terminology or not.

Keywords: terminology, language, knowledge, semantic technologies, ontologies, terminological databases, knowledge bases, ontoterminology.

Abstract

Heutzutage werden Wissenseinheiten typischerweise im Internet weltweit ausgetauscht und die neuen semantischen Technologien stellen für Terminologen Herausforderung und Chance zugleich. Ontologien im Sinne von *Knowledge Engineering* sind wertvolle Ressourcen im sogenannten *Semantic Web*, weil sie die deskriptiven Metadaten von Begriffen bereitstellen, welche zur Erkennung von Terminologien und deren Verknüpfung an den jeweiligen Fachwissenshierarchien ermöglichen.

Diese Arbeit dient der Rechtfertigung der Rolle der Terminologie(wissenschaft) als Schnittstelle zwischen Sprache und Wissen im Zeitalter der Globalisierung sowie zur Feststellung, ob die verfügbaren Wissensmodelle (SW) die Anforderungen von Terminologen und Terminologienormen erfüllen oder nicht.

Schlagwörter: Terminologie, Sprache, Wissen, semantische Technologien, Ontologien, Termdatenbanken, Wissensdatenbanken, *Ontoterminology*.

Table of contents

List of Tables

List of Figures

List of abbreviations in alphabetical order

A-Box	Assertional Box
AI	Artificial Intelligence
CDM	Conceptual Data Model
DB	Database
DL	Description Logic
FD	Functional Dependence
GG	Generative Grammar
GOLD	General Ontology for Linguistic Description
HTML	Hypertext Markup Language
ISA	International Federation of the National Standardizing Associations
ISO	International Organization for Standardization
KB	Knowledge Base
KOS	Knowledge Organisation System
LDM	Logical Data Model
LGP	Language for General Purpose
LSP	Language for Special Purpose
NL	Natural Language
RDF	Resource Description Framework
RDFS	RDF Schema
SGML	Standardized Generalized Markup Language
SW	Semantic Web
T-Box	Terminological Box
TDB	Terminological Database
TE	Terminological Entry
TMS	Terminology Management System
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

Introduction

Although terminology, as it is most commonly understood, is rather a recently born discipline, it has been practiced for a long time. Early suggestions (1785) indicate that the naturalist Duhamel du Montceau introduced the term *nomenclature* to designate the concept of ‘art of classifying the objects of a science and naming them’ (cf. Edo Marzá 2009:83). Some good examples of terminology practice in the 18th century are the works undertaken by Lavoisier and Berthollet in chemistry and by Linné in botany and zoology. By then, these scientists had the need of fixing denominations for the scientific concepts they were using. When science became later of international interest in the 19th century, scientists of different fields started to feel the need of normalising and agreeing upon the new denominations, which would enable them to communicate in a precise way and without ambiguities (cf. Cabré 1999:1).

That terminology cares for proper communication between specialists of specific domains and that there should be standards to do so was true then and still remains so. However, there is a fundamental difference between then and today, namely the interconnected and interdependent world we live in. Computers and the Internet have been in use for a considerable time, enabling communication in such a way that virtually everyone can communicate and share knowledge with any other person across the globe. As a consequence, terminology is gaining more and more importance for being the interface between knowledge and language in this globalized area. Terminological practices can be said to be facing new challenges with the latest evolutions in Artificial Intelligence as well as in Information Science & Technology. In the last decade, there has been a shift in the approach of the World Wide Web and the *Web 3.0* was born. It is the so-called *Semantic Web,* best-known for its alleged capability of transferring knowledge in a similar way that humans do. Information about a particular domain is saved here under knowledge bases, which are decentralized repositories that typically have ontological concept systems as underlying models.

Terminology as a science studies the structure, formation, development, usage and management of terminologies (designations) in various subject fields (cf. ISO 1087-1:2000, 3.5.2) while taking concepts – units of knowledge – as basis and structuring them into corresponding systems according to their relations.

In this context, today's question addresses the need of sharing knowledge instead of merely data across the globe, which naturally implies the sharing across the Web. Thus, the main research of this thesis focuses on the assumable relationship between the terminology world and the Semantic Web.

The general objectives of this research include establishing if there is a relation between the terminology world and the Semantic Web (paradigms) as well as discovering the challenges and problems they bring with themselves. The specific objectives (SO) of this research are to determine:

- If the framework/s of the Semantic Web and that of terminology are the same, similar or different (SO-1).
- If the terminology world and the Semantic Web can profit from each other (SO-2).
- If the knowledge models available (for universally conceptualizing the world) meet the requirements of terminologists and those of norms on terminology (SO-3).
- How traditional concept-oriented termbases and ontological knowledge bases are typified (SO-4).

Outline

The topics on terminology and the Semantic Web are captured in this thesis as follows:

Chapter 1

Based on the premise that terminology is the interface between knowledge and language, this chapter provides an overview of different linguistic contributions in relation to language, knowledge and meaning.

The relativistic position of Humboldt and Sapir-Whorf with respect to language is introduced. The analysis of language structure is covered by Chomsky's Generative Grammar and the cognitive aspect of language is presented according to Fillmore's well-known theories of Case Grammar and Frames Semantics. Language is also appreciated by means of semiotics offering an insight into the theories of de Saussure, Ogden and Richards as well as Wüster. The last section presents the main difference between natural language and formal language as well as a brief overview of semantics (meaning/sense).

Chapter 2

This chapter briefly presents the differences between terminology and lexicography to then offer an insight into terminology principles based on the corresponding ISO norms.

A particular emphasis is given to the characteristics and properties involved in the process of conceptualization. Concept systems are materialized as structures ordering domain concepts according to the relations among them and this is treated in the second section. Subsequently, termbases are introduced as a means for terminology management. Best practices for the conception of terminological databases are promoted and the needed data models involved in their development process are exemplified.

In order to establish the differences between a traditional concept-oriented termbase and an ontological knowledge base, section 2.2 illustrates the concept system on which the created termbase (section 2.3) and the comparable ontology (section 4.3) were based.

Chapter 3

This chapter provides an outline to the vision and technologies of the Semantic Web. The main differences between the Web 2.0 and the Web 3.0 are discussed over the introduction of the Semantic Web Stack by Berners-Lee while explaining the role semantics plays in the whole. Furthermore, the Resource Description Framework is exemplified as the basis for data interexchange in the Semantic Web.

The second section delves into various related aspects of knowledge and into the three paradigms commonly used to formally represent knowledge, namely Semantic Networks, Frames and Description Logic. A particular focus is placed on the last two, as they are well-known for providing advantageous features for not only representing knowledge but also reasoning over it based on ontologies.

Chapter 4

This last chapter is dedicated to ontologies as means to organising knowledge on the Semantic Web. The concept *ontology* as suggested by Gruber is explained. It follows a characterization of conceptualization from the point of view of information science for which the proposition of Stock and Stock (2013) is introduced. Furthermore, a distinction between ontologies and knowledge bases is done as well as the different types of ontologies are presented.

The second section offers an insight into Roche's theory of *ontoterminology* (2007-2015) that explains the epistemological principles of concept theory to bringing up the foundations of ontology and terminology together under the same paradigm.

Section 4.3 is a description of the semi-formal domain ontology created (in Protégé) by the author of this research in an attempt to verify whether terminology can be divided into a conceptual and a linguistic root as suggested by Roche.

5 Conclusions

This chapter includes the final conclusions while answering the questions posed in the introduction.

6 Limitations and further research

The different limitations of the studies involved in this research are noted and an overview of future research is provided.

1 Language paradigms

If you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart.

Nelson Mandela

1.1 Relativism

Language has been an interesting field of study for centuries and the science dedicated to studying it is known as linguistics (synchronic focus). Interestingly, before the 20th century, language was rather analysed from a historical point of view (diachronic focus), hence the term philology was commonly used to refer to the science in charge of investigating language.

A traditional approach to conceiving language dates back to the 19th century, when the philosopher Wilhelm von Humboldt (1767-1835) promoted the idea of language being the 'formative organ of thought'. In this sense, language plays an intellectual activity that comes to expression by means of sounds and writings without which the act of thinking could not be perceived as such.

> Sprache ist die grundlegende Voraussetzung des Denkens und damit auch den konkreten Klang der Sprache, den Laut, als Grundlage menschlicher Gedankentätigkeit. (Humboldt, in: Humboldt-Werke, 1963)

Humboldt considers language as the means of thinking, and the act of thinking as the actual tone of language. In this sense, there is no thought without language and there is no language without thought. Furthermore, he states that languages may differ one from another but thought and language (in the sense of 'parole', – language in EN) remain interdependent in any language (in the sense of 'langue' – speech/ to speak in EN). That is to say, every language is a representation of the world view a community belongs to and how it is used as well as the thoughts it evokes depends on the culture a person is immersed in.

This argument is most often associated with the Sapir-Whorf- hypothesis that was first discussed by Edward Sapir (mentor) in 1929 and became popular in the 1950th by Benjamin Lee Whorf. Their hypothesis is most commonly known as a principle of linguistic relativity, which holds that the way nature is categorized (how it is grammatically organized and how people refer to it by means of concepts) causes speakers of different languages to conceptualize and experience the world differently.

> Menschen, die Sprachen mit sehr verschiedenen Grammatiken benutzen, werden durch diese Grammatiken zu typisch verschiedenen Beobachtungen und verschiedenen Bewertungen äußerlich ähnlicher Beobachtungen geführt. Sie sind daher als Beobachter einander nicht äquivalent, sondern gelangen zu irgendwie verschiedenen Ansichten von der Welt. (Whorf, 1963, in: Radegundis, 1994:27)

According to this passage, it is evident that Sapir-Whorf relate the meaning of a word (in the framework of a person's worldview) to grammar. Although a person might be capable of reflecting on the meaning of a word the grammatical foundations of it are strictly beyond the scope of conscious reasoning for allowing a common interpretation.

This era marked by relativism supports the integral premise that the knowledge a person has about the world is key to conveying meaning of any kind. And precisely these notions of world knowledge and its conceptualization are fundamental in this research.

1.2 Language from a grammatical and a cognitive viewpoint

Following the time-line in linguistic research, Noam Chomsky's theories not only mark the linguistic approaches between 1950 and 1970 but are the trigger for later linguistic models.

Chomsky dedicated his efforts to analysing the structure of language in view of what his professor's theory (namely the *immediate constituent analysis*) was lacking. Chomsky's Generative Grammar (hereinafter GG) introduces the notions of *deep structures* and *surface structures*, where the former can be seen as the conceptual basis that all languages have in common and the latter as the distinctive structure that depends on the language to which it is applied (cf. Vater 1996:77-79).

The first version of Chomsky's GG deals with the so-called deep structures that generate the syntax of sentences. For a fact, this grammar is called *generative* as it simulates the underlying dynamic process of generating sentences. Chomsky's self-coined term *phrase-marker* implies that such markers are generated by rules, which can in turn generate new sentences. In contradistinction to traditional sentence parsing (i.e. syntactically classifying a sentence into subject and predicate along with their respective constituents), phrase-markers are supposed to replace a sentence. Taking the sentence (S) /I have a house/ as example, the phrase-markers *noun* (N) and *verbal phrase* (VP) along with their respective constituents translate the sentence into the figure below:

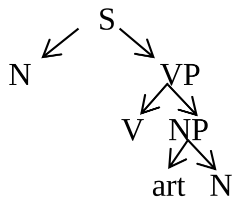


Figure 1: Deep structure of a sentence (according to Chomsky's GG)

Naturally, other linguists criticized this rationale for not taking semantics into consideration. The problem on such a deep structure is that it can stand for any content. Further analysing the same example from above, this deep structure can be well transformed into the surface structure of the proposed original sentence as well as into the surface structure of a sentence like /She has a brother/. For this reason, Chomsky revised his own theory and, in a second version, added semantics to the equation. Deep structures are then not only generated with phrase-markers but also with lexicon rules, which are part of a person's innate body of linguistic knowledge (with a semantic and a phonological component). Chomsky argues next that such rules would guarantee no more changes in the meaning of a sentence when associating a deep structure to a surface structure (cf. ibidem).

Nevertheless, many linguists did not find Chomsky's Generative Grammar theory satisfactory because it offered more exceptions than rules and was not properly extendable to semantics. However, his approach becomes important in computer science, since this is an attempt to model human language with a formal grammar by describing hierarchies and their relations.

Ronald Langacker and Charles Fillmore were the first linguists that included a true semantic value as well as a cognitive one in their approaches. Langacker (1987) proposed a theory called *Cognitive Grammar* that treats the grammar of languages as symbolic systems. He assumes that a linguistic sign has not only a semantic but also a phonological component both being motivated by general cognitive processes (cf. Faber, 2012:95-97). Thus, he states that knowledge conceptualization (characterized by the same semantic/conceptual terms) must be semantically and grammatically consistent. In some way, this grammar can be related to Humboldt's and Chomsky's theories in that it explains that language is a direct reflection of the human mind.

The transformational grammarian, Charles Fillmore, proposed in 1968 his theory called *Case Grammar* which he further expanded in 1971. This revised model highlighted the fact that semantics can also anticipate the syntactic value of a proposition. His language analysis starts at the verb of a proposition and tries to find the relationships between syntactically related meanings of verbs. Fillmore discovered that verbs are generally followed by noun phrases and tried to systematically identify which patterns relate to which meanings. In doing so, he came to acknowledge that verbs allegedly have the need of agents (A), instruments (I) and/or objects (O) to conveying their specific meaning. A particular proposition can fairly differ in its syntax (because the focus on the pieces of information changes – e.g. by focusing on who, what or by which means – but the meaning of the verb remains the same. Although the state or action described by a same morphological verb is what makes up the deep structure of a proposition the semantic value at the surface structure is maintained. Therefore, the syntactic structure of a proposition can be predicted by semantic participants. In the following propositions it is possible to identify the same semantic values by contrasting syntactically different propositions of the same lexical verb:

(1) Case closed.
(2) A clue closed the case.
(3) Mary closed the case with a clue.

According to Fillmore, the patterns (so-called *case frames*) of these propositions will respectively be as follows + [__O], + [__I, O], + [__A, I, O]. The different uses of the verb are arranged as case frames that help organizing the verb as a single combined lexical entry. (cf. Cook, 1922:32-33). In this, Fillmore tries to make clear the importance of semantic roles and how they motivate syntactic as well as morphological constructions. That means that the subject of a verb will be hierarchically determined by the case frames it allows. In the proposition 3 above, the subject of the verb 'to close' is the agent Mary. However, if the verb does not select an agent the subsequent case frame in the hierarchy to represent the subject will be the instrument as shown in proposition 2. Or else, if the verb does not take up an instrument the object will be promoted as subject like in proposition 1.

Fillmore continued developing his theory, which in 1976 was newly designated by the name *Frame Semantics.* It also concentrates on meaning rather than form, whereby a word (not only verbs) specifies the frame from which meaning is viewed. This theory of lexical meaning takes as a premise that a word encodes a concept, which can be interpreted when establishing the relations to other words. Thus, semantic frames are the common backgrounds of knowledge against which the meanings of words are interpreted.

This descriptive framework gives account for the interrelations between language structure and conceptual structure, whereby the meaning of a word that represents a category of experience – specifying a certain perspective from which the frame is viewed – comes with understanding the conceptual system. This can be exemplified with Fillmore's (1977, in: Saeed, 2009:390) famous commercial event frame:

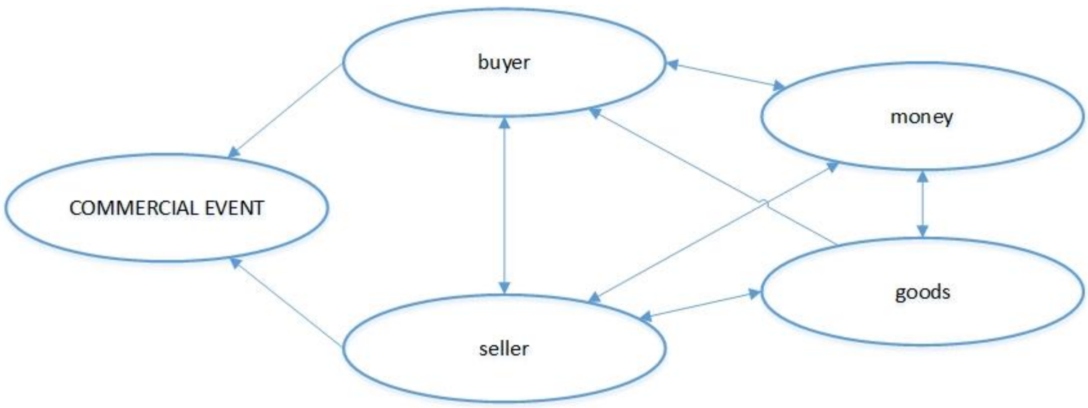


Figure 2: Exemplification of the commercial event frame (according to Fillmore 1977)

The commercial event depicted above shows that the agent **seller** sells to a **buyer** the object **goods** in exchange of the instrument **money**. All the words (frame elements) involved in the event specify a different perspective from which the frame is viewed:

- A **seller** sells goods to a buyer.
- The **buyer** buys goods from a seller.
- The **goods** cost money.
- The **seller** charges the **money**.
- The **buyer** pays money to the seller.

This commercial event evokes the 5 verbs (to sell, to buy, to cost, to charge and to pay) within one common frame by establishing the interrelations between the different frame elements (in terms of situational roles – unlike case grammar, where they are designated as semantic roles). In this regard, Fillmore argues that understanding the meaning of a word presupposes the understanding of the whole frame in which the word is circumscribed. Consequently, when introducing any one concept into the system all other become immediately available.

1.3 Language from a semiotic viewpoint

The next interesting approach towards the aims of this thesis is called semiology (from the Greek semeîon, 'sign'), which was introduced by Ferdinand de Saussure (1857-1913) who is known as the founder of linguistics. Nowadays, the concept behind de Saussure's ideal is referred to as semiotics which, just like linguistics, is broken down into syntactic (relation among or between signs), semantics (relation between signs and the objects they refer to) and pragmatics (relation between signs and the communication context).

De Saussure's definition for *sign* can be seen as the two sides of the same coin, i.e. the *sensorial image* ('image acoustique') of a word ('**signifiant**' – signifier in EN) on the one side, and the content of the word ('**signifié**'– signified in EN) on the other side. The signifier behaves as a pure arbitrary sound chain that brought up together with the corresponding conception of an object (concrete or abstract) makes the linguistic sign (cf. Pelz 1996:43-45).

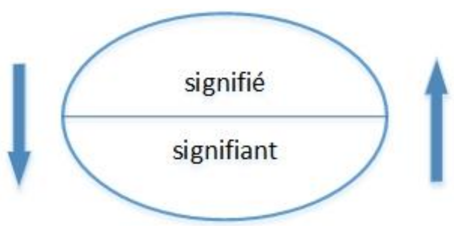


Figure 3: A linguistic sign according to de Saussure

De Saussure argues that a linguistic sign – as part of a language – can be assigned a certain sensorial image to reflect a certain concept by comparing and contrasting these 2 sides to those of another linguistic sign (i.e. difference between things is what creates meaning). For instance, the signifier for and signified of a *chair* differ from the signifier for and signified of a *table*. In the same way he claims that a word can be attributed meaning when it is used in discourse.

Another contribution de Saussure made to this respect is the differentiation between 'langue' (language) and 'parole' (speech/ to speak). The act of speaking can be seen as the result of a social and an individual fact. The former implies the common language (code) a community disposes of as a medium of communication and the latter is the competence a person has to making use of that common language – revealed namely through the speech. Thus, these 2 facts cannot be conceived separately but as a whole: for language to occur, a person chooses from the available lexicon a particular linguistic sign to express his thought in the form of a sound. (cf. Pelz, 1996: 57-61).

Taking de Saussure's static model as inspiration, Ogden and Richards (1923) developed a more dynamic one in the form of a triangle adding a third component to the linguistic conception of a sign.

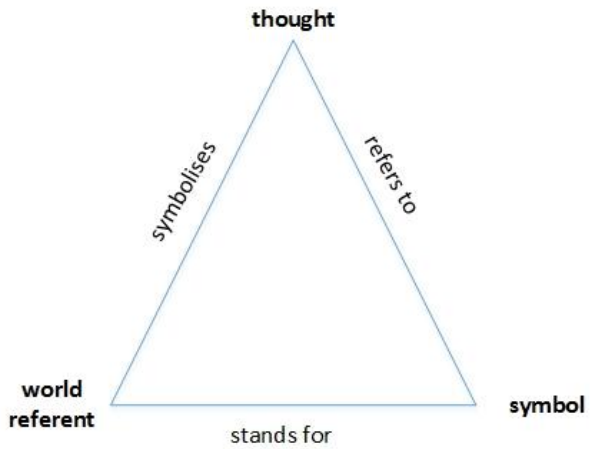


Figure 4: Semiotic Triangle from Ogden & Richards

As it can be seen in figure 4, the third component is the world referent that further specifies the meaning of a word. The relationship between a symbol and the world referent can only be established via the thought. Unless a word is used by a speaker to refer to an object of the world, its meaning cannot be directly conceived. Consequently, the linguistic sign comes to be because it is recognized and understood as such enabling its interpretation to define its meaning. It could be argued that there is some influence from Humboldt, as the world referent is nothing but the image of a thing in the framework of a person's worldview. For being the basis for terminology, this denotational model will be explained more in detail in the next chapter, where also the two different perspectives (onomasiological and semasiological) for studying the relationship between words and meanings will be introduced.

Continuing the timeline of the most prominent contributions regarding language from a semiotic point of view, the Austrian industrialist and later acknowledged terminologist Eugen Wüster (1898-1977) proposed another relevant model for terminology that is based on the semiotic triangle from Ogden & Richards.

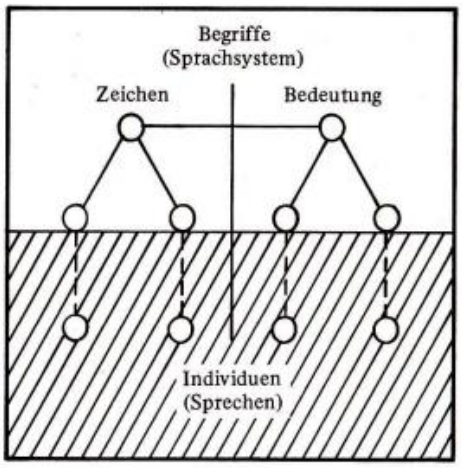


Figure 5: Four Fields' Word Model from Eugen Wüster (in: Arntz and Picht, 1991:39)

In terms of de Saussure, it could be argued that this model takes as well account of the notion 'langue' – depicted in the white section of figure 5 – and 'parole' – lined section at the bottom of the figure. Wüster's explanations (1969) clearly show how this four fields' word model takes the triangle from Ogden & Richards as basis:

> Die obere Hälfte des Schaubildes entspricht dem Sprachsystem. In ihm ist jeweils einem Begriff –d.h. einer „Bedeutung"– ein anderer Begriff (und zwar ein Lautbegriff oder in Schriftbildbegriff) als Zeichen („Bezeichnung") bleibend zugeordnet. […] Die untere Hälfte des Schaubildes stellt die wahrnehmbare Wirklichkeit dar. In ihr entsprechen jedem Begriff viele individuelle Vertreter („Realisierungen"). (in: Arntz and Picht, 1991:40)

Wüster recognises the value of a concept as a meaning within the system of a language, whereby the concept is assigned a symbol that can be either its phonetic or written form. Furthermore, he states that any one concept is realised according to the individual perception of reality. In short, our interpretation of the world is primarily transmitted and conveyed through verbalization.

1.4 Natural language, formal language and semantics

Since the focus of this thesis lies on terminology and the Semantic Web, it is essential to address the dichotomy between natural language (hereinafter NL) and formal language (hereinafter FL) as well as the notion semantics.

Although the concept *language* has been analysed in depth so far, it will be now examined from the viewpoint of the International Organization for Standardization (ISO). To start with, the broader term l*anguage* is defined in ISO 5127:2001 as a 'system of signs for communication usually consisting of a vocabulary and rules', whereby the same norm explains that a *sign* is 'any physical phenomenon interpreted to convey a *meaning'*. Just like the different linguistic theories above suggest, the essence of this definition lies in the fact that language cares for communication by means of signs. The distinctive characteristics that make up natural and formal languages can be appreciated by further narrowing the general concept. Language is considered natural when it is naturally given or acquired, i.e. NLs have grown historically, as it is revealed in ISO 5127:2001: 'which is or was in active use in a community of people, and the rules of which are mainly deduced from the usage'. On the contrary, a formal language is one it is invented for some purpose or other (defined in ISO/IEC/IEEE 24765:2010 as 'a language whose rules are explicitly established prior to its use')[1]. While it is indisputable that NLs have always existed, since they satisfy the basic human need of communicating with each other, it is admissible to think that FLs were invented at a specific time (Chomsky's formal grammar theory of 1950 can be seen as an example of FL modelling English) and that they exist to fulfill specific needs. In general, FLs

[1] It should be noted that ISO/IEC 2382-1:1993 (in: ISO/IEC/IEEE 24765:2010) considers *artificial language* as a synonym of formal language.

are constructed language systems based on logic that can be characterized for being clear, explicit and easy to verify.

Interestingly and despite the differences, NL and FL have much in common in so far as they can both be analysed and they make a distinction between syntax and semantics. In the framework of this research, semantics plays a major role for studying the agreed upon meaning that comes instinctively when relating one particular word to a particular fact in the world. There is no doubt this is one of the challenges terminology tries to overcome and next chapter encompasses semantics from a traditional terminological viewpoint. In an analogous manner to the categorization of language as natural or formal, semantics offer two different approaches to studying *meaning,* which will be hereinafter treated as traditional and formal semantics. On account of that, the German logician, mathematician and philosopher Gottlob Frege (1848-1925) claims that meaning can be appreciated from an *extensional* as well as from an *intentional* point of view (cf. Ebert et al., in: Carstensen *et al.*, 2010:330-339). For example, by paraphrasing the meaning of the word *house* as 'building to live in', it is possible to establish that the word (in discourse) has a referent in the world ('denotatum') thanks to this meaning. However, reference objects do not make up the *linguistic meaning* of a word; on the contrary, the linguistic meaning naturally concerns the lexeme ('abstract unit generally associated with a set of forms sharing a common meaning'[2]). Although referencing is indirectly implied in the meaning of a word in discourse (as instance), the reference object itself is not included. Consequently, Frege distinguishes between *reference* ('Bedeutung') with respect to a word as truth value in this world, and *sense* ('Sinn') with respect to a proposition as truth condition (cf. ibidem).

In terms of the author of this research	In terms of Frege	Reference	Scope
linguistic meaning = sense	'Sinn'	Intentional meaning	the way an expression represents the referenced thing
conceptual meaning = meaning	'Bedeutung'	Extensional meaning	results from things, which are denoted by an expression in a possible world

Table 1: Interpretation of G. Frege's theory of meaning

[2] As defined in ISO 24613:2008, Language resource management – Lexical Markup Framework (LMF).

Arguably, traditional semantics (applied to NL) implies talking about the world by showing how words relate to reality (intentional meaning) and that formal semantics derives from language being a reflection of the conceptual structures conventionally agreed (extensional meaning).

1.5 Summary

This chapter offered a brief look at different perspectives for analysing words, concepts and thinking as part of language. The principle of linguistic relativity (section 1.1), expressed firstly by Humboldt but typically associated with Sapir-Whorf, holds that the structure of a language imposes restrictions on how speakers of a nation conceptualize their world and consequently, on their way of thinking. This led to the Generative Grammar rationale of Chomsky. On the one hand, he identified a common conceptual factor in the linguistic system (deep structure) and a distinctive one that characterises the different languages (surface structure). Although not quite complete, these concepts are not only relevant for being an attempt to formally analyse language but also because they involve the idea of a store of lexicalized concepts with which a speaker of a language can associate both structures.

Language as a semiotic system (section 1.2) is introduced by de Saussure's paradigm that considers a linguistic sign having a sensorial image (signifier) and a content value (signified). Similar to Chomsky, de Saussure argues that language should rather be explained as an individual fact ('parole') and not a social one ('langue'). This research does not take into account; whether a speaker's knowledge (competence) or the predetermined language use (performance) – aversely favoured by these two linguists – should be the focus for analysing language, because the latter cognitive approaches have naturally rejected them. In this regard, Langacker explains syntax and lexicon as symbolic systems in that they are motivated by general cognitive processes. Fillmore later developed a grammar that also takes semantics into consideration, which anticipates in turn the syntactic value of a proposition. This results finally in Fillmore's well-known Frame Semantics' proposition that concentrates more on meaning than form while stating that a frame is the common background of knowledge against which the meanings of words can be interpreted.

How words are used to refer to things was demonstrated with the referent component proposed by Ogden and Richards. This served not only to introducing the foundations of traditional terminology but also to explaining the theory of meaning as proposed by Frege. In this regard, traditional semantics was associated with natural language and formal semantics, with formal languages (representational meaning).

2 Terminology science

At the same time, new concepts and abstractions flow into the picture, taking up the task of describing the universe without reference to such time or space – abstractions for which our language lacks adequate terms. Benjamin Whorf[3]

2.1 Principles of terminology

In 1931 Eugen Wüster laid the theoretical foundations for terminology with his first major publication 'Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik' (International Standardization of Language in Engineering, Especially in Electrical Engineering). Five years later, on the grounds of Wüster's contribution, a Technical Committee (TC 37)[4] dedicated to terminology was founded in the framework of the International Federation of the National Standardizing Associations (ISA) (cf. Felber; Budin 1989:243).

The interdisciplinary and multidisciplinary character of concept-centred terminology is explained by Wüster in 1974 in his article 'Die allgemeine Terminologielehre – ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften' (The general theory of terminology – a border area between linguistics, logic, ontology, computer science and objective sciences) (cf. ibidem, p. 19). That Wüster is considered the founder of terminology is probably based on many of his achievements that include his unfinished work 'Einführung in die allgemeine Terminologielehre und terminologische Lexikographie' (Introduction to the General Theory of Terminology and Terminological Lexicography), which was published posthumously in 1979 (cf. ibidem, p. 58).

All aspects of terminology will be hereinafter introduced with a special focus on the existing norms as defined by the International Organization for Standardization (hereinafter ISO), which has been the official appellation of ISA since 1947. Naturally, norms evolve over

[3] http://www.brainyquote.com/quotes/keywords/concepts_4.html
[4] The scope of the ISO/TC37 is standardization of principles, methods and applications relating to terminology and other language and content resources in the contexts of multilingual communication and cultural diversity (description as provided by the responsible committee – ISO/TC 37 Terminology and other language and content resources –).

time; therefore, any references made to norms will cover mainly those as of 2000 for keeping in view the most recent ones.

The science of terminology is defined by ISO 1087-1:2000 as the 'science studying the structure, formation, development, usage and management of terminologies in various subject fields'. The same norm explains terminology as being a 'set of designations belonging to one special language', whereby *special language* is considered a 'language used in a subject field and characterized by the use of specific linguistic means of expression'.

Note that in the above definition for *terminology* the term used is *designations* and not that of *words*. *Words* are part of a person's general knowledge, that is to say, they are normally used in the context of *Language for General Purpose* (hereinafter LGP) and they are studied using a semasiological perspective, namely studying first the linguistic form of the sign to then see how it is understood. Typically, the relationship between *words* and *meaning*[5] is always found in dictionaries. For example, when looking for the word *table*, the following distinctions associated with this general word were found on the *English Longman* dictionary: 1. furniture, 2. list, 3. maths, 4. group; and all expressions that contain it such as 'to turn the tables (on sb.)'[6]. It can be concluded that dictionaries indicate the meaning and use of a word by means of paraphrases and examples.

In the context of *Language for Special Purpose* (hereinafter LSP) it is though common to talk about *designations* and their relationships to *concepts*. This kind of information is normally found in termbases where a *concept* is the basis of an entry (how a terminological entry looks like will be treated in section 2.3) and the *designation* can be seen as the outcome. As pointed out in section 1.3, the semiotic triangle from Ogden & Richards reveals the role of each of these terminological components.

[5] Note that no difference between *sense* ('Sinn') and *meaning* ('Bedeutung') in terms of Frege's is done here. Unless clearly specified, the term *meaning* will be used throughout the research in its traditional 'arbitrary' way.

[6] Example based on the information for the word 'table' found on the dictionary of contemporary English *Longman* (1995), Third edition. Page 1466.

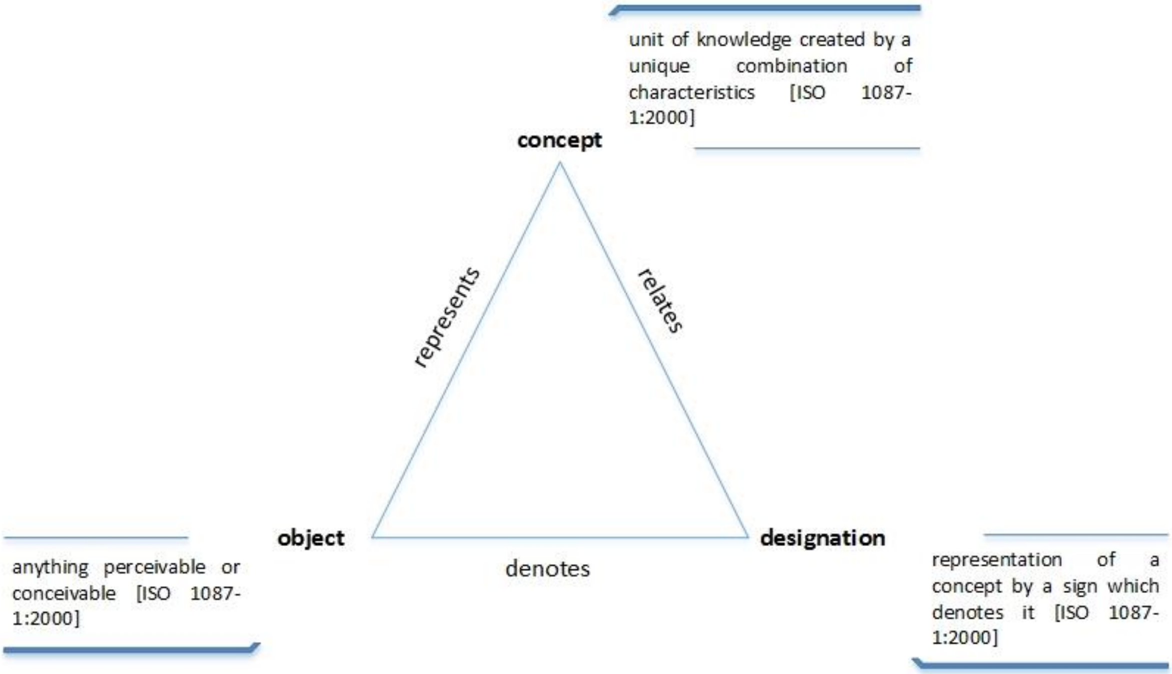


Figure 6: Semiotic triangle for terminology purposes

Figure 6 illustrates that any **object** in the real world is represented in our minds by a **concept**, which has a relating **designation** that simultaneously denotes the object. It can be alleged that organizing information in such a concept-oriented manner is closely related to the fact that specialist would first make new discoveries and later agree on how to designate them. Thereupon, terminology science deals with specialized language and the traditional approach to terminology work adopts an onomasiological perspective, namely asking how an object is delimited to creating its corresponding mental representation and finally attributing it a designation.

To conclude the dichotomy *words-meaning* and *concepts-designations*, lexicographers are concerned with the former while terminologists deal with the latter. The following table summarizes the main differences of these two close-related sciences:

	Terminology	**Lexicology**
Motivation	New *terms (designations)*	New *words*
Science	A science in itself	Branch of Linguistics
Language use	Language for Special Purpose	Language for General Purpose
Method. approach	Onomasiological	Semasiological
Repository	termbase (concept-oriented)	dictionary, glossary (word-oriented)

Table 2: Dichotomy between *terminology* and *lexicology*

Rendering our attention exclusively to terminology now, it is necessary to analyse further in detail the 3 components of the semiotic triangle.

As shown in figure 7 below, at the apex of the triangle is the **concept**, which is the basis of terminology, and by definition a 'unit of knowledge created by a unique combination of characteristics' (ISO 1087-1: 2000). It should be noted that concepts are independent of the terms they are designated with and the context they are used in. Although it might be hard to imagine that a concept exists when there might be no designation for it, this can again be easily appreciated in our previous example of specialists discovering new matters. On this ground, concepts are at a higher level of abstraction and the process involved in 'generating' this mental representation is known as *conceptualization*. Once the corresponding abstraction of the objects involved took place, concepts can be defined and described by their characteristics resulting in a distinction between general and individual concepts. For instance, the animated **object** *dog* can be categorized as 'a carnivorous mammal with all its possible variations in terms of breed, size, colour, etc.'; which corresponds in turn to the unit of knowledge representing a **general concept**. Further to this example, within the specific subject field *American film industry*, the also animated object *Beethoven* would be correspondingly categorized as 'the St. Bernard dog of the Newton family of the comedy film (1992)', therefore is here the **appellation** Beethoven an **individual concept**.

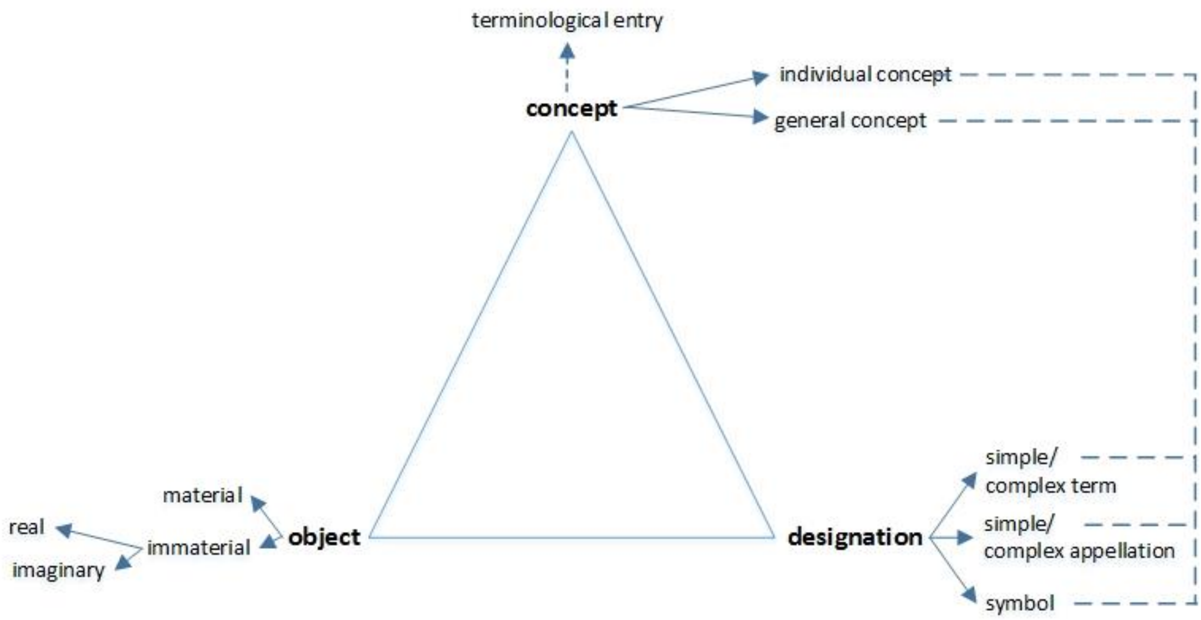


Figure 7: Terminological considerations

In contrast to those theories presented in Chapter 1 that consider thought and language interdependent, concepts are defined by ISO as not bound to any particular language. That is to say, concepts remain immutable no matter in which language they are designated, since – as pointed out before – they come to be through a high level conceptualization of the objects they refer to. However, the ISO norm 1087-1 goes in line with Humboldt by stating that concepts are 'influenced by the social or cultural background which often leads to different categorizations' (ISO 1087-1: 2000).

According to the normative terminology, an *object* is 'anything perceivable or conceivable' (ISO 1087-1: 2000). As it can be seen in the flowchart below, **object properties** and **concept characteristics** play a crucial role in terminology.

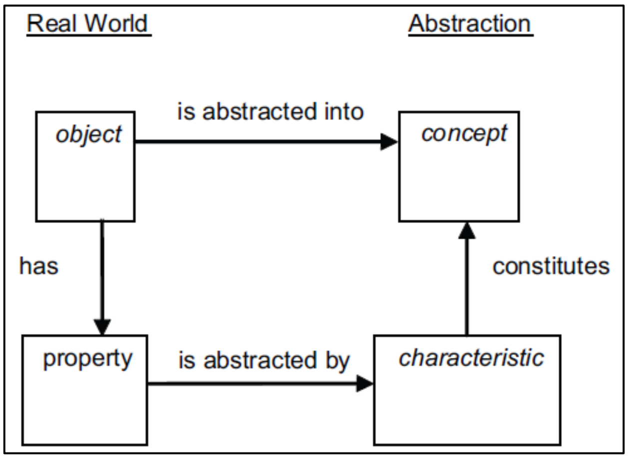


Figure 8: Nature of characteristics (ISO 704: 2009)

The ISO norm 704 points out the object whose properties need to be first identified and determined in order to classify the object(s) of the real world (**material** or **immaterial**) as part of a class (or subclass) of a particular domain. Additionally, the real world object has to be translated into a higher level of abstraction to form the concept. The properties of the object(s) are abstracted as unique combination of characteristics, which in turn creates the unit of knowledge (concept). Because these concept characteristics are naturally given, they are known as *characteristics by intension* and are the pillar of terminology work.

> The set of *characteristics* that come together to form the *concept* is called the *intension* of the *concept*. The set of *objects* conceptualized as a *concept* is known as the *extension* of the *concept*[7] (ISO 704: 2009, p. 6).

[7] As proposed by Frege (refer to section 1.4) and, as it will be later seen in detail, in the domains of IT and logic, it is common to focus on the extension of concepts as it provides the means to identifying necessary, sufficient and essential characteristics of objects.

It is worth mentioning that these characteristics by intension need to be further scrutinized to establish which of them are considered *delimiting* and/or *shared* in relation to other concepts.

To finalize describing the components of the semiotic triangle it is necessary to examine what *designations* are for. As its corresponding verb suggests, *to designate* is 'to represent something in a particular way, for example with a **sign** or **symbol**'[8]. Thus, in the context of terminology, a *designation* is used to represent concepts by linguistic (**terms** or **appellations** – see figure 7) or non-linguistic means (symbols[9]) (cf. ISO 704: 2009). Finally, the term *term* (verbal designation of a general concept in a specific subject field – ISO 1087-1:2000 –) will be henceforward used to make allusion to any kind of designations.

2.2 Concept systems

As stated in ISO 704: 2009, a *concept system* is 'a set of concepts structured according to the relations among them'. The detailed analysis of characteristics does not only allow to determine the specific relation a concept has to another one, but also to fix the exact position of those domain concepts within the system. For this, the relationships between concepts can be of *hierarchical* or *associative* nature. Whereas the latter is of a pragmatic type as it cares for the establishment of concept relationships on the basis of thematic correlations, the former is more interesting as it is additionally subdivided into *generic* (.) and *partitive relations* (-). On the other hand, unlike part-whole-relations, generic relations express the passing of characteristics from one concept to its subordinates. Accordingly, concept systems are classified as *generic concept systems*, *partitive concept systems*, *mixed concept systems* (include both types of hierarchical relations) and *associative concept systems*.

It is important to remember that concepts are designated by terms according to the specified domain. To build any kind of concept system it is essential to first choose the domain of interest and then analyse the interlinked concepts within that particular domain.

[8] http://www.macmillandictionary.com/dictionary/british/designate_1
[9] 'designation by means of letters, numerals, pictograms or any combination thereof' (ISO 5127:2001)

Because the world object participates in the previously explained conceptualization process (where properties are abstracted to create the concept), a different abstraction of the same object(s) comes to pass according to the domain involved; hence, a same object in different domains may be characterized by diverse intensions and extensions (cf. ISO 704:2009).

Figure 9 below shows the concept system of ‘ontoterminology’, which was especially modelled to later create its corresponding terminological database and ontology.

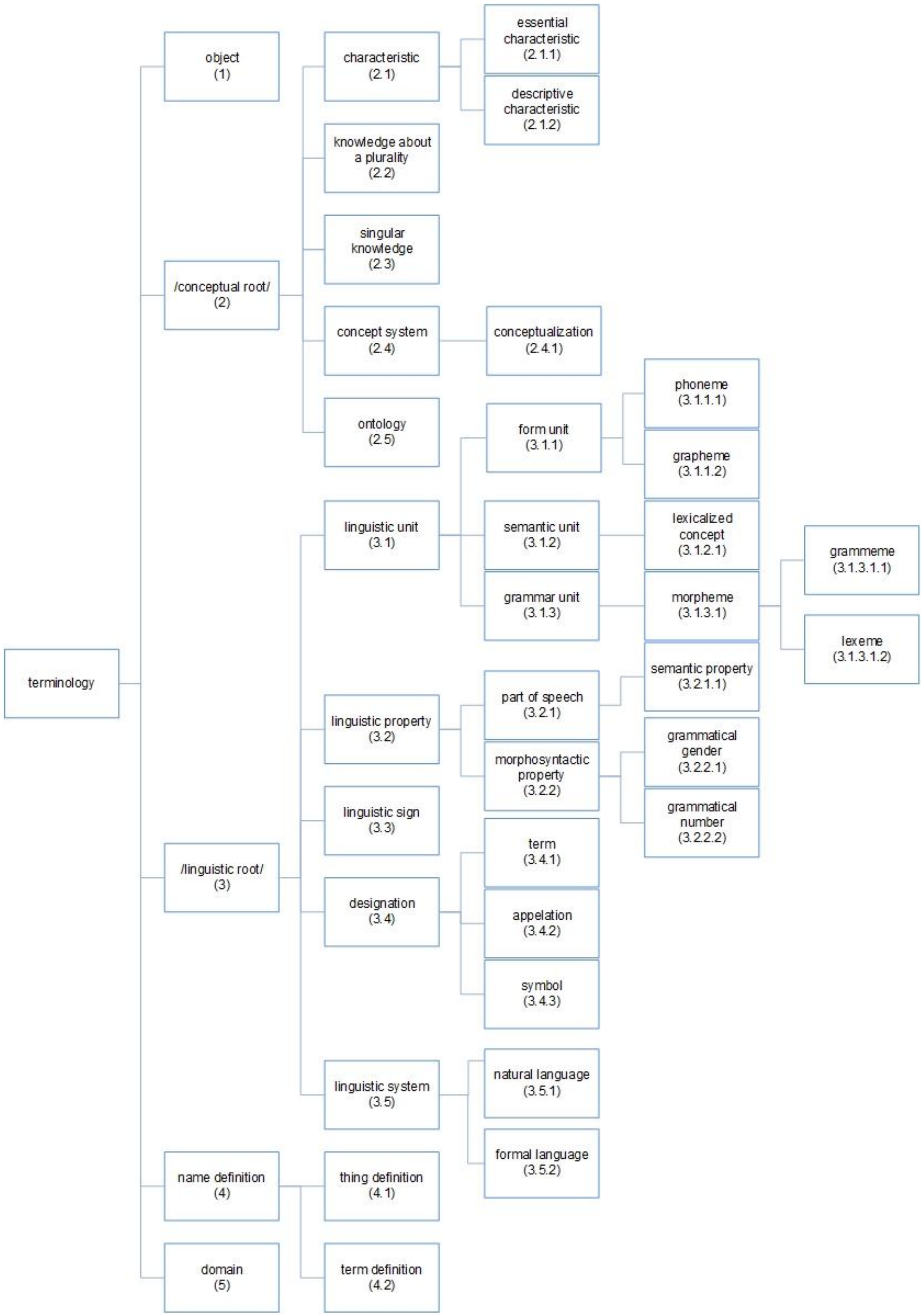


Figure 9: Concept system modelling the 'Ontology of Ontoterminology'

2.3 Termbases

After having assigned each concept of the domain a specific place in the system, the information so far gathered should be ordered and recorded in a repository – and further enriched – for later use. Such repositories are best known as termbases, which are namely databases (hereinafter DB) conceived for terminology management.

The way a termbase (hereinafter TDB) is designed varies radically from business analyst to business analyst. The author of this research believes this is the case because business analysts are not necessarily specialists in terminology and, although business needs may differ, the metamodel of a TDB should typically follow the principles of the semiotic triangle. In doing so, a better exchange of terminological entries between different terminology management systems (hereinafter TMS) can be achieved. In this regard, Schmitz[10] states that a TDB has to be defined in such a way that a concept is autonomous of its designation(s) in any language, whereas the latter has to also be independent but yet conserve its concept orientation, i.e. one terminological entry (hereinafter TE) per **concept**.

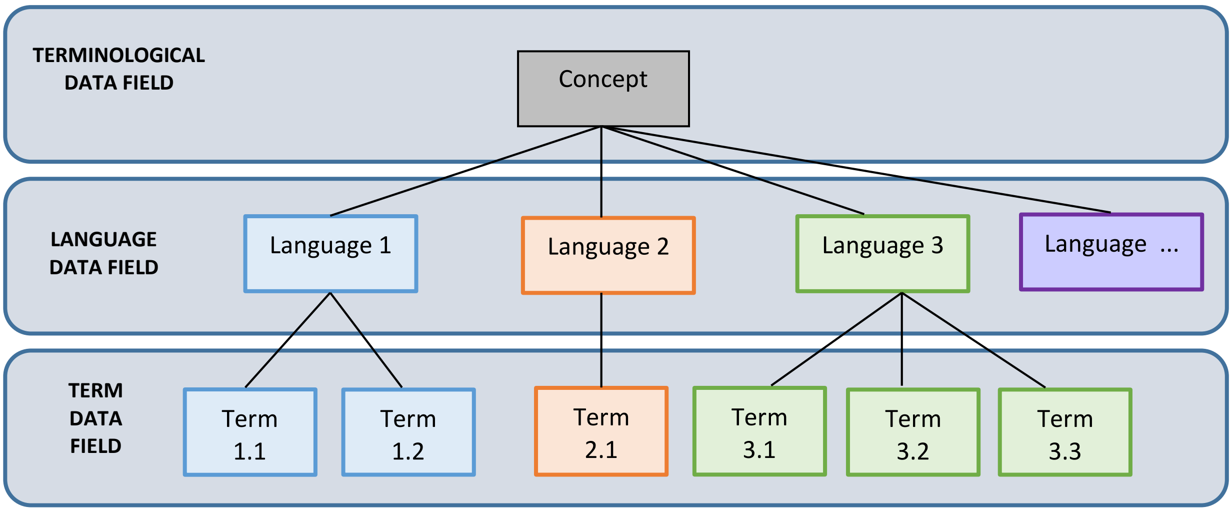


Figure 10: Concept-orientation and term autonomy of a TE

Although the figure above shows that a concept can be represented by more than one **term** per **language**, best practice is to apply the univocity principle according to which a

[10] Course material of the lecture *Terminology Science* ('Terminologiewissenschaft') offered at the *Cologne University of Technology, Arts and Sciences* (2012).

concept should ideally not be designated by synonymous terms to avoid ambiguities and so guarantee clear communication.

The specification of the different **data fields** illustrated in figure 10 – for which the result will be various *data categories*[11] – is also part of the creation process of a termbase. Equally essential is that the TDB cares for *data granularity* and *data elementarity*. According to ISO 26162:2012, the former refers to the 'degree of precision of data' and the latter implies the 'principle whereby a single data field shall contain only one item of information'. For example, when modelling a TE it is recommended that the **term data field** permits a fine definition (hierarchically ordered) of various data categories such as *part of speech*, *grammatical gender* and *grammatical number* and not simply one data category called *grammar* to cover all. On the other side, data elementarity is achieved like depicted above in figure 10: at the *term data field* there are various fields for each language so that each of their corresponding designations (be it a term, its synonym, its abbreviation or the like) is saved under a separated field and not altogether. Another aspect is the simplicity or complexity of data categories. Complex data categories include open and closed ones that are respectively characterised for including data type with unforeseeable values (data type normally defined as *string* – text) and data type with predetermined values (data type normally defined as *string/integer* – pick list). In contrast, a simple data category has no set of valid value meanings (cf. ISO 12620:2009, 3.1.12). As it will be seen later, the typology of data categories is normally defined when specifying the *logical data model* of a database.

The above introduced criteria can be seen as the minimum requirements that a TDB has to fulfil in order to facilitate the data exchange between different systems. It is recalled that *formal* implies the use of an invented language, in this case, for information systems to enable the communication and processing of information. As wit as previously anticipated, behind a DB there is a metamodel, which aims to formally define what data[12] will be contained and how a datum will be used and related to one another. A *data model*[13] sets the underlying structure of the subject field in question (here: terminology) and is the result of

[11] 'result of the specification of a given data field' (ISO 1087-2:2000)
[12] 'representation of information in a formalized manner suitable for communication, interpretation and processing' (ISO 1087-2:2000)
[13] 'abstract model that describes how data is represented and used' (ISO 25964-1:2011)

the processes involved in designing/describing a TDB – identification and conceptual organization of information. Since information systems are normally designed in such a way that data and data processing are separately conceived, their design method commonly consists of 3 different stages: the conceptual level, the logical or organizational level and the physical level of a data model.

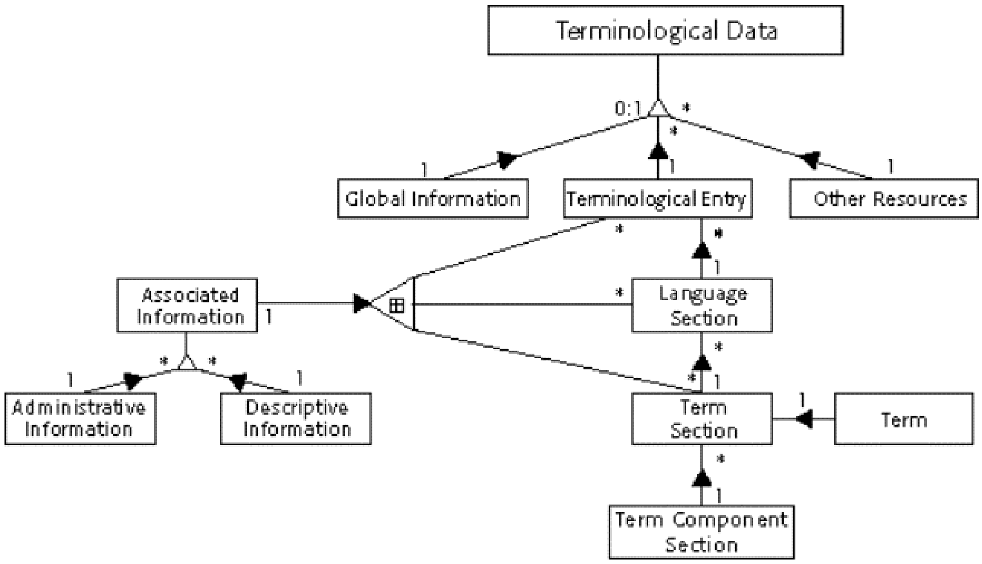


Figure 11: Standard conceptual data model of a TDB (metamodel from ISO 16642:2003)

The figure above shows the first design stage, which represents a high level abstraction of the entities that define the structure of a terminological database as dictated by the Terminological Markup Framework (TMF)[14]. Each entity[15] is unique as well as it is to be described by a set of data, which are included in the second design level that concerns namely the *logical data model*[16] (hereinafter LDM) of a TDB. Figure 12 shows an example of the LDM (based on relations consisting of attributes as well as identifiers and data carried by certain associations) that could represent one part of the conceptual model illustrated above in figure 11 while indicating which entity's data is deemed identifier so that it can be source of the functional dependences (hereinafter FD) (cf. Neumann, 2012).

[14] TMF is a metamodel that allows the definition of different terminological markup languages for specific purposes.

[15] 'any concrete or abstract thing that exists, did exist, or may exist, including associations among these things' – an entity exists whether data about it are available or not – (ISO 11179-1:2004, in: ISO 23081-2:2009)

[16] 'data design, that takes into account the type of database to be used, but does not consider means of utilisation of space or access' (ISO/TR 17185-3:2015)

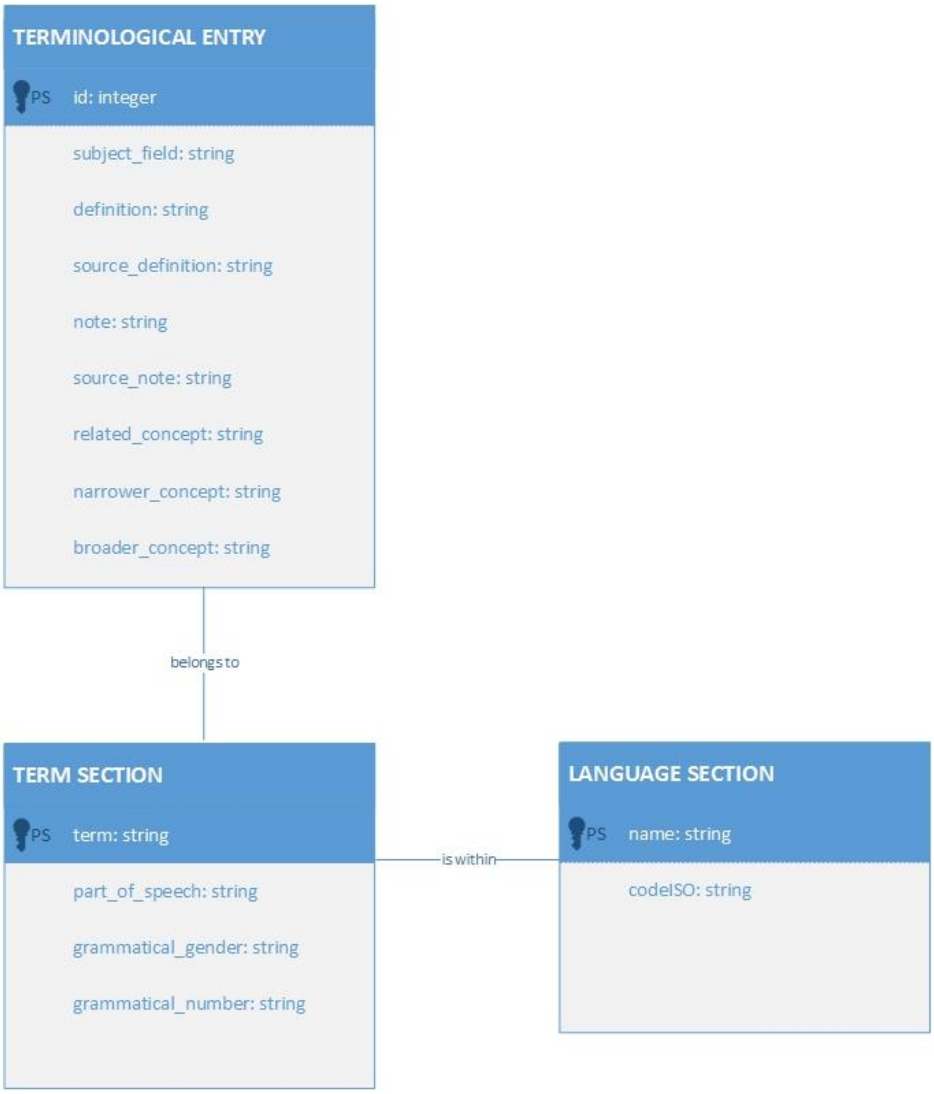


Figure 12: Logical data model of a TE

This figure shows the set of data that describes each entity as well as the corresponding entity's identifiers[17]. For example, the entity **term section** is identified by a **term**, which is characterized by **part_of_speech**, **grammatical_gender** and **grammatical_number**. In this way, this entity's identifier (term) is source of the FD in the association **is within** with the entity **language section** – whose identifier **name** is considered the object of the FD. Considering for example figure 10 in relation to this explanation it is possible to say that only instance **2.1** of the entity **term section** can be used to find **2** as instance of the entity **language section** and not that of **1** or **3**.

[17] 'mnemonic string used to refer to the data category' (ISO/IEC 11179-3 in: ISO 12620:2009)

Finally, the physical data model involves implementing the 2 preceding data models in the database of choice. However, this exact implementation is not relevant for this research and will therefore not be covered.

For the purposes of this thesis, the TDB that corresponds to the concept system presented in figure 9 was built in SDL MultiTerm 2011 and takes the following points into consideration:

- Since concepts are considered units of knowledge, they are defined in English and identified by a unique ID at the terminological entry level (i.e., there is no definition at the term level for every other language).
- While the TEs contain terms in English, French and German, the respective concepts are defined in English for being the world's most widely used language in informal as well as in formal communication.
- Since the richness in natural languages is inevitable and even precious, the principle of univocity was not taken into consideration. A minimum amount of synonyms was included while indicating, where possible, which synonym has normative character.
- For the intended comparison with an ontological knowledge base, data categories at term level were restricted to the essentials. Specifically, this data field admits the categories *note*, *context*, *term type, normative use* (Boolean data type for implicitly indicating that the term was found in an ISO or DIN norm), *grammatical number* and *grammatical gender*. The distinction between the data categories *syntactic unit* and *part of speech* follow the descriptive terminology as proposed by the *General Ontology for Linguistic Description* (GOLD 2010)[18].
- When no term equivalent was found for French or German, a prospective term was proposed and correspondingly indicated under the data category *note* at term level. Naturally, no context is available in these few cases.
- When no already established definition was found, a definition was recreated from available explanations. This was explicitly indicated in the *note* data category pertaining to the linguistic independent level.

A sample entry is included in *appendix A* and the TDB is to be found on the attached CD.

[18] See http://linguistics-ontology.org/gold

Languages
English
French
German

Entry Structure	Mandatory	Multiple
Entry level		
ID		•
Subject Field	•	•
Definition	•	•
Source		•
Note		•
Source		•
Related Concept		•
Narrower Concept		•
Broader Concept		•
Language level		
Term level		
Note		•
Source		•
Context		•
Source		•
Normative Use		
Term Type		
Part of Speech		
Syntactic Unit		
Grammatical Number		
Grammatical Gender		

Descriptive Fields

Name	History	Type	Picklist Values
Broader Concept		Text	
Context		Text	
Definition		Text	
Grammatical Gender		Picklist	feminine | masculine | neuter
Grammatical Number		Picklist	singular | plural
ID		Text	
Narrower Concept		Text	
Normative Use		Boolean	
Note		Text	
Part of Speech		Picklist	noun | verbal | adjectival | adverbial
Related Concept		Text	
Source		Text	
Subject Field		Picklist	Terminology | Information Technology | Knowledge Engineering | Linguistics | Information Science
Syntactic Unit		Picklist	noun phrase | verb phrase | adjective phrase
Term Type		Picklist	short form | abbreviation | acronym

Figure 13: TDB's definition

2.4 Summary

The first part of this second chapter (in straight relation to section 1.3) presented some facts regarding the founder of terminology Eugen Wüster, and the importance of his contributions that gave rise to treating terminology at international level and with normative character. Precisely for that reason, most of the topics here described find their basis on the following norms:

- ISO 704 (2009): Terminology work – Principles and methods
- ISO 1087 (2000): Terminology work – Vocabulary
- ISO 26162 (2012): Systems to manage terminology knowledge and content – Design, implementation and maintenance of terminology management systems
- ISO 12620 (2009): Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources

The comparison between terminology and lexicography aimed at highlighting the different perspectives from which the encoded linguistic knowledge of words (sense) and designations (meaning) is considered. Most importantly, be they general (semasiological approach) or specialized (onomasiological approach) 'words', they are both linguistic units with linguistic properties.

In relation to the conceptualization process to which objects of the real world are subjected, it was noted that terminology work focuses on the characteristics by intension.

As it was brought to light, the main purpose of terminology is to clarify concepts that are designated by terms, which are in turn used to represent concepts when communicating. On the other hand, it was also revealed that concepts do not exist isolated but in interrelation while forming concept systems (section 2.2) within the involved subject field. Such conceptual systems are classified by ISO as *generic*, *partitive*, *mixed* or *associative*. In view of the intended comparison between a traditional concept-oriented termbase and an ontology (as a conceptual system underlying a knowledge base), the end of section 2.2 illustrated a concept system reflecting the concepts dealt with in the ontology.

The principles of terminology work were dully presented according to the corresponding norm for the conception of a *termbase* (section 2.3). Although such a repository is *concept-oriented*, it offers *term autonomy* while caring for *data granularity* and *data elementarity*. Furthermore, DB are characterised for having an underlying structure that calls for a formal definition at three levels, which are commonly known as *conceptual*, *logical* and *physical data models*. In contrast to conceptual modelling within the framework of terminology (where the focus lies on the system of concepts itself and its relationship to other concepts – concept intension), it was shown that in the framework of information technology the modelling implies the formal definition of a TDB. More specifically, concept data modelling has to do with the structuring of the data itself (specification of data fields) and not any concept.

3 About semantic technologies

Properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.

Berners-Lee, Hendler and Lassila (2001)

3.1 The Semantic Web – Web 3.0

Semantics does not only play a crucial role in human communication but also for machines just as it is stressed by the appellation of the new Web itself. By means of formal semantics, the Web 3.0 seeks to make the conceptual meaning of natural language expressions readable by machines. For a fact, this concept was proposed by Tim Berners-Lee and Ora Lassila who also coined the term *Semantic Web* for it and described it as an extension of the *Web 2.0,* 'in which information is given well-defined meaning, better enabling computers and people to work in cooperation' (in: Casellas, 2011:4).

Going back in time, the *Standardized Generalized Markup Language* (SGML), which is considered the first description language created to electronically save the layout and structure of documents, became a norm (ISO 8879) in 1986. Not later than 1989, Berners-Lee (physicists and computer scientists) invented the famous *World Wide Web* and introduced the well-known *Hypertext Markup Language* (hereinafter HTML)[19]. Shortly after it was necessary to separate data structure from data layout, and the formal *eXtensible Markup Language* (hereinafter XML) was introduced by the World Wide Web Consortium (W3C) in 1998 as a specification of SGML (cf. Hanke, 2006:7-12). Thanks to XML, which works with tags for document mark-up, it has been so far possible to provide unstructured text with additional information (metadata) that helps structuring and exchanging data between different information systems. However, Berners-Lee and Lassila soon (2001) came to acknowledge that information systems were not capable of accurately understanding each other's data just by sharing a common markup language.

Search engines typically work with those *keywords* chosen by a person who is looking for a particular piece of information on the Web. The problem is that many of the hits will

[19] HTML as document description language serves for graphically displaying information on a browser page (Web 1.0).

plausibly not make any sense of what that person had in mind for the simple reason that there is too much ambiguity in NL. It is no easy task for a person to think up the exact *lexical item* that will be contained in the document he wants to retrieve. The fact that search engines of the Web 2.0 mainly rely on syntactic matching (as linguistic units) is already implied in the term lexical item. For example, while having in mind the concept 'a structure that spans and provides a passage over a road, railway, river, or some other obstacle'[20] a person might introduce in a search engine the keyword *bridge.* The retrieved information will probably include as well documents containing the acronym *BRIDGE* (that stands for *Biotechnology Research for Innovation, Development and Growth in Europe*) or bridge as in *dental plate,* as in *part of a string instrument*, etc. Therefore, it is much desirable for these engines to allow intelligent semantic searches, and literally, the aim of semantic technologies is to structure information in such a way that it facilitates its automated analysis by formally defining the entities in the world (as will be described in section 3.2).

It could be then argued that the approach of the Web 2.0 is semasiological as it follows the principles of lexicography in that it treats words, namely keywords to which many meanings can be attributed. On the contrary, it can be claimed that the Semantic Web proposes an onomasiological framework by concentrating on concepts represented by terms. *Terminological knowledge* (within the meaning of descriptive data) allows in this way to identify polysemic terms and associate their concepts to their corresponding domain knowledge hierarchies. Thus, the exemplified lexical item *bridge* could be then represented by their corresponding concepts in terms of formal knowledge enabling concept-term association within the related domains (i.e. including the needed conceptualization); in this particular case, within the domains of *construction*, *biotechnology*, *dental medicine* and *music* respectively.

The key of such intelligent results lies in the processing of *knowledge* (instead of text – strings), which will be covered in the next section. For this, the languages of the Semantic Web are organized in interdependent layers where each of them profit from the capabilities of the layer below, as illustrated in the following figure.

[20] Definition taken from the EU's multilingual termbase IATE (Inter-Active Terminology for Europe): http://iate.europa.eu/SearchByQuery.do (15-08-2015)

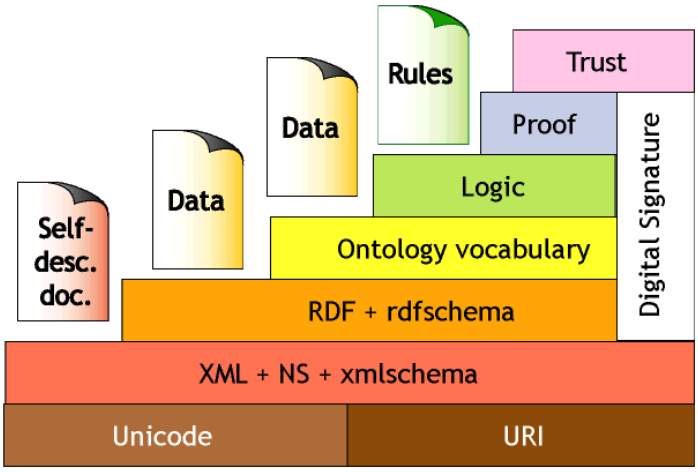


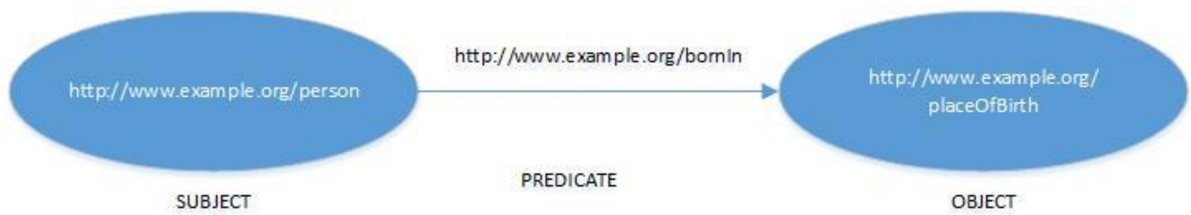
Figure 14: Semantic Web Stack by Berners-Lee[21]

Due to the different language characters there are, **Unicode** is used as standard for encoding international character sets allowing all NL to be used on the Web. On the other side, data can be stored in different ways (by rows, by columns, by cells, etc.) and to guarantee their interchangeability, a server has to be able to find and relate data from different DBs irrespective of where and how the data was saved. In this respect, Allemang and Hendler (2011:7) bring to reflect a characteristic of the Web for which they used the AAA slogan: 'Anyone can say Anything about Any topic', i.e. anyone can write any kind of information on the Web. It is not only that the Web contains an enormous amount of information but also that the storing of this information radically varies. A first measure to deal with this problem was the introduction of *Uniform Resource Identifiers* (**URI**s) that facilitates the distribution of the entities we talk about on the Web. Thanks to them it is possible to globally identify and associate data from one *resource*[22] to the one from another resource. On the contrary, the Web 2.0 uses *Uniform Resource Locators* (URLs) to refer to webpages among a distributed network where webpages can be associated to one another. (cf. Allemang and Hendler, 2011:31-40). In an analogous manner to the individual definition of tags in XML, different Web authors might select different URIs for the same real-world resource. Notwithstanding, the Semantic Web applies a feature to assuming that different persons can refer to a same resource on the Web with different identifiers. This feature is known as *Nonunique Naming Assumption* (cf. ibidem, p. 11).

[21] Retrieved from http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html (05-08-2015)
[22] According to Allemang and Hendler (2011), the term *resource* designates the things in the world and define them as 'anything that someone might want to talk about' in the Web.

On the second layer is the descriptive **XML,** the kind of formal language that characterizes both the Web 2.0 and Web 3.0. Being XML a structure language – that provides a syntax for content structure within documents –, it needs a schema (**xmlschema**) to specify the overall structure of an XML document (thus the annotation **self-descriptive document**) where normally all its elements and their attributes – as well as the relationships between them – are accordingly identified and defined. Yet, this language does not convey any meaning about its own structure. Since XML permits a person to define his own tags (within the meaning of labels), *namespaces* (hereinafter **NS**) are meant to distinguish the meaning between 2 tags that are equally named. For a fact, if two authors define a tag as <product> </product>, but the one means by it the manufacturer of the product, and the other one the reseller of the product, a prefix makes sure that the NS for product can be clearly allocated to manufacturer and reseller respectively: <manufacturer: product> </manufacturer: product> and <reseller: product > </reseller: product > (cf. Hanke, 2006:9-12, 49).

At the third layer of the Semantic Web, data structure *meaning* is expressed by the *Resource Description Framework* (hereinafter **RDF**) that builds on XML principles for data interchange. This simple standard modelling language describes the distributed data on the Web in a syntax independent way allowing a meaningful indexing[23]. It is important to note here that XML is only one of the various syntaxes an RDF-based model can be represented in. It is therefore said that RDF defines data models independently of their syntax. An RDF-based model expresses the meaning of its structure in the form of triples where things in the world are encoded like in a typical syntax analysis (subject, predicate and object) but by means of URIs that provide a machine-readable description. The following statement */Carolina Dunaevsky was born in Córdoba/* will be annotated on a Website in such a way that a person is related to its place of birth.



[23] Indexing is one of the main tasks in the field of Information Retrieval (IR), which, up until the use of semantic technologies, mainly relies on a rather general linguistic analysis of the words contained in documents (cf. Weller, 2010).

Figure 15: Sample of an RDF-triple

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:ex="http://www.example.org/">
    <rdf:Description rdf:about="http://www.example.org/person">
        <ex:bornIn>
            <rdf:Description rdf:about="http://www.example.org/placeOfBirth"></rdf:Description>
        </ex:bornIn>
    </rdf:Description>
</rdf:RDF>
```

Figure 16: XML syntax of the RDF-triple above

The RDF ensures that concepts are understood as such – and not as merely text strings – while being linked to a unique definition that everyone can find on the Web. Thanks to the fact that each resource on the Web is identified by an URI it is possible to dereference them to get other related data.

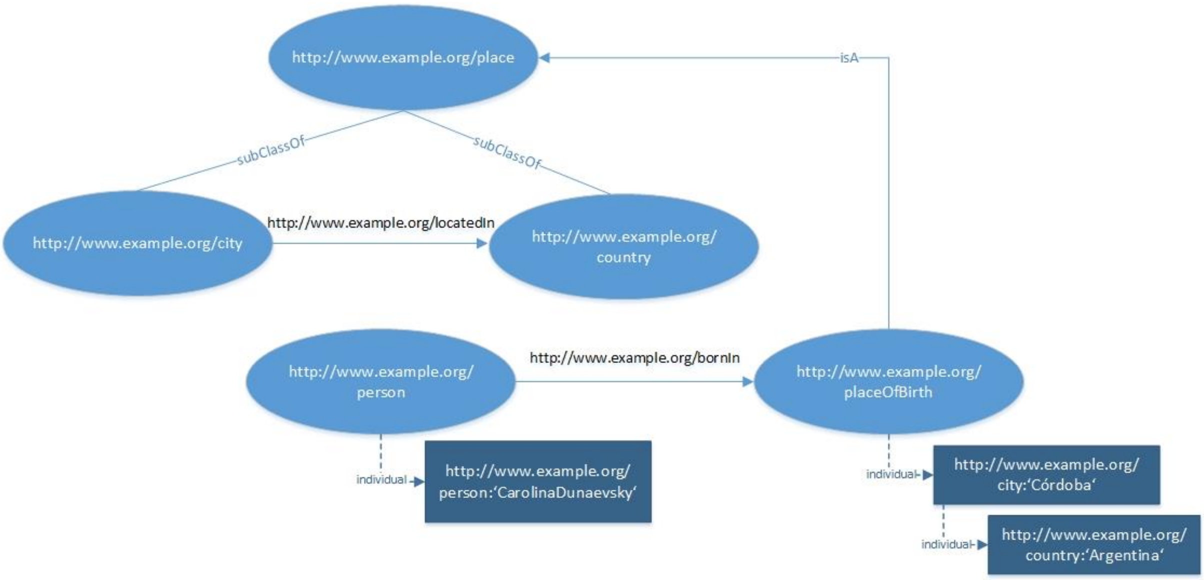


Figure 17: Many RDF-triples interconnected

Figure 17 shows that the resource labelled **placeOfBirth** is a type of **place**, which is in turn sub classified into **city** and **country**. At the same time, the resource **city** is defined as being located in the resource **country**. It is so possible to know that the individual (depicted in rectangles) **Carolina Dunaevsky** is a **person** having **Córdoba** as place of birth, which is in turn a **city** located in **Argentina**.

The purpose of an *RDF Schema* (hereinafter **RDFS**) is similar to the one of an xmlschema, i.e. to describe properties and classes. However, the distinguishing characteristic is that RDFS provides semantics for generalized-hierarchies (taxonomies) of the properties and classes it represents. Classes are used as the triple subject and object while a property constitutes its bilateral predicate. By specifying for example the *domain* and *range* of a predicate, the kinds of members of a class that subjects and objects can take can be restricted (cf. Noy and McGuinness, 2001:4).

Figure 18: RDFS domain and range

The triple in figure 18 states that the class **city**, which is denoted by the **subject** whose predicate is **rdfs:isLocatedIn** (as an instance of rdf:Property) is the domain. Reciprocally, the class **country,** which is denoted by the **object** whose predicate is also rdfs:isLocatedIn is the range. The restriction implies that a city is located in a country and does not allow for a country to be located in a city.

The *Web Ontology Language* (**OWL**) depicted on the 4th layer of the Semantic Web stack is another means to defining properties and classes but with a wider vocabulary. This language enables for example the specification of relations between classes (e.g. if they are disjoint, equivalent), of cardinality (e.g. the maximum or minimum of resources a class can admit), of equivalence (if a class has an equivalent class), of characteristics of properties (e.g. if a property can have at most one range – functional property), etc. While OWL builds on RDFS, the kind of specifications just mentioned provide the means for much fuller reasoning and inference within a knowledge base (cf. Allemang and Hendler, 2011).

The Semantic Web is further characterised by the kind of assumption that OWL can behold, namely the so-called paradigm *Open World Assumption.* This means that information on the Web remains true until the contrary is proved (the information available at one specific moment cannot be taken as all the information that might be available). It is always to be assumed that new information may anytime be discovered making previous information on the Web obsolete. In contrast, the paradigm *Closed World Assumption* (basis of the Web 2.0) holds that everything is to be assumed false if it is not directly specified in a distributed network of information (cf. Weller 2010:127).

For the Semantic Web to behave in the desired way, its underlying data has to be offered in structured collections with set of inference rules so as to manage automatic reasoning over data and this is mainly achieved thanks to the four top layers in figure 14. There is not much explanation available for these layers, thus, what follows is the definition as provided by the *World Wide Web Consortium*: 'The **Logic** layer enables the writing of rules while the **Proof** layer executes the rules and evaluates together with the **Trust** layer mechanism for applications whether to trust the given proof or not.'[24]

Reaching the end of the Web stack, **digital signatures** are defined by Berners-Lee, Hendler and Lassila (2001:3) as another vital feature for being 'encrypted blocks of data that computers and agents can use to verify that the attached information has been provided by a specific trusted source'.

3.2 Knowledge, its representation and its organisation

In the previous chapters it was demonstrated how language was conceived at various times, how terminology came to be and how the needs influenced its use and application. In this regard, the notion *concept* was introduced in chapter 2 already using the definition relevant to terminological needs, i.e. a concept being 'a unit of knowledge created by a unique combination of characteristics' (ISO 1087: 2000). While todays' definition of concept makes allusion to 'a unit of knowledge'; in 1990, a concept was defined – also by the international norms – as 'a unit of thought constituted through abstraction on the basis of properties common to a set of objects' (ISO 1087:1990). The gradual shift of the definition from *unit of thought* to *unit of knowledge* highlights the change in the terminological paradigm. It is therefore our intention in this section to address the notion *knowledge* per se and the general dimensions of the correlations between knowledge representation and knowledge organisation (systems) to later present the role of ontologies in the Semantic Web. The

[24] http://www.w3.org/2001/12/semweb-fin/w3csw (07-08-2015)

following figure offers an overview of all aspects related to knowledge that will be further described in this section.

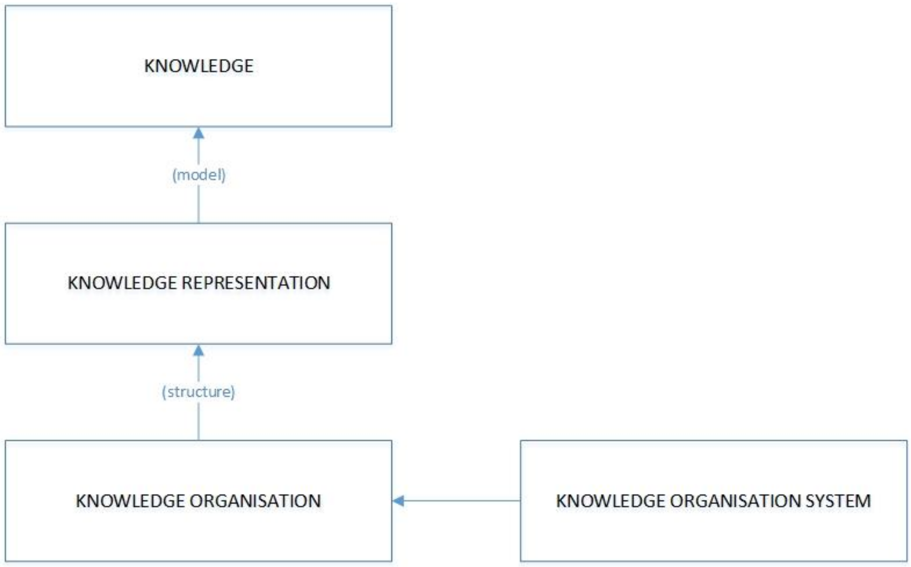


Figure 19: Knowledge and its scope

Knowledge Organisation Systems (hereinafter KOS) cover all concept systems and terminologies used for ordering and retrieving **knowledge**[25]. For this reason, under KOS it is common to find *authority files*, *glossaries*, *dictionaries* (e.g. lists of terms sometimes accompanied by their definitions), *taxonomies* (classifications), *semantic networks* and *ontologies* (which provide the relationships between terms and concepts), among others. These have naturally different prototypical data models (cf. Gail 2000).

The relation between *knowledge* and *organisation* brings us back to Aristotle (384–322 B.C.E.) who tried to understand the organisation of the natural world in a logically-organized manner. Being considered the precursor of the theory of knowledge, this philosopher and scientist believed that our appreciation of the world is influenced by our own experiences, i.e. we acquire knowledge thanks to our reasoning capability that enables us to associating new or already known objects with contexts or experience[26]. As it will be explained later, Roche (2015:138) highlights in this regard the epistemological principles on which concept theories actually rely.

It is recognized that knowledge needs some means of expression to become evident and, as it was shown, it is communicated among people through NL. However, in terms of formal

[25] 'cognizance which is based on reasoning and passes verification' (ISO 5127:2001)
[26] Cf. https://es.wikipedia.org/wiki/Epistemología (20-09-2015)

communication there is something else to it. According to its own epistemological nature, knowledge remains something rather abstract but it needs some kind of formalism for its representation if it is to be understood by both humans and machines. For that reason, there is also the distinction between formal and informal **knowledge models**. Allemang and Hendler (2011:15) state that a model 'help people understand their world by forming an abstract description that hides certain details while illuminating others'. Models meant for human communication offer a particular advantage and disadvantage over the same characteristic, which is the interpretation of information (i.e., knowledge that is communicated). As opposed to machines, people can associate different information and make conclusions, i.e. humans can and do interpret information but in a subjective way, since each person is influenced in his thinking by his own environment and background knowledge. To this, Allemand and Hendler (2011:19) further state that 'when a model relies on particulars of the context of its reader for interpretation of its meaning' it is then considered informal.

On the contrary, formal knowledge models aim at explaining facts about the world without subjectivism so as to enabling computers to answering and solving complex problems encountered in NL. For this, knowledge needs to be conceptually structured and formally expressed.

> Wissen […] ist etwas, das einem System zugeschrieben werden kann, ohne auf die konkrete Form zu verweisen, in der es realisiert ist. Zudem ist Wissen unendlich und zeigt so das Vorhandensein sowohl von Struktur- als auch von Verarbeitungsaspekten: Wissen liegt nicht nur explizit vor, sondern kann auch durch Inferenzprozesse erschlossen werden.
>
> Wissen ist ebenfalls nicht gleichzusetzen mit „Information" […]. Information ist vielmehr das Bindeglied zwischen Daten und den Strukturen, die abstraktes Wissen realisieren: Daten sind dann Information, wenn sie als Instanzen schematischer Strukturen erkannt werden. (Carstensen, in: Carstensen *et al.*, 2010:533)

Carstensen (2010) suggests that abstract knowledge (i.e., non-linguistic) can be systematized by arranging data according to a scheme. This can be immediately related to a data model as introduced in section 2.3 for TDBs. However, data stored in a DB still requires the addition of knowledge so that the values held inside can be specified and constrained. Carstensen therefore explains that knowledge is not only something that lies available in an explicit way but also something that can be implied. For revealing implied knowledge it is necessary to define the meaning of data within the framework of its interrelationships with other data. Along these lines, the concept that describes this whole is a field in Artificial Intelligence[27] (hereinafter AI) called **knowledge representation**, which is mainly concerned with formally explaining the constructs that make up knowledge, namely (cf. ibidem, p. 534):

- *Concepts* being the representatives of the *entities* in the world,
- *Instances* as individual occurrences of concepts,
- *Attributes* as the underlying properties of entities,
- *Relations* that hold between the entities in the world denoted by terms, and
- *Rules* in terms of relationships between facts.

Table 3 below summarizes the different paradigms of knowledge representation that make possible building up a structure of knowledge with the aforementioned constructs (cf. ibidem, p. 535).

	Semantic networks	**Ontologies**	
		Frames	**Description Logics**
characte-risation	object/concept	Frame	concept (class)
	node/arcs	Slot	role (property)
		Filler	
			individual
assumption	closed world assumption	open world assumption	open world assumption

Table 3: KOS with focus on terms, concepts and their relationships

Semantic network is apparently the most basic knowledge representation paradigm that emphasizes the connections between terms and concepts. These networks encode knowledge (in a graphic) by means of labelled nodes that are connected with directed arcs, which establish the relations between different objects or concepts. In a similar way, Collins and Quillian (1969) proposed a hierarchical network model of semantic memory for computers where information is stored in categories. These categories are in turn logically related to each other forming a hierarchy of broader and narrower ones. The main focus of

[27] 'branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement' (ISO/IEC 2382-1:1993, in: ISO/IEC/IEEE 24765:2010).

this theory lies in evaluating the relationship of *linguistic units* in terms of their content. Concepts or objects in the network are classified according to their content and related to each other by some common content that does not need to be repeated for each pertinent node, since a fact is stored at the highest level to which it applies allowing general facts between nodes to be reciprocally inferred. This is how these two authors explain how we humans structure and save information in our minds (*human associative memory*). Equally important to them is the time it takes a person to retrieve information from his memory (*reaction time*) according to where in the structure the data is located. By taking these reflections into consideration, Collins and Quillian attempt to solve natural language problems at computer level. (cf. Collins and Quillian, 1969:240-247).

Analogously to Chomsky's GG theory, the problem of the knowledge that can be retrieved via such a network is that the arcs do not distinguish between the different contents that a node can stand for leading then to semantic confusion. Although this solution is commonly applied in the semantic indexing of documents, it does not put formal semantics into use.

In the bigger picture of KOS, knowledge is accordingly stored in knowledge bases (hereinafter KB). Because **Semantic Frames** and **Description Logic** are both paradigms for representing knowledge in a specific area while consciously separating content from structure, they are listed in table 3 under the same category, namely **ontologies**. For the moment being it is sufficient to know that the term *ontology* originally comes from philosophy and that it found a specific application in computer science as a means to viewing, organizing and conceptualizing a domain. The details of this matter will be addressed in section 4.1.

The need of also capturing the declarative part that semantic networks were lacking was acknowledged by Brachman (1979). He came to the conclusion that a KB has to distinguish between two components, namely a *terminological box* (hereinafter T-Box) and an *assertional box* (hereinafter A-Box). While the latter contains statements that describe domain concept hierarchies, i.e. concepts, roles and their relationships with a focus on the attributes of the underlying logical system; the former represents the assertional knowledge, i.e. the belonging of individuals – defined as instance manifestations according to the statements defined in the T-Box – in relation to their concepts (cf. Baader et al., 2008:135-138). In general, a KB can be typified as being a centralized repository for information about a particular subject. As part of an AI expert system, a knowledge base is a dynamic machine-readable resource that makes sense of what is defined inside (by means of explicit formal semantics it is possible to establish the link between the knowledge constructs and their interpretation in the world). Accordingly, the T-Box purely defines *terminological knowledge* adopted in the modelled domain and the A-Box interprets if an individual proposition can be true or not (*factual knowledge*). Brachman distinguished our conception of the world independent of any natural language as ontological knowledge, and it's structuring as epistemological knowledge. Moreover, he states that better results can be obtained if the T-box and A-box are separated so that they can be manipulated either together or separately according to the particular needs. (Brachman: 1979, in: Carstensen, 2010:536).

A particularly interesting paradigm for representing the nature of things is Fillmore's cognitive theory (see section 1.2). Recalling the broad outlines of *Frames semantics*, a frame represents an entity[28] in the world as a set of attributes (slots) and associated values (fillers), whereby a frame is represented by a concept that can in turn be used to fill the aforementioned slots.

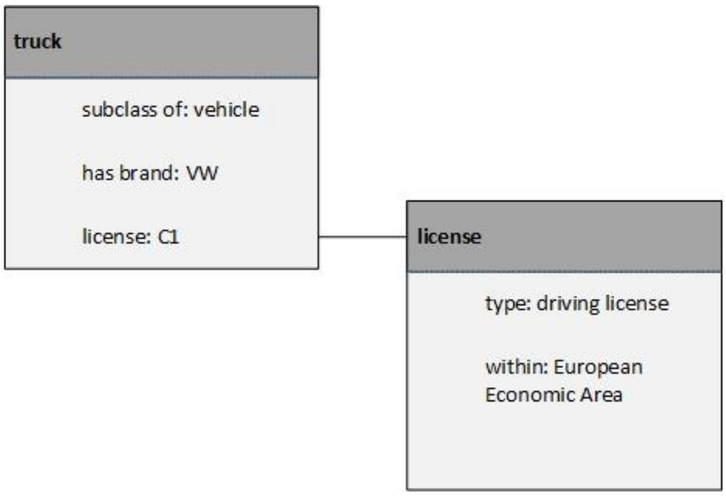


Figure 20: Example of a simple semantic frame

[28] 'conceptual semantic unit that typically functions as a participant' Note: An entity is an individual such as a person, organization, physical object, or logical entity, as well as, on occasion, a number, quantity, dimension, or a reification of an eventuality, a property, or a quality, e.g. emotion (anger, love), the value of a colour, etc. (ISO 24617-4:2014)

Minsky (1975:104) describes frames as ‘a data structure representing a stereotyped situation’, i.e. a taxonomy where slots’ default characterise the prototype of a frame as whole making the relevant knowledge about an entity directly available through its scene. The slots themselves can be framed (like **license** depicted in figure 20) and their default values – to fill unspecified aspects – are mutually inherited through abstraction. Unlike semantic nets, this paradigm does not only offer a definitional component but also a descriptive one in relation to the way frames are to be used.

Representing domain knowledge can also be done by means of **Description Logic** (hereinafter DL). In contradistinction to frame semantics – that typically uses First-Order Logic (FOL) –, DL offers wider reasoning capabilities with an emphasis on the properties of the underlying logical system. Already the fact that DL was previously called *terminological logic* brings out the importance of terminology for computational purposes and, more specifically, for the Semantic Web. A concept hierarchy in DL is built out of *atomic concepts* that are characterised just like in chemistry[29]. That is to say, a concept (in chem., *an element*) has a valence according to its combination possibilities with other concepts (in chem., *atom*, i.e. ‘the smallest component of an element having the chemical properties of the element’[30]). For instance:

land vehicle : vehicle with wheels

water vehicle : vehicle with tracks

+ some property restrictions

= amphibian vehicle

(: wheeled or tracked vehicle that can be used on land or water)

The concepts **land vehicle** and **water vehicle** are both interpreted as sets of individuals of the class **vehicle** and their properties are used as roles for binary relations. In addition, and just like in frame hierarchies, atomic concepts in DL inherit characteristics from those

[29] Cf. https://en.wikipedia.org/wiki/Description_logic (20-09-2015)

[30] http://dictionary.reference.com/browse/atom?s=t (31-08-2015)

concepts higher up in the taxonomy. In this example, the narrower concept **amphibian vehicle** is naturally subsumed by the broader concept **vehicle**.

These two last paradigms are represented by formal languages, namely RDF/RDFS and OWL that were already presented when analysing the *Semantic Web Stack* in section 3.1. For the most part, RDF/RDFS is associated with frame semantics and OWL is the representational language, which is in turn basis of the DL paradigm. In appendix B are listed the major differences these two languages offer in relation to their expressions.

Although out of the scope of this research, it should be mentioned that predetermined knowledge structures need some reasoning mechanisms to produce useful outputs. In this regard, there are different reasoning mechanisms (e.g., FaCT ++, Hermi) that check the consistency of the descriptions used in a KB and allow the inference of individuals through defined classes.

3.3 Summary

In the same way that there would be no terminology without language, language is the fabric of the Web. Consequently, this chapter focused on the semantic technologies that concentrate on language as the interface between knowledge and terminology (as the set of designations belonging to one special language).

Section 3.1 exposed that the Web 2.0 is already a primary source for storing and accessing data. For the simple reason that its origin dates back some decades, the data saved under the Web has increased to the extent that it is a neither obvious nor an easy task to manage. Most of the Web's content is ready for the humans' eye, i.e. not for machines to manipulate it meaningfully. When typifying XML, indirect reference was made to the origins of the Semantic Web due to the problems that this second decade of the Web posses with its syntax-based approach. Even though XML offers normative data structure – including metadata – to the Web (unlike HTML that only provides layout), it does not contribute on making structure understandable, i.e. it does not capture the meaning of documents and neither enables a proper sharing of knowledge across the Web. The information saved on the Web 2.0 holds on to the *Closed World Assumption* paradigm presuming that what is not known at a specific time to be true must then be false.

On the contrary, the Semantic Web has the potential to becoming a powerful instrument for offering explicit semantics to its data and the various encoding schemes in which they are represented. As exposed in section 3.2, the concept *Knowledge Representation Systems* is not really new; it just needs some adaptation to the new context the Web is evolving to. These systems have been typically centralized but the overwhelming amounts of data on the Web (which recalls Allemand and Hendler's – 2011:7 – slogan, 'Anyone can say Anything about Any topic'), makes such an approach unmaintainable. Consequently, semantic technologies promote the use of interlinked data through Unique Resource Identifiers and documents marked-up with semantic information.

Within the framework of the different knowledge paradigms it was possible to distinguish how some KOS are semantically loaded. Network-based structures represent knowledge in a dispersed manner, i.e. all relationships are based on a central concept. Furthermore, the first attempts of semantic networks (considered graph-based semantic representations – section 3.2) that date back to the 1960's may be related to Chomsky's GG. This presumable association is done due to the fact that semantic networks also formally model NL by hierarchically describing *propositions* (in terms of what can be true or not) and the relationships between the elements that make their structure.

Frame-based structures go further in schematically organizing knowledge by means of concepts that establish different relations among them according to the various contexts in which an entity can be analyzed. This theory is quite relevant for knowledge representation as it treats concepts as frames that form (domain) systems through which the meaning of a word can be understood. It is possible to argue that this descriptive framework gives account for the interrelations between language structure and conceptual structure in a similar way that terminology does. At the same time, for applying knowledge to both a particular viewpoint and many viewpoints, this second paradigm seems to be a right approach for the Semantic Web to overcome the challenge of solving problems (by processing of knowledge instead of text).

In this respect, it was brought to light that ontologies are a central pillar in the architecture of the SW since they provide the basis for automated reasoning in a domain by asserting the existence and kinds of relations hold between entities. In the context of KOS, Frame Semantics and Description Logic are both prototypical knowledge data models that view, organize and conceptualize specific domains while consciously separating content from structure. Despite the fact that these two paradigms structure knowledge with different constructs, they both rely on the extensional approach applied to ontologies. On the grounds that ontologies are being used to provide knowledge input to databases and applications – enabling search engines in the Web 3.0 to answer in a more meaningful way and consequently give rise to a vast and complex global knowledge base –, the next chapter is rendered to them and its application towards terminology.

4 Terminology and the Semantic Web

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data.

Tim Berners-Lee (2006)[31]

4.1 Ontologies for organising knowledge on the Web

The term *ontology* (from Greek '*ont'*, which means *to be*) is actually borrowed from philosophy and refers to the study of the nature of the world itself. Without going into details about the many ways there are to describe this concept, the interesting application for this thesis lies in the field of AI where the most quoted definition is the one from the computer scientist Gruber (1993:1): 'ontology is an *explicit specification* of a *conceptualization'.*

Under *conceptualization* suggests Gruber the abstraction of world knowledge in terms of entities: the *things* that exist in the world, the *relationships* they hold and the *restrictions* they pose to each other. Because of its own abstract nature, the need for a *specification* of the conceptualization seems evident. As it was pointed out in the framework of terminology, conceptualization is uttered in NL including the problems that an informal language system carries with it. However, the term *explicit* was included in Gruber's definition in combination with the term *specification* to highlight the need of representing a conceptualization in FL that is free from subjectivism. In this context, the author of this research believes that Gruber intentionally did not coin a new term, he rather availed of the epistemological root of the ancient term *ontology* to defining it in the domain of *ontology engineering*.

Moreover, the International Organization for Standardization has also given some thought to this and proposes the following definition for *ontology*:

> 'formal representation of phenomena of a universe of discourse[32] with an underlying vocabulary including definitions and axioms that make the intended meaning explicit and describe phenomena and their interrelationships' (ISO 19101-1:2014)

[31] http://www.w3.org/DesignIssues/LinkedData.html (02-10-2015)

[32] 1. 'view of the real or hypothetical world that includes everything of interest' (ISO 19101-1:2014), 2. 'the collection of concrete or abstract things that belong to an area of the real world, selected according to its interest for the system to be modelled and for its corresponding environment' (ISO 15531-42:2005)

When it comes to conceptualizing a domain, concepts and their relationships play the most important role. A *concept* represents a *set of things* within a specific domain. Identifying how a *thing* is member of a class allows subscribing a concept in two ways. The proposition from Stock and Stock (2013), which adapts the well-known semiotic triangle of Ogden and Richards to information science purposes, illustrates such a conception.

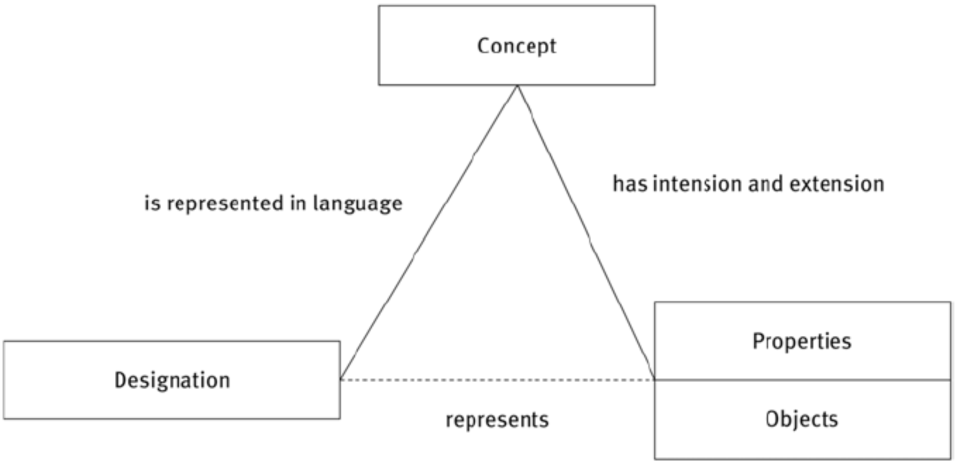


Figure 21: The semiotic triangle in Information Science (Stock and Stock, 2013:532)

On the one side, a **concept** is here understood as a *set of objects* that, after the conceptualization process, results in its **extension**. When a concept is defined by its extension, only the *essential properties* of the set of objects are observed. On the other side, the **intension** of a concept is a description of both the necessary and sufficient properties that an **object** – pertaining to the same set – needs to fulfill to be member of the set of objects. The relation between a **designation** and a concept is literally depicted in figure 21 with the indication '**is represented in language**', i.e. a concept is represented in a language system (be it NL or FL) by designations. (cf. Stock and Stock, 2013:531-533).

The development of an explicit ontology involves as a first step the identification of concept(s) as *class(es)* after having determined the extensional properties from the *thing(s)* involved in the specific domain. It is secondly needed to establish the relations that might exist between classes and those between the properties of classes. For all practical purposes, concepts are organized in a taxonomy that accounts for superordination and subordination. Among the most widely used forms we find the '*is a kind of*' and '*is part of*' relations; more specifically, relationships expressing hierarchy and meronymy respectively. At the same time, it is possible to relate classes that do not belong to a same hierarchical structure but have semantic or contextual similarities and such kind of relationships *are* called *associative*. Relations can also have *characteristics* that capture further knowledge about the relationships between concepts. Examples of these are *cardinality* (the number of values allowed in a relationship), the different degrees as to how a relationship can hold on a concept (e.g. universally, optionally, with some restrictions), etc.[33]

When the conceptualizing process is performed in such a way that it materializes the steps just presented, i.e. after having descriptively defined classes, properties and characteristics of the corresponding interactions, the result is an ontology.

It should be noted here that an ontology and a knowledge base are not the same, i.e. the former may be part or not of the latter. This clarification is not only generally necessary but also specifically applied to Table 3, where *instances* were listed under the ontology-based KOS. *Instances* are not part of an ontology; when the conceptualization of a domain is made concrete and it includes instances, then a knowledge base is created.

Ontologies solve ambiguity problems in that they take into account structure and semantics enabling common access to information, which makes forthwith different systems interoperable. In the context of the Semantic Web, ontologies care for reusability. An ontology is reusable because it is not meant to satisfy just one application (as is the case for example of a database's schema). On the contrary, an ontology can be reused in different applications as long as the purpose for which the ontology was created conforms to the ulterior objective of the application. Beside the fact that ontologies can vary in coverage and level of detail, ontologies can be of four different types according to their possible usage (i.e. how an ontology is treated) (Mizoguchi *et al.,* in: Lacasta *et al.*, 2010:3).

Typically, ontologies can define concepts in general, a domain, a task or even the concepts from a particular domain and task. Thus, it is customary to talk about *top level ontologies* when they represent general concepts that do not depend on a particular domain, i.e. they define common high level concepts like space, time, matter, object, event

[33] These are some among the many possibilities (features) available in *Protégé* 4.2.0, the ontology-editing environment chosen for the creation of the ontology that will be described in section 4.3.

or action. There are also *domain ontologies* in charge of defining those representational terms of a specific domain and, on the other hand, *task ontologies* when the representational terms refer to a particular task or activity. If it is necessary to define concepts that depend both on a particular domain and task, concepts respond to roles played by domain entities while performing certain tasks, then *application ontologies* are well suited (cf. ibidem).

In the preceding chapter some attention was rendered to two ways of capturing knowledge, i.e. formally and informally. While an ontology can be designed informally – using descriptions in natural language and/or mind map building tools, it is more profitable to encode it in a formal knowledge representation language that is machine computable. Thereby, the building of an ontology should permit for it to be communicated to humans as well as for it to be interpreted by information systems.

4.2 Ontoterminology

After having separately covered the linguistic principles that can be associated with some of the fundamentals of terminology, the basis of terminology that relies on the clarification of concepts as units that convey knowledge, the paradigms of representing structures of knowledge for computational purposes, and finally having covered how ontologies provide the best framework for web content organization, what follows is Roche's rationale called *ontoterminology*. This approach seems rather interesting, as this author tries to merge the terminological with the ontological focus while simultaneously emphasising the distinctions of the proper character of each of the components making up the semiotic triangle.

Roche brings to reflect that in today's digital era we find ourselves dealing with different automated means that in fact depend on how their *concept systems* are represented, such as in the case of translation tools, semantic and multilingual search engines and knowledge management, among others. He thus suggests that the need of 'formalizing their respective concept systems associated with a computational representation (operationalization of terminologies)' is best covered by ontologies. (cf. 2015:128-129).

As it was introduced in section 2.2, concept systems are created through conceptualization. Roche (2007) adds to this regard that conceptualization should be defined in a formal language according to epistemological principles and not linguistic ones.

> '[…] Cette conceptualisation n'est pas du ressort de la linguistique. Sa définition relève d'une démarche épistémologique dans son appréhension des objets du monde et d'une démarche logique et computationnelle dans sa formalisation et sa représentation à des fins de manipulation' (Roche, 2007:4)

This experienced author in AI explains the nature of each vertex of the semiotic triangle by attributing them a specific language description. On the one side, Roche insists on the fact that concepts exist first and then a designation is found for them. Although this is nothing new, he considers that concepts need to be expressed in a different language other than the natural one, which should in turn be capable of representing the entities of the world in a universal manner.

Thus, in view of the concepts' extra linguistic nature, concepts are – according to Roche – to be represented by means of FL ('langage de répresentation' – that aids the computational representation of abstract knowledge) and viewed as logical specifications of the *language of intellection* ('langue de l'intellection'). Because terms are expressed and determined in context, they exist in discourse and can only be defined by NL as linguistic explanations of themselves. (cf. ibidem, p. 4-7).

While concepts build a conceptual semiotic system themselves, terms are part of a linguistic semiotic system and this is what Roche's *double semiotic triangle* below seeks to disclose:

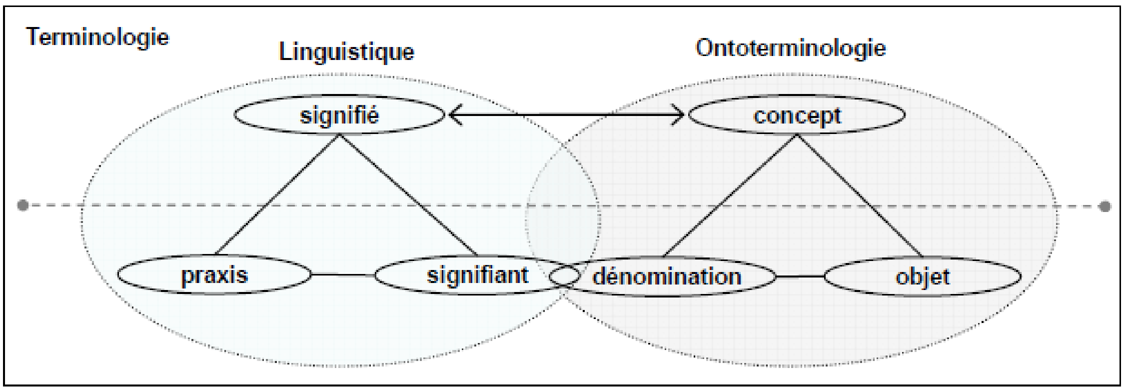


Figure 22: Double semiotic triangle from C. Roche (2007)

With this figure, Roche tries to show that concepts can be positioned within a system at an ontological level making them comprehensible independent of any NL. He further states that terms come to play only at a linguistic level, since that is the way a term can be related to the concept for which it stands (lexicalization of a concept). Precisely because a concept is abstract knowledge specified by a *definition*, drawing upon a FL becomes essential for its full representation. Therefore, the definition[34] of concepts (*logical specification*) needs to be treated from an ontological/epistemological viewpoint. (cf. Roche, 2015:131-136).

Natural Language cannot be standardized (for example, the attempts to unify the world languages into Esperanto did not result), yet a formal language offers that possibility. Moreover, given that terms are the ones to pose the problems – because of all the ambiguity that NL brings with itself – and given that science in general tries to explain the phenomena of the world based on assumptions until otherwise proven (and in the context of information science we even dispose of different FL to represent world knowledge); Roche (2015:133) argues that it would be best to concentrate on the standardization of *concept's definition* and their identifiers to express reality in a consensual way. Reality is always reality, no matter in which natural language it is manifested and no matter if it can always be put in 'words' or not. Therefore, this author is convinced that reaching a consensus on how to formally model reality will permit a clear and unequivocal communication both between humans and computers.

In the same line of thinking is Costa (2011), who also agrees that terms are elements of a lexical field and that their natural habitat is the discourse/text where they find their *referential meaning*. She acknowledges terminology as a science dealing with language and knowledge, which is expressed by concepts allowing the evaluation and validation of definitions in natural language. Similarly, Sager (1990:13) considers that, although terminology is a science in itself, the latest developments in different disciplines may contribute to a wider use of terminology thanks to its multidisciplinary character. He identifies terminological units as being linguistic entities (relation to linguistics), concept entities (relation to ontology -cognitive sciences) as well as communicative units.

[34] 'representation of a concept by a descriptive statement which serves to differentiate it from related concepts' (ISO 1087-1:2000)

4.3 An ontology of (*Onto)terminology*

This last section describes the semi-formal ontology created by the author of this research, which aims at verifying whether terminology can be divided into a conceptual and a linguistic root as suggested by Roche, and so determine if the paradigm of o*ntoterminology* can serve as a link between knowledge obtained from terminological work and ontological knowledge.

The *domain ontology* hereinafter described encompasses ontological (conceptual root of terminology) and linguistic concepts (linguistic root of terminology) along their definitions and the relationships between them. While the representation of the *conceptual root of terminology* is purely based on Roche's rationale, the representation of the *linguistic root of terminology* is inspired in some parts of the *General Ontology for Linguistic Description* (GOLD)[35], which is considered an emerging standard for descriptive terminology in linguistics.

The author of this research recognizes the developed ontology as semi-formal. For it to become formal it still needs some refinements that can be done with the advanced features offered by Description Logic, which is out of the scope of this thesis. Notwithstanding, the ontology was created by means of the open-source ontology-editor *Protégé 4.0.2,* using the W3C's OWL-formalism for its representation.
Figure 23 below offers an overview of all the classes and properties that participate in the aforementioned ontology.

[35] GOLD 2010: http://linguistics-ontology.org/gold

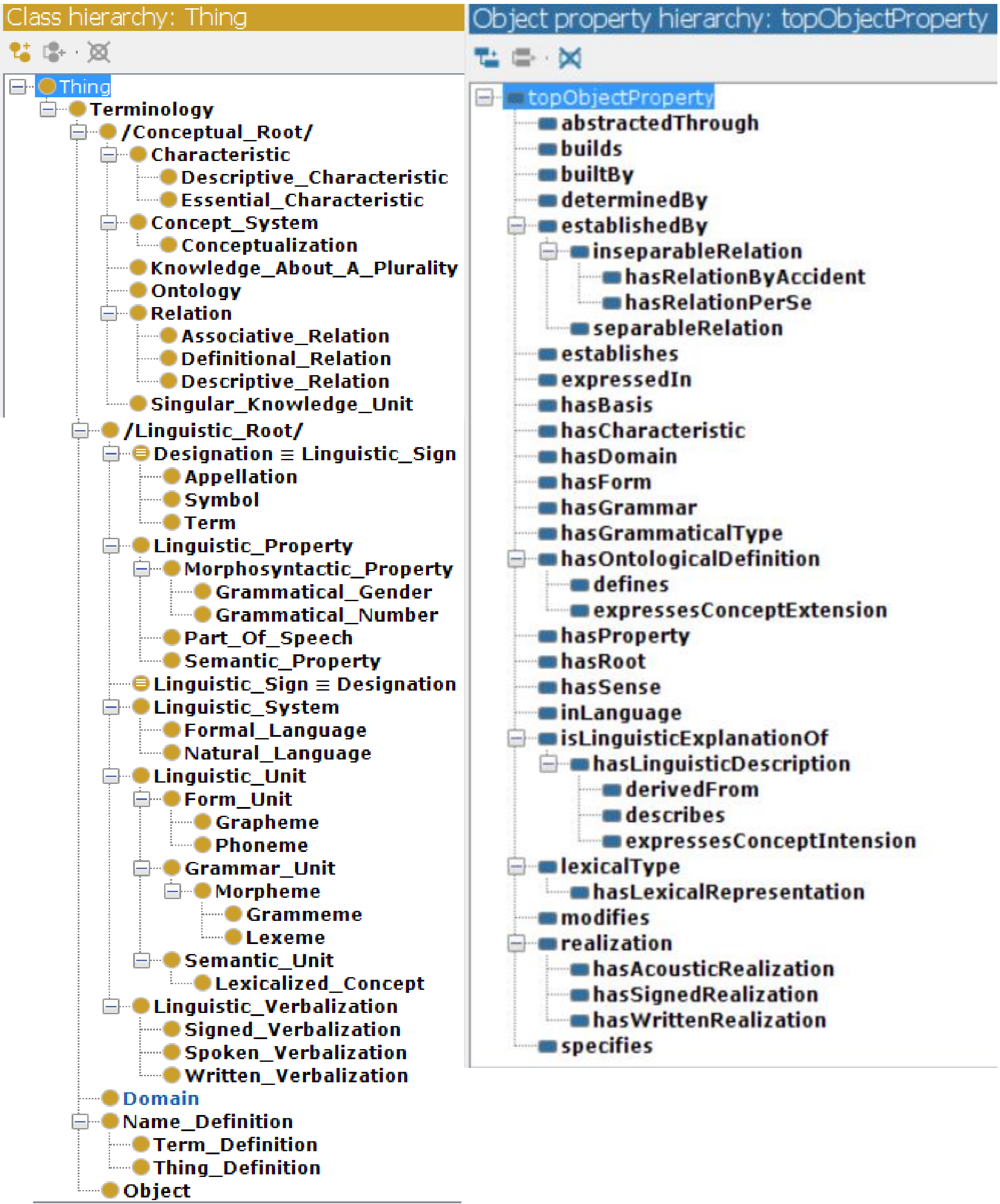


Figure 23: All classes and properties within the ontology of *ontoterminology*

Figure 24 depicts the top level concepts of both linguistic and conceptual systems: the concepts relating to the **Linguistic Root** are shown at the right-hand side of the figure and those of the **Conceptual Root** at the left-hand side.

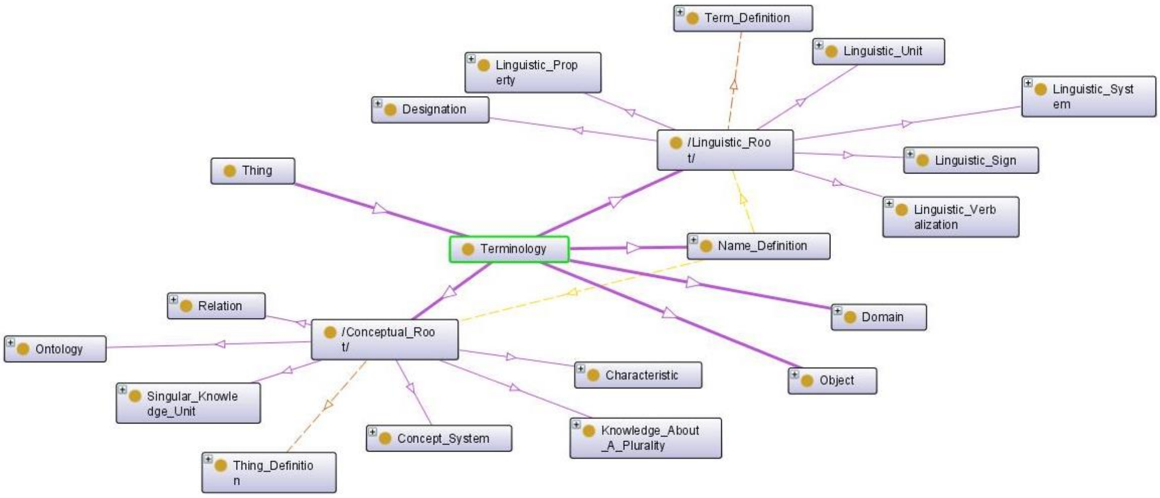


Figure 24: Ontology graph visualization with OntoGraf

Just like for any ontology, the mother root class here is **Thing**, i.e. the superclass of all classes (at the same time, *Nothing* is the subclass of all classes under *Thing*). All concepts are centrally grouped into classes where **Terminology** is the superclass that includes the subclasses (illustrated in bold Lila arrows) **Object**, **Domain**, **Name Definition**, **/Linguistic Root/**' and **/Conceptual Root/**[36] as well as their respective members (e.g. **Concept System** is member of the class **Conceptual Root**) forming a hierarchical taxonomy.

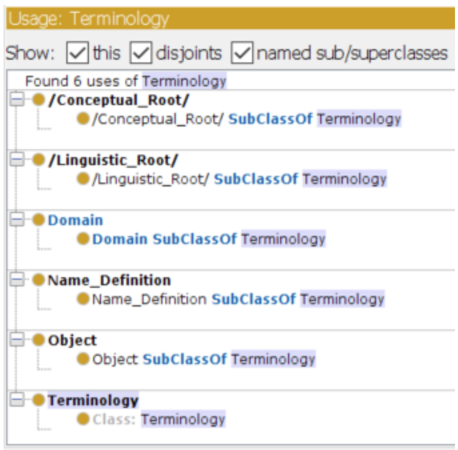


Figure 25: Usages of the class **Terminology**

Focus of this ontology is the concept **Terminology** and the list illustrated in figure 25 includes the main usages identified for it as a class. Concretely, the aim here is to present

[36] The classes '/Linguistic Root/' and '/Conceptual Root/' – between slashes – were exclusively included for the purposes of classification.

terminology as a representational vocabulary that takes into account the ontological nature of concepts as well as the linguistic origin of terms. For this, and considering the fact that terminology is a totality of concepts and terms in a specific field, the class **Domain** is correspondingly subscribed to the class **Terminology**. The class **Object** builds another mainstay for being a world referent that is first conceptually conceived, which may in turn be latter linguistically verbalized.

On the same principle operate many different kinds of relationships that are established between classes, i.e. classes are formally defined by properties. In this sense, what was presented as *subject* and *object* associated by a *predicate* in the Resource Description Framework is the basis used by the Ontology Web Language to enabling the definition of classes in the Protégé environment. It is recalled that the features of the Semantic Web build on each other; consequently, being OWL at a higher level in the hierarchy, this language naturally builds on the principles of RDF(S). For clarity purposes and, following the usual guidelines within this ontological context, all kinds of classes have been capitalized and all kinds of properties are strings whose different lexical components are delimited by the capitalization of the start of the lexical component that follows in the string.

Under the class **Name Definition** in figure 26 (which holds the class dichotomy between **Thing Definition** and **Term Definition**) it is possible to appreciate the relation between *concept* and *term* as identified by Roche (2015:134-136):

> 'A *name definition* arbitrarily links the term with the concept. […] Even though, in view of a sign's arbitrary nature, the name definition presupposes a certain independence of the concept system on the linguistic system, we must take into account the existing significations attached to terms through usage. […] A name definition has the merit of making us choose a term as a sign that grants access to knowledge of a subject field. It helps distinguish terms that designate in usage from those that denote outside discourse.' (2015:135)

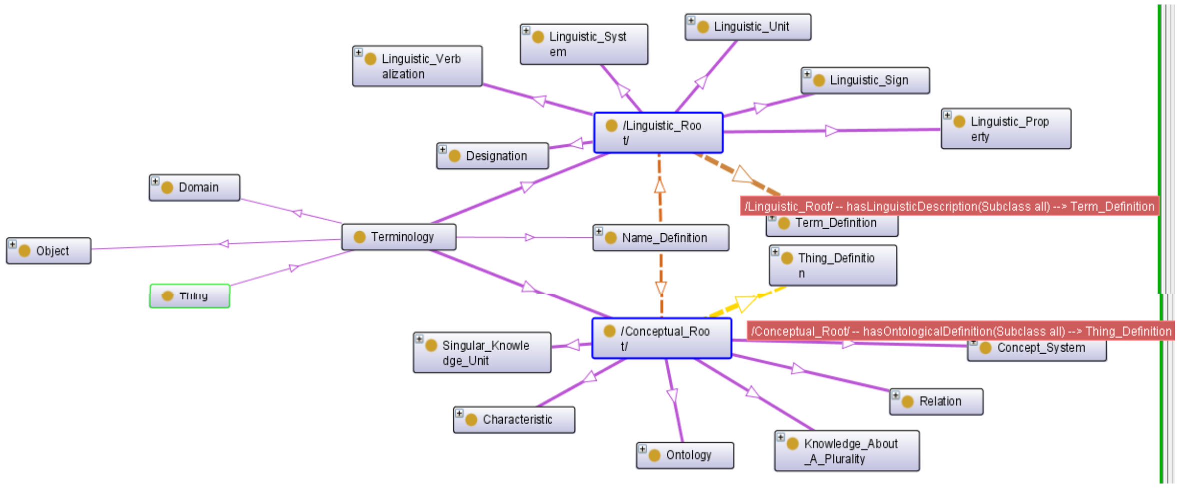


Figure 26: Definition of properties shown in the red rectangles

The yellow and orange dotted arcs in the figure above show the relationships both root classes have to the class **Name Definition** through their corresponding children, namely **Thing Definition** and **Term Definition**. At the same time, the red rectangles specify that the properties **hasOntologicalDefinition** and **hasLinguisticDescription** determine the relations hold between the classes **Conceptual Root** and **Thing Definition**, as well as between the classes **Linguistic Root** and **Term Definition** respectively.

The hand-right column in figure 27 offers not only an explicit definition of the properties but also a specification of which class makes up the origin **domain**, and the destination **range** of the relationships involved. For example, the property **isLinguisticExplanationOf** has as *domain* the class **Term Definition**, and as *range*s the class **Designation**. In other words, the concept 'Term Definition' from Roche's rationale is defined as being a 'Designation' that gives the meaning of a word in discourse.

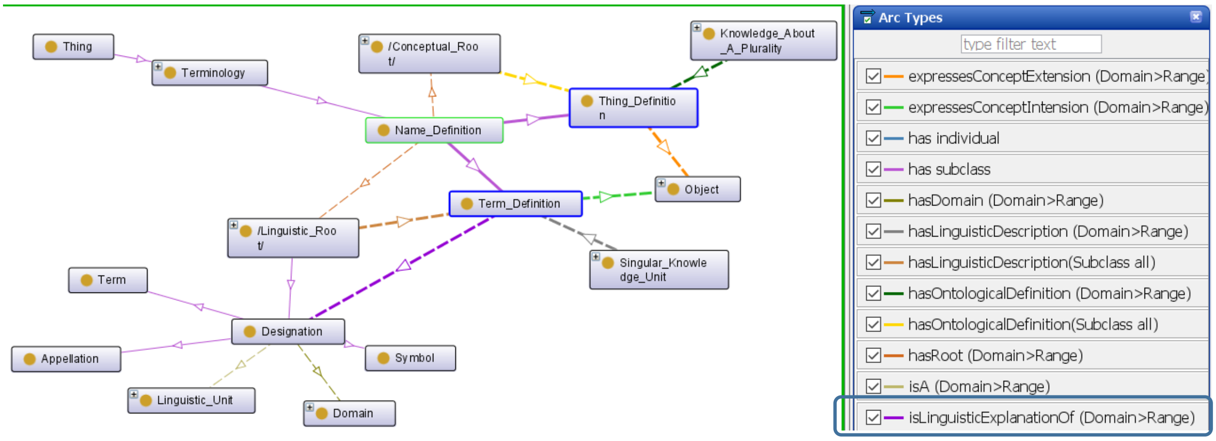


Figure 27: Specification of properties (arc types)

By the same token, the class **Thing Definition** is defined in such a way that – being subclass of **Name Definition** while having a broader relation with the class **/Conceptual Root/** – its property **hasOntologicalDefinition** (for which 'Thing Definition' is the *range* of the relation to the class 'Knowledge About a Plurality') is further divided into the subproperties **defines** (essential characteristics) and **expressesConceptExtension** to represent the notion of a concept by the characteristics of an object (see figure 28 below). As it will be seen next, characteristics play a fundamental role to establishing the relation to the conceptual or linguistic root.

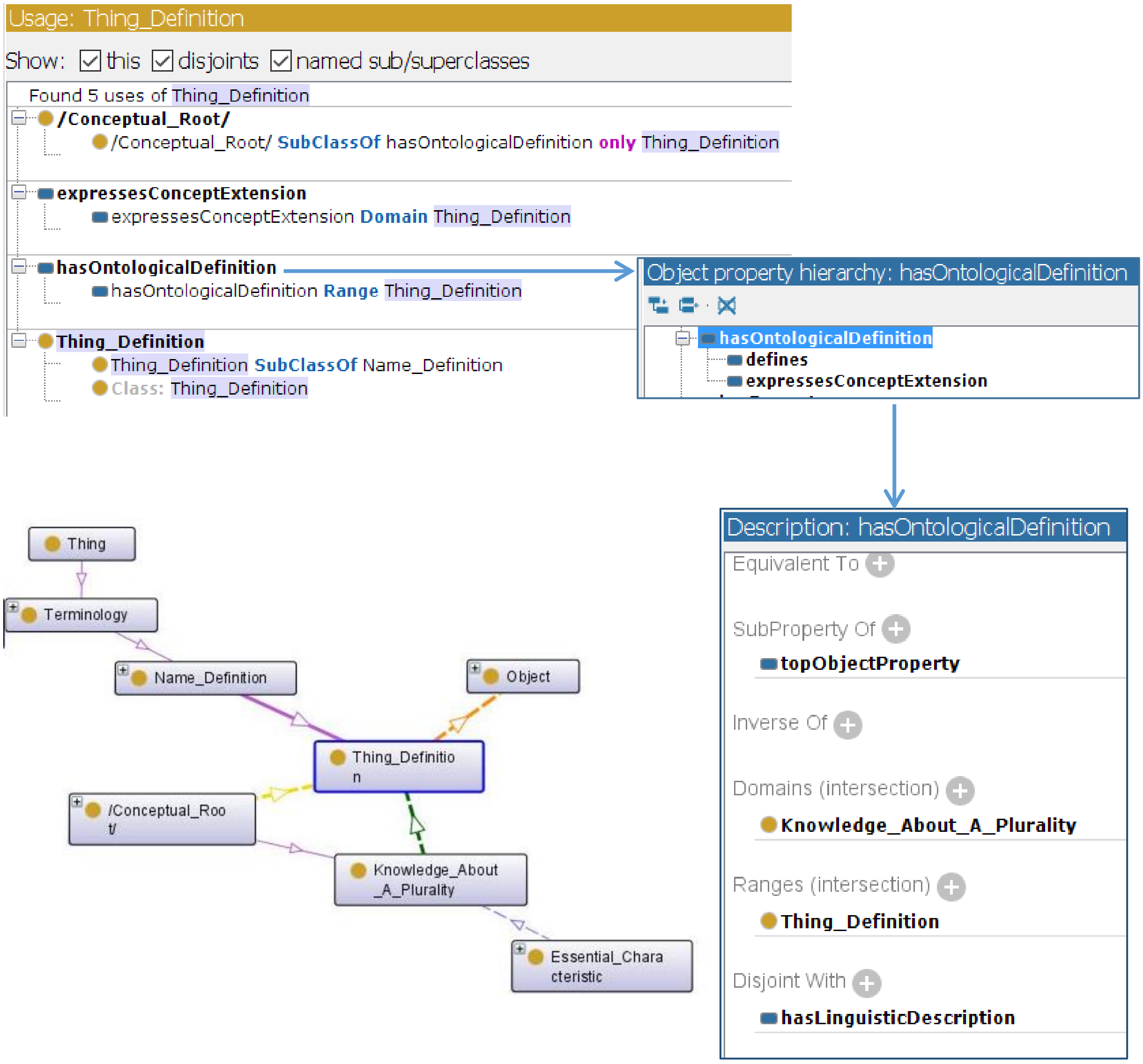


Figure 28: Definition of the class **Thing Definition**

Other than as recommended by the norms that define terminology work as concerned with the intension of concepts, *ontoterminology* is based on the *extension of concepts*. According to Roche (2013:32), object characteristics are of *essential* and *descriptive* character while establishing inseparable relations to a *unit of knowledge*. While essential characteristics build *definitional relations* by their very nature, descriptive characteristics establish *descriptive relations* because of their accidental origin. Consequently, the conceptual root of this exemplified ontology represents the class **Knowledge About a Plurality** as defined by **Essential Characteristics** obtained from the nature of the object on the one hand, and the class **Singular Knowledge** as defined by those **Descriptive Characteristics** of the object on the other hand.

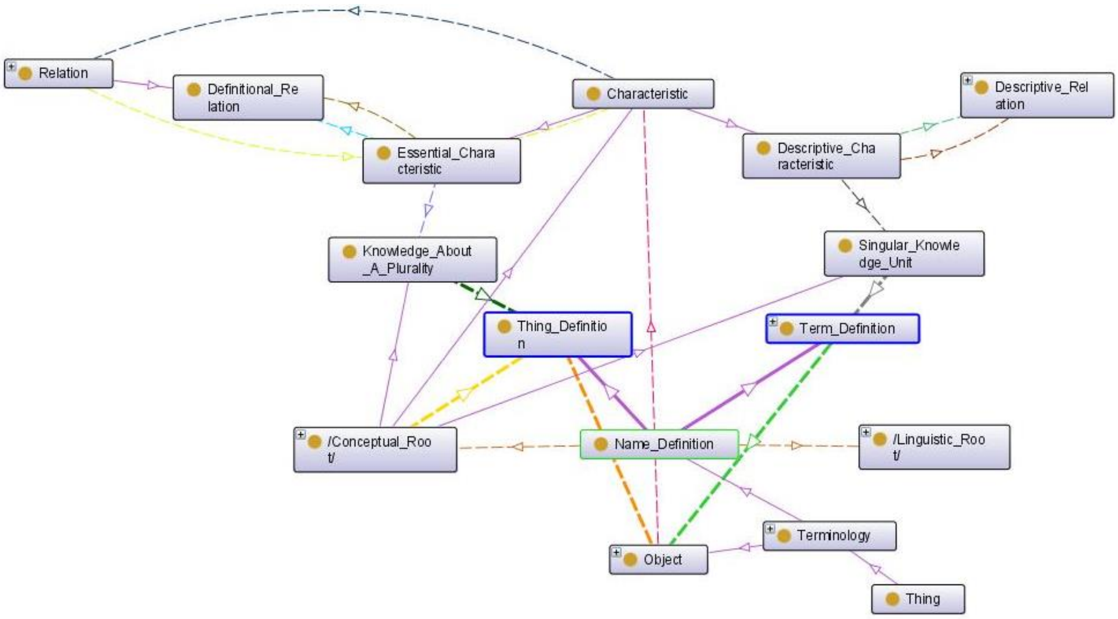


Figure 29: Definition of the extension (***Knowledge About A Plurality***) and intension (***Singular Knowledge***) of a concept

Given that the classes **Thing Definition** and **Term Definition** are siblings (i.e. they share the same parent class **Name Definition**), their corresponding properties **hasOntologicalDefinition** and **hasLinguisticDescription** are defined as disjoint with each other so that the classes in question do not subsume contradictory characteristics like it is shown in figure 30 below.

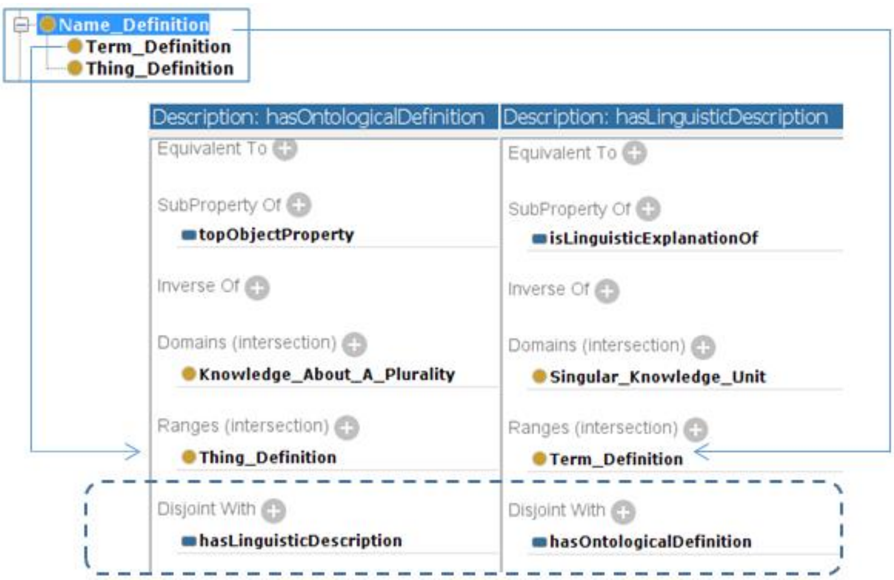


Figure 30: The classes **Thing Definition** and **Term Definition** defined as disjoint

Since ontologies help building concept systems on epistemological grounds, the whole equation just described requires the class **Concept System**, which

- is built by the relations just mentioned (Relation -- builds (Domain>Range) --> Concept System) and
- finds its basis in the class **Ontology** (Concept System -- hasBasis (Domain>Range) --> Ontology).

To observe Gruber's definition of an ontology, in figure 31 it is also possible to see the class **Conceptualization** related to the class **Ontology** (Ontology --specifies (Domain>Range)--> Conceptualization) as well as it is immediately connected to one of the subclasses of the **Linguistic System**, namely **Formal Language**. Although the uttering of a conceptualization may vary in form as linguistic unit from one language to another, concepts are herewith abstracted in a language independent way.

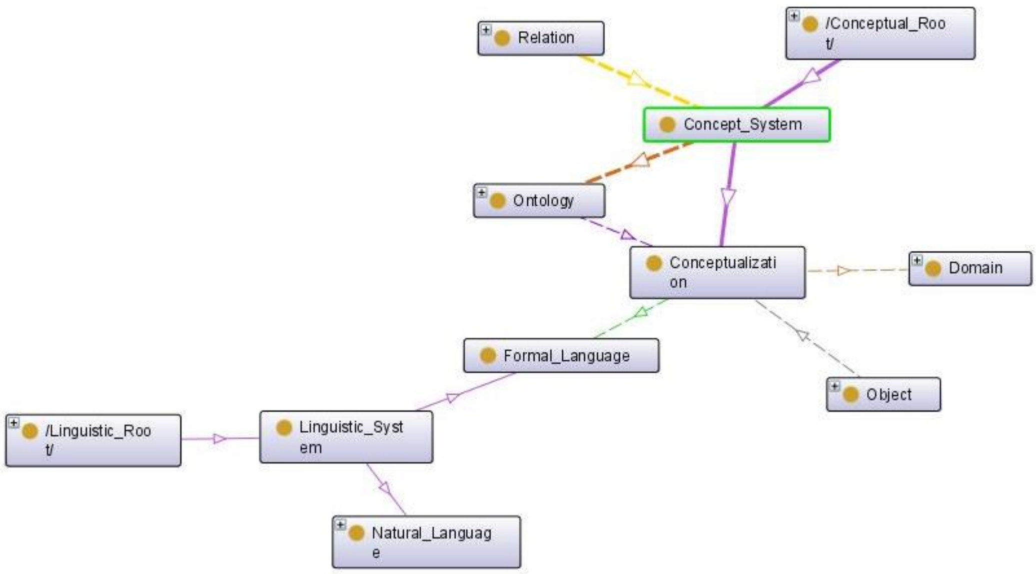


Figure 31: *Ontology* for the creation of *concept systems* on epistemological grounds

Thus, the description of the *conceptual root* of terminology is completed. To continue, the *linguistic root* will be now asserted.

The main actor within the **Linguistic Root** is the class **Object** whose definition is given as the expression of the concept intension through the class **Term Definition** (Term Definition --expressesConceptIntension (Domain>Range) --> Object). At the same time, the latter is defined as being the linguistic explanation of the class **Designation**[37] (Term Definition -- isLinguisticExplanationOf (Domain>Range) --> Designation). Be it general or specialized, a word exists as linguistic sign whose abstract structure is characterized by a form (linear characteristic of a sign for being a sequence of phonemes or graphemes), a grammar and a sense component. Therefore, the domain ontology represents the class **Designation** as being equivalent to the class **Linguistic Sign.** With such a definition, the designation inherits all characteristics of a linguist sign as presented in the figure below:

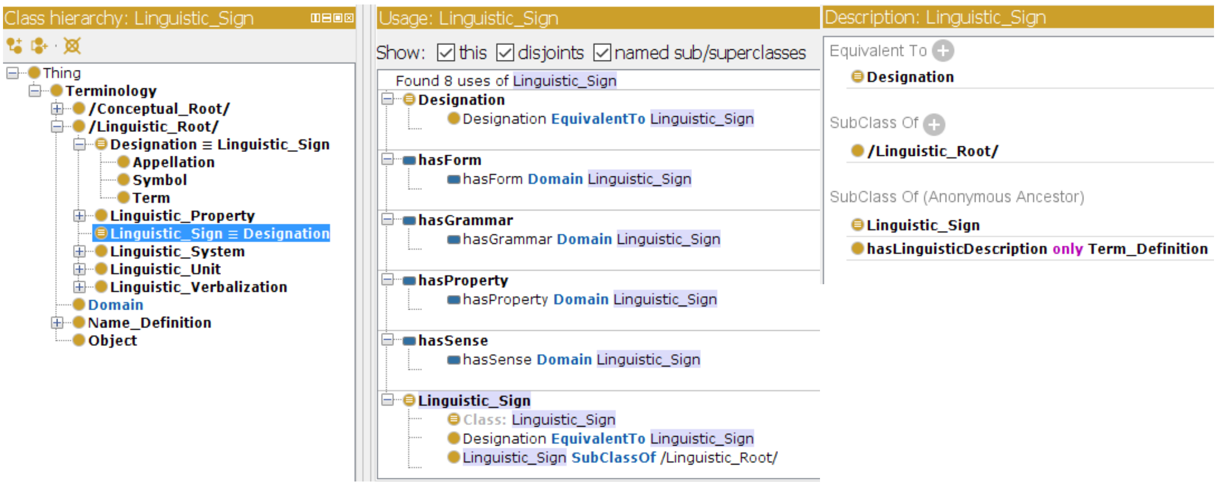


Figure 32: Hierarchy, usage and description of the class **Linguistic Sign**

Following the recommendations from GOLD 2010, the class **Linguistic Unit** was included in the ontology for referring to 'any entity which constitutes the focus of an enquiry'[38]. In line with that principle, the linguistic root of the ontology builds on the **Linguistic Unit** class that contains the subclasses **Form Unit**, **Grammar Unit** and **Semantic Unit.** These three subclasses are specified as the ranges of the corresponding relationships (**hasForm**, **hasGrammar**, **hasSense**) they establish to the class **Linguistic Sign**. Because these units can be further analysed into properties, the class **Linguistic Property** is also taken into consideration for the definition of the linguistic sign (Linguistic Sign -- **hasProperty** (Domain>Range) --> Linguistic Property).

[37] Regardless of how *designations* may be classified (i.e., *term*, *appellation* or *symbol*), they encode linguistic knowledge.
[38] Refer to http://linguistics-ontology.org/gold/2010/LinguisticUnit

Figure 33 shows that the class **Linguistic Unit** participates in a linguistic system, namely in that of **Natural Language** (Linguistic Unit -- inLanguage (Domain>Range) --> Natural Language).

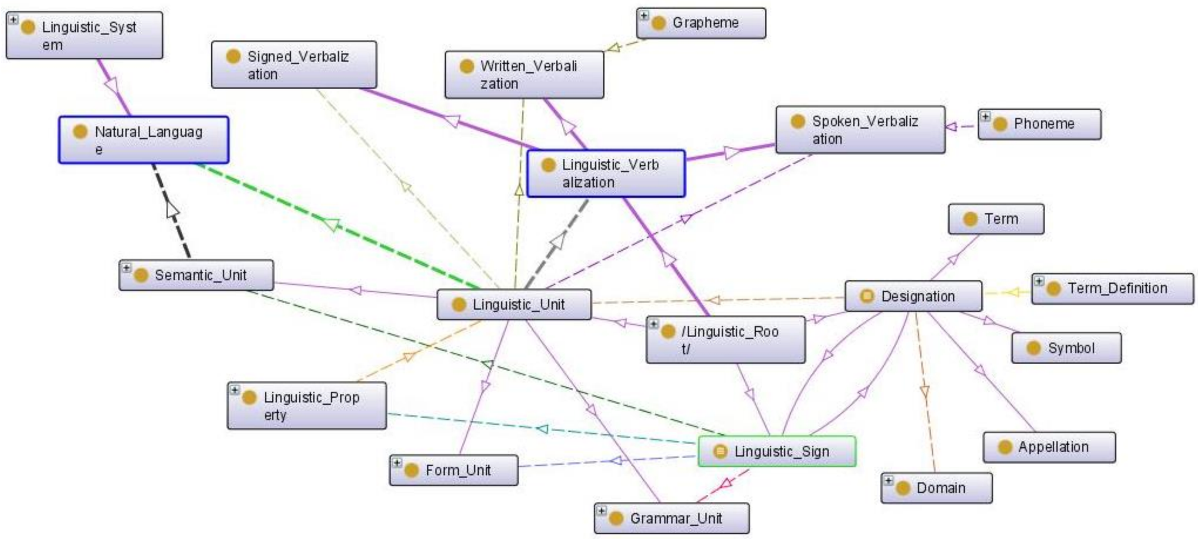


Figure 33: A **Linguistic Unit** is uttered in **Natural Language**.

It is thereupon possible to find the relation between the class **Linguistic Unit** and its verbalization through the property **realization** that contains some further subproperties for the corresponding linking of the classes.

The table below summarizes that the relation holding between a linguistic unit and its 'physical form' may be either a spoken, written or signed verbalization. When a linguistic unit is acoustically realized the form is a **phoneme**. Analogously, when a linguistic unit has written realization its form is a **grapheme**.

Linguistic Unit – Domain	**Realization – Property**	**Verbalization Unit – Range**	**Form Unit – Domain**
	Realization – Property	**Verbalization Unit – Range**	
	hasWrittenRealization	Written Verbalization	grapheme
	hasAcousticRealization	Spoken Verbalization	phoneme
	hasSignedRealization	Signed Verbalization	-

Table 4: Verbalization and realization of a *linguistic unit*

There is in fact more to say about the linguistic root, however, to conclude describing the fundamental aspects that shall verify if ontoterminology can serve as a link between terminology science/work and ontological knowledge, it is necessary to retake the class **Linguistic Sign**.

A linguistic sign can naturally only be part of the linguistic root within this ontology and its class was defined as having an information value, whereby it establishes a relationship to the class **Semantic Unit** (Linguistic Sign -- **hasSense** (Domain>Range) --> Semantic Unit) in light of the following considerations:

- Because the class **Semantic Unit** is the 'sense' component of linguistic signs – in terms of the author of this research according to Table 1, which should in turn reflect Frege's conception of '*Sinn*' –, it expresses a **Lexicalized Concept**, which, according to GOLD 2010, is a 'grammatical or semantic category expressed by a *lexical item*'[39]. Consequently, the class **Lexicalized Concept** establishes a relation to the class **Lexeme** (Lexicalized Concept -- derivedFrom (Domain>Range) --> Lexeme).
- Because the sense of a linguistic sign has been conventionally agreed by the users of each specific language community, it necessarily relates the intension of a lexicalized concept to Natural Language (Semantic Unit -- determinedBy (Domain>Range) --> Natural Language).

Furthermore, there is the fact that concepts also have an information value for being a unit of understanding. The conceptual root within this ontology reflects the extensional meaning – in terms of Frege's conception of '*Bedeutung*' as shown in Table 1 – as part of formal semantics deriving from epistemological structures that behold the truth value in this world. Hence, the truth-value of a concept is given by its ontological definition (Knowledge About a Plurality -- hasOntologicalDefinition (Domain>Range) --> Thing Definition), which is in turn a formal definition (Conceptualization -- expressedIn (Domain>Range) --> Formal Language) obtained from the corresponding abstraction of object's properties (Ontology -- specifies (Domain>Range) --> Conceptualization).

[39] Refer to http://linguistics-ontology.org/gold/2010/LexicalizedConcept

The here described ontology is to be found on the attached CD. The read-file .owl can be opened with a text editor such as Notepad or directly in Protégé 4.2 (InstallAnywhere Web Installer - http://protege.stanford.edu/download/protege/4.2/installanywhere/Web_Installers/).

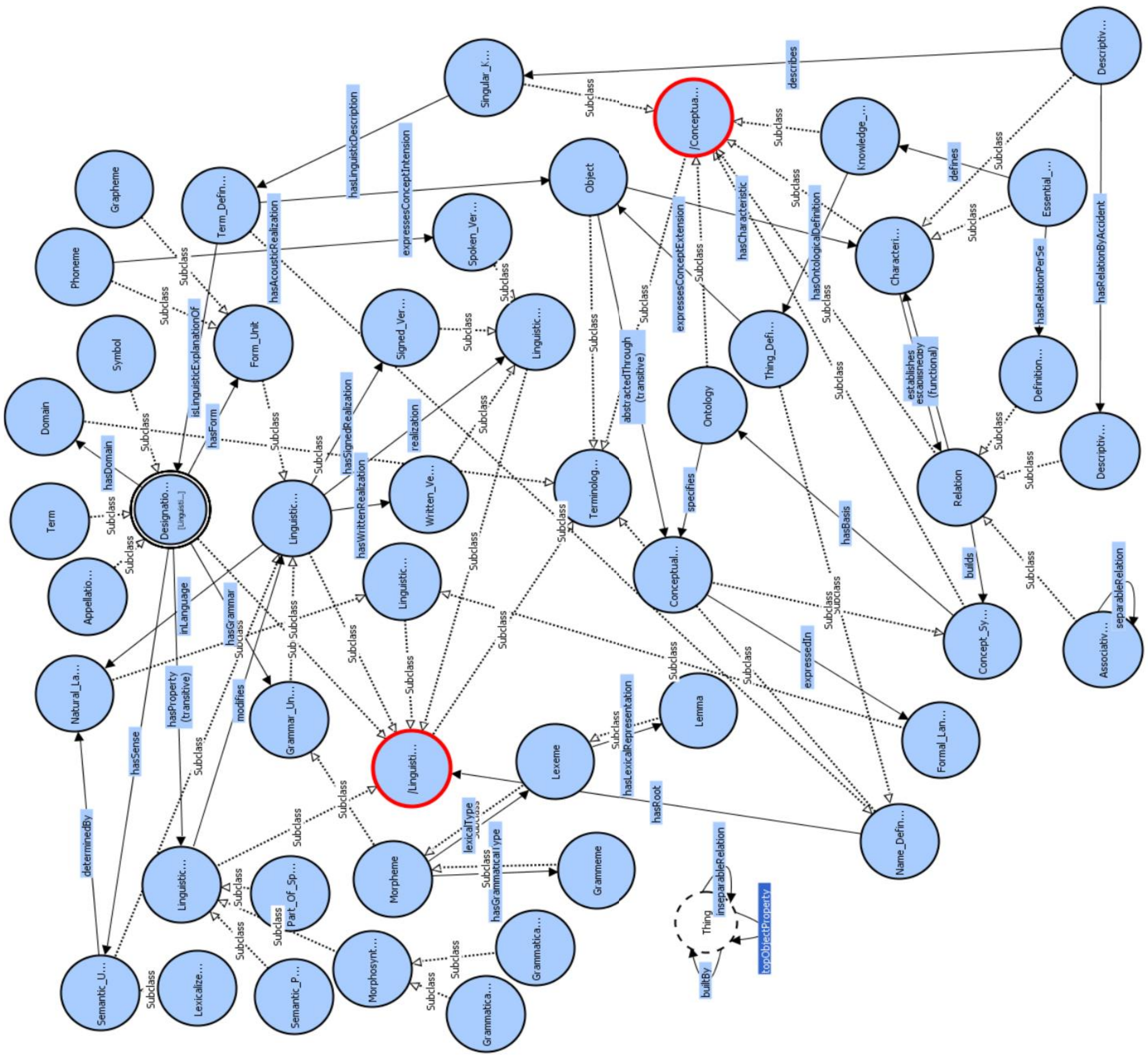


Figure 1: OWL visual representation of the whole ontology showing all classes their respective relationships

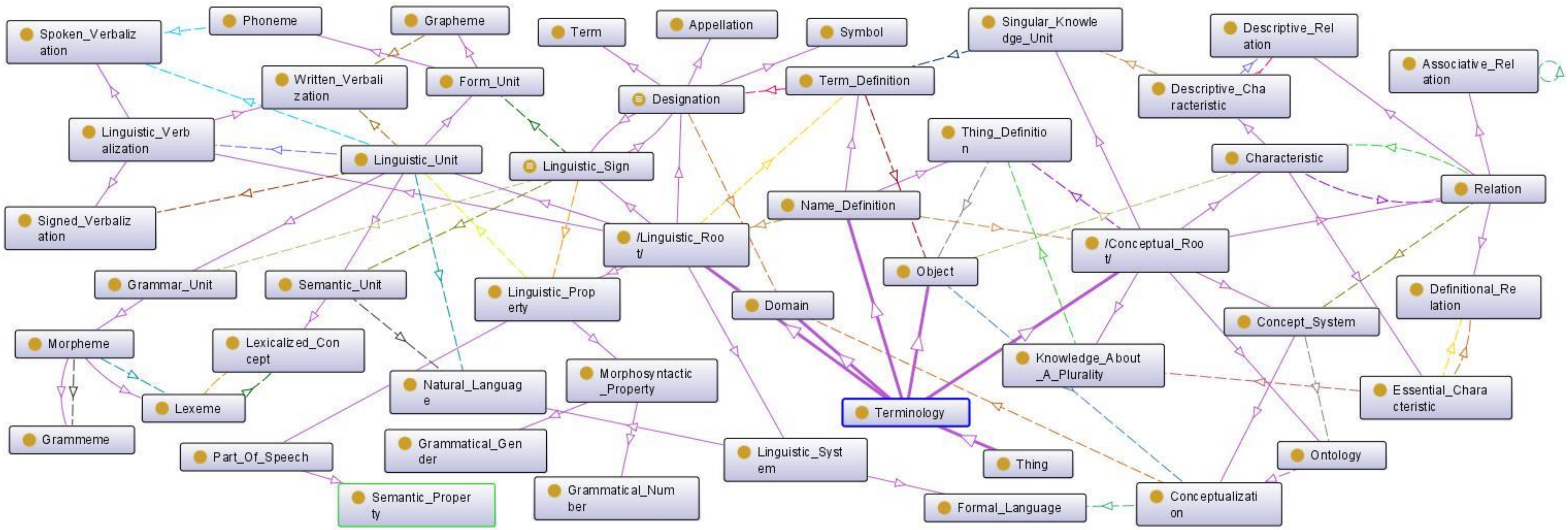


Figure 35: OntoGraf visual representation of the whole ontology showing all classes

4.4 Summary

As it was previously exposed in chapter 3, despite the fact that Frame Semantics and Description Logic structure knowledge with different constructs, they both rely on the extensional approach applied to ontologies. It could thus be speculated that the extensional approach used for developing ontologies is related to Frege's theory of meaning (section 1.4), which is in turn correlated to formal semantics. In this respect, Stock and Stock (section 4.1) explain Ogden and Richards' semiotic triangle applied to information science while taking such epistemological principles of concepts into consideration. The properties of a set of objects are abstracted in such a way that the extension of a concept is determined by essential characteristics. On the other hand, those characteristics, which account for the necessary and sufficient properties so that an individual object can belong to the same set, are part of the intension of a concept. Beyond the terminological distinction between meaning ('Bedeutung') and sense ('Sinn') made by Frege, the most important aspect to note is that the former, i.e. the extension of a concept, results from the own nature of things; while the latter, namely the intension of a concept, concerns the way a referenced thing might be expressed as.

In this way, it is possible to elucidate Gruber's definition of an ontology (section 4.1) as a specification of an abstract model of knowledge (the domain conceptualization) that is independent of any particular context. At the same time, this can be supported by the descriptions offered in section 3.2 relating to knowledge being something abstract that, in the context of AI, involves representational structures for information systems to formally capture, read and interpret data and information. It is a fact that ontologies are different in scope according to their intended meaning and are as well classified as top-level ontologies, domain ontologies, task ontologies or application ontologies.

Just as in section 2.2 was highlighted the fact that in terminology work it is extremely important to first determine the domain for establishing the relations in a concept system since the objects involved may be differently conceptualized according to the specific area they are circumscribed in, this is also important when creating ontologies. Despite the differences on the intended use of an ontology, ontologies are all equal in their ability to host knowledge.

A further relevant aspect covered is that an ontology is not equal to a knowledge base. The former is a set of concepts within a domain and the relationships between these concepts, which can be contained in the latter. A knowledge base typically includes the ontology proper as well as instances and a set of rules – proper to any KOS – that constrains the values of classes and those of instances. Moreover, the explicit specification of an ontological abstraction can be done by different formalisms like the Web Ontology Language that structures knowledge on the fundamentals of Description Logic. In addition, this section covered the fact that knowledge representation languages vary in terms of expressiveness (accuracy for the description of the components of an ontology), ease of use and computational complexity.

For the very reason that ontologies offer new ways to creating concept systems, Roche (section 4.2) introduces the so-called o*ntoterminology* paradigm. Roche argues that *terms* are part of the actual spoken/written language ('langue d'usage'), which can be associated to the conception of a *linguistic sign* from the point of view of de Saussure. In order to verify whether terminology can be divided into a conceptual and a linguistic root as suggested by this logician, the author of this research proposed in section 4.3 an ontology of ontoterminology. The nature of both conceptual and linguistic roots of terminology were schematically organized offering formalized definitions of the concepts involved while making explicit their attributes as well as the various relations that exist among them for linking them into an ontology-based conceptual system (section 4.3).

5 Conclusions

The efforts of including semantics in different language theories have shown the importance of explaining the relations between words (be they specialized or general) and the reality they account for. As could be observed, different specialists deal with concepts as *units of knowledge / units of thought* that require a means of expression for communication.

Taking into consideration that terminology is a many-faceted subject; the focus of this research was mainly terminology as a resource and terminology as a set of methodologies and procedures to be used in creating this resource. However and naturally, this immediately involved the aspects of terminology as an academic discipline and undoubtedly as a factor in communication. In this regard, the pragmatic aspect of communication was not taken into analysis; instead, the focus was set on language as the most valuable medium for communicating our understanding of the world.

5.1 Paradigms

Whilst terminology cares for precise and unambiguous communication, the vision of the Semantic Web is that of endowing information on the Web with metadata, which shall convey its meaning (by further describing or annotating it) and thus, enable information retrieval, reuse and integration.

(SO-4) Traditional concept-oriented termbases vs. ontological knowledge bases

As it was demonstrated, terminological databases and knowledge bases are both concept-oriented and deal with specialized domains. Thanks to the experience gained while creating the terminological database and the ontology for the domain of ontoterminology it was possible to establish the following differences between these two kinds of repositories:

It is true that a TDB is a repository for storing concepts as units of knowledge as well as it is true that these concepts (saved under terminological entries) are just data to 'the eyes of computers' – as implied in the very designation *terminological database*. Strictly speaking, no knowledge can be automatically retrieved from these concepts since terminological databases do not support formal representation of knowledge. The data model of a TDB is always defined for a particular application; therefore it can be neither globally shared nor reused. At the same time, traditional terminological data models are built by autonomous term entries that are independent from the content in other entries. While terminology is a means in itself, e.g. for translation, terms are typically found in discourse within a TDB. Finally, a TDB can be characterised as a data repository that presents concepts, terms and their multilingual equivalents in isolation with no semantic interconnection.

On the contrary, ontologies provide a formal and machine manipulated vocabulary that supplies a specification of the characteristics discovered during the abstraction process to which things in the world are subjected. Such formally defined vocabulary that underlies a knowledge base cares for knowledge inference as well as it constraints capturing background knowledge about the domain. Because concepts as units of knowledge are defined by their own epistemological nature – that does not vary, ontologies have no need of contextual information. Ontologies can also be reused in different applications as long as the purpose for which the ontology was created conforms to the ulterior objective of the application. Consequently, ontologies can be characterized as an artefact that represents concepts as 3D networks with several different aspects from which one can request and obtain linguistic and cognitive information in a well-structured, logical manner.

5.2 Challenges and problems

(SO-1) The Semantic Web and terminology science/work have similar but not identical frameworks

It could be argued that terminology and the Semantic Web share a similar paradigm in that they both focus on units of knowledge whose analysis is fundamental to facilitate concept clarification. They concentrate on terminological knowledge (be it monolingual or multilingual), which permits the description of the objects that exist in the world for which an explicit and common conceptualization is needed. The distinctive aspect between these 2 worlds is the way knowledge is conceptualized.

Both ontology and prescriptive terminology aim at standardization. However, while terminologists attempt to regulate language (in this case, the vocabularies of special languages), ontologists try to normalize our perception of the world.

From the observations performed when applying the created concept system (section 2.2) to the different frameworks of this research the following conclusions are made:

Creating a concept system in the framework of terminology implies concentrating on the system of concepts itself and its relationship to other concepts (concept intension). Despite the fact that the realm of ontology construction is as well the system of concepts itself, it focuses on the relationship to objects (concept extension: concept definition by specific differentiation). Within the context of information technology, a formal definition of TDBs is needed, whereby concept data modelling has to do with the structuring of the data itself (specification of data fields) and not any concept.

As previously stated in the paradigms, terminology and the Semantic Web have different origins, i.e. they came to be out of different needs. In this sense, it is worth mentioning that traditional terminology only serves to solve communication problems amongst humans and ontologies are meant to facilitate the interchange of knowledge between agents as well as their reuse in different systems.

Another intrinsic challenge has to do with the different expert profiles these two ‘disciplines’ count with. It cannot be expected to obtain a common result from profiles as varied as those of linguists, terminologists, ontology engineers, logicians, specialists in information science and/or AI. While to the eyes of terminologists the intension of a concept is fundamental, the extension of concepts is what the others specialists take as premise for analysing knowledge.

(SO-3) Requirements of terminologists and those of norms on terminology are not met

Although terminology in the framework of the Semantic Web is used as a resource, the methodologies and procedures used for creating an ontology are not exactly the same as the ones implied in terminology work.

By concentrating on the epistemological nature of concepts, terminology work would be independent of linguistic aspects (no context would be necessary to explain term related phenomena). Naturally, and as suggested by Roche, this would imply the revision of the norms on terminology so that instead of an intentional approach towards concepts an extensional one is taken into consideration to comply with the epistemological principles upon which ontologies rely. It is noteworthy in this context that any KOS can be ontologized for being structured terminology systems.

Notwithstanding, it is plausible to conclude that the available models today do not meet the requirements of terminologists and that it would be no easy task to change the paradigms on which terminologists and ontology engineers respectively base their approach to knowledge, namely because they start from different premises about their own experiences on knowledge.

(SO-2) Connection between the terminology world and the Semantic Web

The connection between the terminology world and the Semantic Web might be ensured by nothing but ontological vocabularies published in any of the semantic-rich representational languages available.

General or specialized words respond purely to the *linguistic meaning* (refer to Table 1) of concepts. That is to say, they are an expression of the human thought and feeling; thereby justifying the ambiguities in vocabulary and grammatical structures (structural and lexical vagueness) as well as the different linguistics variants there might be. On the contrary, a formal *conceptual meaning* (refer to Table 1) is based on the pure epistemological aspect of concepts which respond to a logical system of signs caring for unambiguousness, explicitness and ease of verifiability for the purpose of international understanding. It is so possible to determine the link between knowledge obtained from terminological work (linguistic representation) and ontological knowledge (conceptual representation).

By explicitly stating and agreeing upon the underlying vocabulary (concepts outside discourse) of knowledge-based systems it could be guaranteed that the meaning of data is equally defined and understood in the same way.

Between the SW and the terminology world it could be argued that the ontoterminology paradigm plays a significant role, enabling each field to profit from the other. However, although ontoterminology might be an appealing paradigm for combining terminology and ontology along with the aforementioned benefits, its practical aspect remains questionable. Defining concepts from an ontological viewpoint raises the question of how it is possible to agreeing what is what by nature. Additionally, when looking at terminology as a means in itself, the linguistic context of concepts is something rather necessary. For example, when applying terminology to translation work, any document that needs to be translated is a written expression of its author. Consequently, only through the linguistic meaning of concepts (terms in discourse) it is possible to appreciate what the author of a text might have had in mind when writing.

6 Limitations and further research prospects

Having the author of this research a background in translation and terminology but no previous background in ontology or logic, this thesis does not elaborate in depth or length the aspects of the latter. On these grounds, it was only possible to develop the ontology just to serve the thesis objective of highlighting the limitations and possibilities that could be expected from combining the disciplines of terminology and ontology in the framework of the SW. Hence, future research would include:

- Learn more about knowledge based on logic to develop a rich ontology (with more assertions and complex restrictions on defined classes and properties) as the underlying model of a knowledge base.
- Analyse the possible inferences done from asserted instances when applying corresponding query languages, such as DL and SPARQL.
- Investigate how things defined by their own nature (i.e. concepts linguistically independent defined) could be profitable for applied terminology (from a practical point of view).

Bibliography

Allemang, Dean; Hendler, Jim (2011): Semantic Web for the working Ontologist. Effective Modeling in RDFS and OWL. Oxford: Elsevier Inc.

Arntz, Reiner; Picht, Heribert (1991): Studien zu Sprache und Technik – Einführung in die Terminologiearbeit (2nd Edition). Hildesheim: Georg Olms AG.

Baader, Franz; Horrocks, Ian; Sattler, Ulrike (2008): Description Logics. In: van Harmelen, Frank; Lifschitz, Valdimir; Porter, Bruce (eds.): Foundations of Artificial Intelligence – Handbook of Knowledge Representation. Amsterdam: Elsevier.

Bergman, Mike (2007): An Intrepid Guide to Ontologies. Retrieved from: http://www.mkbergman.com/374/an-intrepid-guide-to-ontologies/ (06-05-2015).

Bergman, Mike (2010): A New Methodology for Building Lightweight, Domain Ontologies. Retrieved from: http://www.mkbergman.com/908/a-new-methodology-for-building-lightweight-domain-ontologies/ (06-05-2015).

Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001): The Semantic Web – A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Article in the Scientific American. Retrieved from: http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html (06-07-2015).

Carstensen, K.-U.; Ebert, C.; Ebert, C.; Jekat, S.; Langer, H.; Klabunde, R. (eds.) (2010): Computerlinguistik und Sprachtechnologie –Eine Einführung. Heidelberg: Spektrum Akademischer Verlag.

Casellas, Nuria (2011): Legal Ontology Engineering –Methodologies, Modelling Trends and the Ontology of Professional Judicial Knowledge. Heidelberg: Springer.

Collins, Allan M.; Quillian, M. Ross (1969): Retrieval Time from Semantic Memory. Journal of verbal learning and verbal behavior. Retrieved from: http://gureckislab.org/courses/fall15/lhc/readings/cq69.pdf (20-05-2015).

Cook, Walter Antony (1992): Case Grammar Theory. USA: Georgetown University Press.

Costa, Rute (2011): Collaboration between terminologists and domain experts: the case of the Portuguese parliament. Retrieved from http://termcoord.eu/wp-content/uploads/2013/10/Rute-Costa_07_10_2013_proofread.pdf (03-03-2015).

Edo Marzá, Nuria (2009): The Specialised Lexicographical Approach: A Step Further in Dictionary-making. Bern: Peter Lang International Academic Publishers.

Faber, Pamela (ed.) (2012): A cognitive Linguistic View of Terminology and Specialized Lang. Berlin, Boston: Mouton De Gruyer.

Felber, Helmut ; Budin, Gerhard (1989) : Terminologie in Theorie und Praxis. Tübingen : Gunter Narr Verlag.

Fikes, Richard; Kehler, Tom (1985): The role of frame-based representation in reasoning. In: Communications of the ACM, Vol. 28, N° 9. Retrieved from http://courseweb.stthomas.edu/djpanton/FikesRoleFrameRep.pdf (20-08-2015).

Gruber, Thomas (1993): A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition, 5(2). Retrieved from: http://www.dbis.informatik.hu-berlin.de/dbisold/lehre/WS0203/SemWeb/lit/KSL-92-17.pdf (16-04-2015).

Guarino, Nicola (2006): Ontology and Terminology. Retrieved from http://eaft-aet.net/fileadmin/files/VAKKI/nicola_guarino.pdf (28-09-2014).

Hanke, Johann-Christian (2006): XML leicht und verständlich. Osnabrück: KnowWare.

Hodge, Gail (2000): Systems of Knowledge Organization for Digital Libraries. In: Council on Library and Information Resources (2014). Retrieved from: http://www.clir.org/pubs/reports/pub91/1knowledge.html (20-08-2015).

Humboldt, Wilhelm von: Werke in fünf Bänden, Flitner, A.; Giel, K. (eds.) (1963): Schriften zur Sprachphilosophie, Vol. III, p. 368-756. Darmstadt: Wissenschaftliche Buchgesellschaft.

ISO norms:

- ISO 1087-1 (2000): Terminology work – Vocabulary.
- ISO 12620 (2009): Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources.
- ISO 26162 (2012): Systems to manage terminology knowledge and content – Design, implementation and maintenance of terminology management systems.
- ISO 704 (2009): Terminology work – Principles and methods.

ISO Online Browsing Platform (OBP), https://www.iso.org/obp/ui/ :

- ISO 19101-1 (2014): Geographic information – Reference model – Part 1: Fundamentals.
- ISO 5127 (2001): Information and documentation – Vocabulary.
- ISO/IEC/IEEE 24765 (2010): Systems and software engineering – Vocabulary.
- ISO 16642 (2003): Computer applications in terminology – Terminological markup framework.
- ISO 15531-42 (2005): Industrial automation systems and integration – Industrial manufacturing management data – Part 42: Time Model.
- ISO 12620 (2009): Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources.

Jurafsky, Dan (2014): Obituary Charles J. Fillmore. In: Computer Linguistics, Vol. 40, N° 3. Retrieved from: http://www.aclweb.org/anthology/J14-3009 (23-07-2015).

Lacasta, Javier; Nogueras-Iso, Javier; Zarazaga-Soria, Francisco Javier (2010): Terminological Ontologies – Design, Management and Practical Applications. New York: Springer.

Lakoff, G. (1991): Cognitive versus generative linguistics: how commitments influence results. In: Language and communication, Vol. 11, N° 1/2. Retrieved from: https://georgelakoff.files.wordpress.com/2011/04/cognitive-vs-generative-linguistics-in-language-and-communication-lakoff-1991.pdf (14-05-15).

Liebowitz, Jay; Beckman, Tom (1998): Knowledge organizations: what every manager should know. Boca Raton, FL: St. Lucie Pres.

Melby, Alan K.: Terminology in the age of multilingual corpora. Retrieved from http://www.jostrans.org/issue18/art_melby.php (21-05-15).

Minsky, Marvin (1975): Minsky's frame system theory. In: Proceeding TINLAP '75 Proceedings of the 1975 workshop on Theoretical issues in natural language processing. Association for Computational Linguistics Stroudsburg, PA, USA. Retrieved from: http://www.aclweb.org/anthology/T75-2022 (20-08-2015).

Neumann, Idriss (2012): Initiation à la conception de bases de données relationnelles avec MERISE (Retrieved from http://ineumann.developpez.com/tutoriels/merise/initiation-merise/ (20-07-2015).

Noy, Natalya F.; McGuinness, Deborah L. (2001): Ontology Development 101: A Guide to Creating Your First Ontology. Retrieved from: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf (09-09-2015).

Pelz, Heidrun (1996): Linguistik: eine Einführung. Hamburg: Hoffmann und Campe.

Pinker, Steven (2000): Wörter und Regeln. Heidelberg: Spektrukm Akademischer Verlag.

Rapaport, J. William (2005): Meinongian Semantics and Artificial Intelligence. Retrieved from http://www.cse.buffalo.edu/~rapaport/Papers/simons.pdf (20-08-2015).

Roche, Christophe (2007): Le terme et le concept: fondement d'une ontoterminologie. Retrieved from http://arxiv.org/ftp/arxiv/papers/0801/0801.1275.pdf (18-05-2015).

Roche, Christophe (2012): Ontoterminology: How to unify terminology and ontology into a single paradigm. Retrieved from http://www.lrec-conf.org/proceedings/lrec2012/pdf/567_Paper.pdf (27-05-2015).

Roche, Christophe (2013): Ontology: Re-thinking Terminology (Applications of Cognitive Terminological Theories in Terminology Management –Zagreb). Retrieved from http://struna.ihjj.hr/uploads/actt/Roche_Zagreb_2013.pdf (22-09-2015).

Roche, Christophe (2015): Ontological definition. In: Kockaert, Hendrik J.; Steurs, Frieda (eds.): Handbook of Terminology, Vol. 1. Amsterdam: John Benjamins.

Saeed, John I. (2009): Semantics. UK: Wiley-Blackwell Ltd.

Schmitz, Klaus-Dirk (2005): Developments in computational terminology management and its influence on terminology science and terminology work. Retrieved from: http://www.eleto.gr/download/Conferences/5th%20Conference/5th_24-11-Schmitz%20Klaus-Dirk_Paper.pdf (01-06-2015).

Sharoff, Serge; Hartley, Anthony (2012): Lexicography, terminology and ontologies. Retrieved from http://corpus.leeds.ac.uk/serge/publications/termino-v3.pdf (03-02-2015).

Stock, Wolfgang G.; Stock, Mechtild (2013): Handbook of Information Science. Berlin/Boston: de Gruyter GmbH.

Stolze, Radegundis (1994): Übersetzungstheorien – Eine Einführung. Tübingen: Narr.

Taye, Mohammad Mustafa (2010): Web-Based Ontology Languages and its Based Description Logics. In: The Research Bulletin of Jordan ACM, Vol. II. Retrieved from: http://ijj.acm.org/volumes/volume2/no1/ijjvol2no1p1.pdf (20-09-2015).

Uschold, M. King; Moralee, S.; Zorgios, Y. (1998): Special Issue on Putting Ontologies to Use The Enterprise Ontology. The Knowledge Engineering Review, 13(1):31.89.

Vater, H. (1996): Einführung in die Sprachwissenschaft. München: Fink Verlag.

Voltmer, Leonard (2005): Werkzeuge für Rechtsdatenbanken – Über computerlinguistische Verfahren zur Untersuchung, Speicherung und Kommunikation rechtlichen Wissens. Retrieved from: http://edoc.ub.uni-muenchen.de/3716/1/Leonhard_Voltmer.pdf (02-03-2015).

W3C (2013): Extensible Markup Language (XML) 1.0. Retrieved from: http://www.w3.org/TR/REC-xml

Weller, Katrin (2010): Knowledge Representation in the social Semantic Web. Berlin/New York: Walter de Gruyter GmbH & Co. KG.

Wright, Sue Ellen; Budin, Gerhard (eds.) (2001): Handbook of Terminology Management (Vol. 2). Application-oriented Terminology Management. Amsterdam/Philadelphia: John Benjamins.

Appendix A

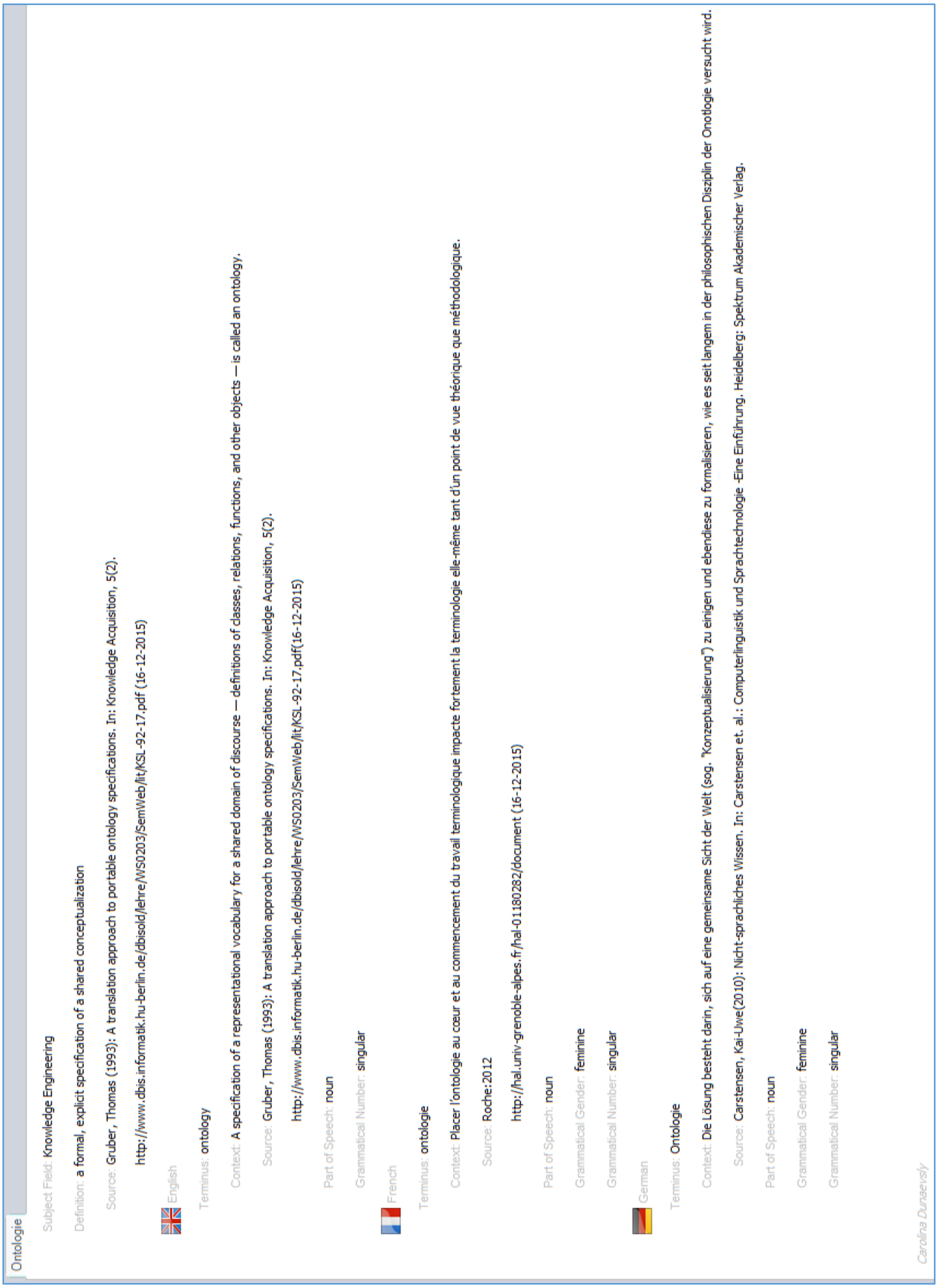


Figure 36: TE sample of the created TDB

Appendix B

The expression	RDF/RDFS	OWL
Class	√	√
rdf: Property	√	√
rdfs: subClassOf	√	√
rdfs: subPropertyOf	√	√
rdfs: domain	√	√
rdfs: range	√	√
Individual	×	√
sameClassAs	×	√
samePropertyAs	×	√
sameIndividualAs	×	√
differentIndividualFrom	×	√
inverseOf	×	√
TransitiveProperty	×	√
SymmetricProper	×	√
FunctionalProperty	×	√
InverseFunctionalProperty	×	√
allValuesFrom	×	√
someValuesFrom	×	√
minCardinality	√	√
maxCardinality	√	√
cardinality	√	√
oneOf	√	√
disjointWith	×	√
complementOf	×	√

Table 5: Differences between RDF/RDFS and OWL (Taye, 2010:15)

Erklärung

Ich, Carolina Dunaevsky, versichere, die von mir vorgelegte Arbeit selbständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keiner anderen Prüfungsbehörde vorgelegen.