OPAALS PROJECT

Contract n° IST-034824

WP6: Socio-Economic Constructivism and Language

Del. 6.3 – Metaphorological Tool Kit

Project funded by the European Community under the "Information Society Technology" Programme

Contract Number: IST-034824

Project Acronym: OPAALS

Deliverable N°: 6.3

Due date: November 2007

Delivery Date: January 2008

Short Description:

Metaphors play a central role in OPAALS and Digital Ecosystems: they do not only represent a specific linguistic form, they also shape how we perceive and interact in our daily lives. This deliverable therefore depicts the first dimension/phase of the development of a metaphorology for Digital Ecosystem research, whereas "metaphor" can also be interpreted as representing language in general.

Author: Oxana Lapteva, Frauke Zeller

Partners contributed: UniKassel

Made available to: Consortium

Versioning		
Version	**Date**	**Name, organization**
V1	15.10.2007	Lapteva, UniKassel
V2	09.11.2007	Lapteva, Zeller, UniKassel
V3	07.12.2007	Lapteva, Zeller, UniKassel
V4	11.12.2007	Lapteva, UniKassel
V5	22.12.2007	Lapteva, UniKassel

Quality check

Internal Reviewers: Ossi Nykänen (TUT)

Dependences:

Work Packages	4, 6, 7, 10
Partners	LSE, CreateNet, T6-Eco, WIT
Domains	Social science: linguistics, computational linguistics, empirical media studies
Targets	All partners

Table of Contents

1) Introduction

Metaphors play a central role in OPAALS and Digital Ecosystems. They do not only represent a specific linguistic form, they also shape how we perceive and interact in our daily lives. Furthermore, they carry a high potential regarding the communication of new and innovative ideas, such as “Digital Ecosystem”, which is itself a metaphor and, simply speaking, aims at transferring certain well-known concepts in order to depict a new concept for the sake of making itself understandable.

According to Lakoff and Johnson (1980), “metaphor is pervasive in everyday life, not just in language but in thought and action” (p. 3), however, to a great extent we are not aware of this due to the fact that language usage is mostly done unconsciously or automatically, not thinking of grammar, syntax or word-choice when communicating in our mother tongues. Hence, discussing a socio-constructivist approach to metaphors in OPAALS means that we are confronted with two dimensions: Firstly, the socio-constructivist potential of language/communication in general (including metaphors) which shapes and constructs our reality (see for example Luckmann, 2005). This means that each partner of OPAALS automatically contributes metaphors from their daily (working) lives to the project community, often in a rather unconscious manner, which transforms OPAALS to a “razzmatazz” of metaphors. Secondly, we deliberately use the socio-constructivist potential of language and try to build a Digital Ecosystems concept throughout metaphorical constructs as an important means to create an interdisciplinary research community.

It goes without saying that the metaphor razzmatazz cannot exclusively be played by ear, as the different concepts, understandings, and actions need to be understood in order to work across different disciplines. To be more precise, the powerful potential of metaphors in terms of linguistic expressiveness and creativity also bears the high risk of misunderstandings and misconceptions when bringing together people from different background. For example, the first results of a survey conducted among the OPAALS researchers (in task T6.1) provided interesting insights into the inherent variety of concepts regarding the same terms. The survey included word associations and used terms that have a central meaning and relevance in the project. The term “network” (“network” is a prominent metaphor used in various research directions, such as software development, language learning, policy and governance, etc.) was among them and the associations so far gathered already prove a strong differentiation between the two main domains social science and computer science. Computer scientists mostly connect “network” with the following characteristics: *connected nodes p2p, on-line, networking, platforms, information flow.* Whereas social scientists used *social network, social capital, community relations,* and *common space* to depict the term “network”. These first findings already provide a first glimpse on the intricateness of language usage in OPAALS and DE in general: the strong potential and power of language in terms of socio-constructivism, and also the infinite variety and creativity that constitutes natural language.

Therefore, a structured and analytic approach to metaphor is necessary which would enable us to apply metaphors to Digital Ecosystems without creating misunderstandings or false expectations. This approach is called “*metaphor*ology” in task T6.2, being a composition of “metaphor” and the greek ending “-ology” meaning “the study of”. Whereas the creative potential and power of metaphor is an important point of departure in our work, we will use for our metaphorology a more applied approach regarding its actual applicability in Digital Ecosystems research (hence the “tool kit”). This means, that we will not focus on the theories and discussions of metaphor in philosophy (see, for example, Derrida, 1982; Morris, 2000; Ricoeur, 1975), but merely take Ricoeur's statement “Il n'y a pas de métaphores dans le dictionnaire”, meaning that there is no such thing as a dead metaphor as an important point of departure for our metaphorology. Specific philosophical elaborations on metaphor, identity, and language, however, will inspire research in work package 10, task 10.6 “Epistemological basis for OPAALS' interdisciplinary research”, and research in work package 12, task 12.1 “Systems theory, associative systems, and autopoiesis” (see for further explanations page 9).

It should be mentioned that this is the first, hence an introductory deliverable in task T6.2 “Theoretical implications for a structural grammar of metaphors (*metaphor*ology) in DEs”, and that a second deliverable is planned in month 30 of Phase II.

Task T6.2 is therefore divided into three different phases and/or dimensions:

1) introductory overview of existing theories on metaphor and computational applications leading to a first <u>draft</u> of a Metahorological Tool Kit (documented in deliverable D6.3);
2) test and refinement phase; data collection phase (documented in deliverable D6.7, and first results presented for example in the OPAALS wiki);
3) the self-referential loop: applying the results from 1 and 2 to socio-pragmatic questions regarding community building, and to the development of an evolutionary framework for language (documented in deliverable D6.7).

More concretely, the first phase/dimension, that is to say this deliverable, focuses on an introduction to metaphor which will lead to a computational framework for metaphor processing. It concludes in a first draft of a Metaphorological Tool Kit, which is a platform for metaphor analysis including a variety of possible methods and techniques from computational linguistics that are relevant for a metaphorology of digital environments. During the second phase/dimension, we will then use this platform in order to expand and refine the framework for its application in Digital Ecosystems. The results of the test and refinement phase will be useful for work package 10, specifically tasks T10.5 “Visualisation of the OKS”, T10.11 “Social networks in cross-domain collaboration, T10.12 “Knowledge models and representations”, T10.14 “Meta-data tagging tools and environments”, and T10.15 “Distributed semantically searchable repository”. Additionally, some of the new partners, among them FAO, have expressed a keen interest in the work of UniKassel on metaphor and

language representation that combines a sound communication theoretic background with methods and techniques from computational linguistics.

We will then use the data/results from our computational framework in the third phase/dimension (the self-referential loop) to discuss from a new and innovative perspective the socio-constructivist character of metaphor and language in general, and its socio-pragmatist and theoretical implications will be analysed and transferred to the OPAALS research network. Ricoeur's emphasis that there are no “dead metaphors” which can be stored in a lexicon supports a strong socio-constructivist notion of metaphor and communication in general as it integrates notions of communication shaping society, of situatedness, and dynamics of language. It also points towards the fact that we cannot simply prescribe certain metaphors for the development of a new community in terms of putting them down in a fixed vocabulary book for Digital Ecosystems. In order to build a new community by means of a socio-constructivist language approach we need to analyse the traditional, well-known language and metaphors used by all stakeholders that will shape a new community (i.e. their individual domain languages). Moreover, we add another complexifying point to the already challenging task of building a community: the building of a community in a digital environment. This means that we need to apply techniques and methods from computational linguistics in order to be able to analyse the actual language and metaphor usage in a community that finds itself in a digital environment, that is to say physical and tangible objects or demarcation lines are not existent. This spatial determinant also influences our communication behaviour, patterns, etc., which is well-documented an analysed in the fields of computer-mediated communication, computer-mediated discourse analysis, and digital media theory.

Our overall metaphorology approach then can be called new and innovative as it develops a sound basis of digitally gathered and enriched natural language data that can be mapped to both social and computer science theories regarding digital community building, etc. This means that we try to avoid a predetermining selection of an either techno-centric or socio-centric approach to communities in digital environments. By means of building upon theories and methodologies from social science we add the technological dimension by a sound computational analysis of a community's language(s). Deliverable D6.7 will then show and discuss the results of our approach and will also connect the results to task T6.6 in work package 6 “Evolutionary framework for language”.

2) A short introduction to metaphor

This chapter offers a first introduction to metaphor. Whereas chapter 3 concentrates on the introduction to different theories and approaches of metaphor research, this chapter aims at providing a first general understanding of what metaphors are, their definitory roots, their influence on communication and culture, what different types of metaphor exist, and finally what a grammar of metaphor is. These different points of view had been chosen as only few among many more aspects of metaphor.

However, as giving a full and exhaustive report on metaphor research is not the intention of this deliverable, the focal points of chapter 2 had been chosen as those which will play an important role either in the course of this deliverable or in future work in OPAALS. For example, the grammar of metaphor gives a first impression of a structured and analytic approach to such a powerful example of the creativity of natural language. By means of this, we hope to create a certain awareness for the Metaphorological Tool Kit, where metaphors are dealt with in a computational environment, which is also based on a structured and analytic approach.

2.1) Overview: definition and roots

Metaphors are a powerful means of human language. We use metaphors in our daily life to describe our world in both familiar and innovative ways. Studying metaphors occurring in written resources and in everyday conversation poses interesting challenges for theories of language and meaning. But what exactly is “metaphor”? The word “metaphor” comes from Greek *metapherein* meaning “to transfer”. The Oxford English Dictionary defines metaphor as:

- A figure of speech in which a name or descriptive word or phrase is transferred to an object or action different from, but analogous to, that to which it is literally applicable; an instance of this, a metaphorical expression.
- Something regarded as representative or suggestive of something else, especially as a material emblem of an abstract quality, condition, notion, etc.; a symbol, a token.

Metaphor can also be defined as a “word or expression that is used to talk about an entity or quality other than that referred to by its core, or most basic meaning. This non-core use expresses a perceived relationship with the core meaning of the word, and in many cases between two semantic fields” (Deignan, 2005, p. 34).

To simplify the provided definitions, we can say that metaphor means the process of transformation through using one entity to describe another. A common understanding of metaphor is the existence of two areas (domains or concepts) in metaphor. Different theories apply different names for these domains and consider different aspects of metaphor. One calls them “tenor and vehicle”, “focus and

frame", another "source and target". A tenor is associated with the principal subject whereas a vehicle is referred to something that is metaphorically attributed to the tenor (Way, 1991). For example, in the phrase *Life is a walking shadow, "*life" represents the tenor and "walking shadow*"* the vehicle.

The roots of metaphor research and definition refer to Aristotle who claims in *The Poetics*: "Metaphor is the application of word that belongs to another thing: either from genus to species, species to genus, species to species, or by analogy" (Aristoteles, 1995, p. 105). Aristotle represents the traditional view of metaphor by understanding it as a formula for achieving more colourful expression.

There are a number of obvious problems with Aristotle's treatment of metaphor (Veale, 1995):

(1) Aristotle does not take into account that many metaphors are semantically related to each other.

(2) Aristotle's theory observes metaphor as much as one term stands for, or re-places another. Metaphor is described as a formula for achieving more colourful expression; as an essential feature of human communication it is not considered at all.

(3) Metaphors have no representational status in themselves. Hence they are not seen as active conceptual agents in the construction of other concepts.

(4) According to Aristotle's theory metaphor exists in the rules of knowledge use, rather than in the knowledge itself.

2.2) Metaphor, communication and culture

Metaphor is a communicative resource by which language users may enhance "the expressiveness of their message through the most economical means available to them" (Charteris-Black, 2004, p. 17). It provides a method for cultures to express their particular identity and ideas. Metaphors allow us to produce new meanings of a term or concept, stretching and enhancing its usage. Moreover, metaphors affect social relations among the participants within organisations, communities or society in general. According to Kövecses (2002), the main focus of metaphors is "the culturally agreed-on conceptual material associated with the source that it conventionally imparts to its targets" (p. 118).

Two different kinds of cultural variation of metaphors have been defined by Kövecses (2002):

(a) cross-cultural (intercultural)

Cultures and languages often differ in their concepts and conceptual metaphors available for the conceptualisation of particular target domains. Kövecses identifies two "categories of causes that bring about cultural variation in metaphor" (p. 186): broader cultural context and the natural and

physical environment. It is obvious that the environment shapes language, primarily its vocabulary. Consequently, it influences the formation, development and death of the metaphors as well.

(b) within-culture (intracultural)

This type of cultural variation has different sources such as human concern and personal history. Regarding Digital Ecosystem research we are faced with both intracultural and intercultural metaphor usage in a multilingual environment. Regarding the intracultural metaphor dimension, Nöth (2006) describes them as "semiospheres" by building on the works of Lotman (1990):

> 'Thought is within us, but we are within thought,' or 'the world is both within us and without us' [1990]. Such enigmatic paradoxes in Lotman's semiotic rhetoric reflect a view of culture as a self-referential system in which semiotic spaces are embedded in more encompassing isomorphic spaces of cultural semiosis. Metaphors, according to this theory, are semiospheres representing mental images by means of verbal signs. Since space plays such an important role in Lotman's theory of metaphor, his own metaphors of culture as semiotic spaces can thus, themselves, be read as self-referential.

The clear reference to systems theory and autopoiesis indicates that the language approach, which can be seen as a red line to be followed in the NoE project, will provide interesting insights regarding the development of an integrative theory for Digital Ecosystems. The notion of culture as semiotic spaces that are self-referential will be elaborated in deliverable 12.1 "Foundations of the Theory of Associative Autopoietic Digital Ecosystems, Part 2", whereas deliverable 1.2 "Foundations of the Theory of Associative Autopoietic Digital Ecosystems, Part 1" contains an introduction to the fields of semiotics and biosemiotics as a promising approach to Digital Ecosystems research and autopoiesis.

However, apart from the importance of verbal representations of spatial relationships there are many other reasons why people use metaphors in their everyday lives. Sometimes, there are no other words they could use to express ideas, feelings or to refer to a specific thing. In general, however, even if we have a choice, we use metaphors in order to communicate, to explain things in a more creative and interesting way (Knowles & Moon, 2006). Therefore, the focus on the aspects of communication and cultural specifics of a community enrich the overall socio-linguistic research of Digital Ecosystems. The reconstruction of metaphorical models used in written and spoken language provides useful information for Digital Ecosystems. Besides the analysis of linguistic features in natural language, these models reflect the organisational-communicative aspects of metaphors (word choice depending on the context and situation, metaphors at different levels of organisational hierarchy, domain-specific metaphors, etc.). Furthermore, they provide the possibility to reconstruct the research participants' metaphorical points of view and overall cultural phenomena. Metaphorical descriptions of relationships in which two people speak about their joint activities offer a further level of analysis (Schmitt, 2005).

2.3) Types of Metaphor

Depending on theories and approaches, metaphors can be seen as similes, analogies, or models. *Metaphors as similes* can be characterised as statements of the form “A is like B” that gives a strong association between A and B (Indurkhya, 1992).

In *analogy*, the association is weaker. Parts of B may be compared with parts of A, but B is not considered to be the same as A. Describing metaphors as analogies, Indurkhya (1992) identifies three classes of them: simple, proportional and predictive. The simple analogies refer to similarities between two objects or situations. Proportional analogies are often characterised as statements of the form “A is to B as C is to D”. The proportional analogies are symmetric, i.e. the terms B,D can be interchanged with the terms A,C, respectively, without affecting the meaning of the analogy. Predicting analogy is defined as “the process of predicting further similarities between two objects or events on the basis of some existing similarities between them” (Indurkhya, 1992, p. 32).

Finally, metaphors as *models* contain an object (situation, phenomenon) being modelled, and there is another object, which is a structured set of symbols, that represents or describes the target. An element of interpretation need to be considered, since the symbols must be interpreted appropriately in the context of the target for the model to be meaningful (Indurkhya, 1992).

Shelestiuk (2006) developed a metaphor taxonomy consisting of three categories: semantic, structural, and functional category.

a) Semantic category

This group falls into subcategories based on the different types of associations:

a.1) Associative links between the vehicle and tenor. These links build several kinds of similarities as shown in Table 1.

Similarity	***Example***
Similarity of functions	The hands of the clock
Similarity of form	A bottle's neck
Similarity of structure and substance	A flood of tears
Similarity of result	He evaporated

Table 1: Similarities based on associative links (cited from Shelestiuk, 2006, p. 337)

a.2) Logic-grammatical meaning

This subcategory reflects a process of nomination regarding the ground of metaphor. The ground

can be of different kinds; Table 2 lists a few examples.

Ground	***Example***
the characteristic of a substance through another substance	'It [the sun] struck upon the hard sand and the rocks became furnaces of red heat' (V. Woolf);
the characteristic of a substance through an action	'. . . there was a stir and bustle among the stars' (F. S. Fitzgerald);
the characteristic of an action through a substance	'They [the waves] serpented towards his feet' (J. Joyce)
the characteristic of an action through another action	'He watched him closely while he excavated his smile' (G. Greene);
the characteristic of an action through a quality	'Mrs. Cloudesley Shove blackens the doorway with her widowhood' (A. Huxley);
the characteristic of a quality through another quality	'. . . by day beside a livid sea, unbeheld, in violent night walking beneath a reign of uncouth stars' (J. Joyce).

Table 2: Metaphors based on logic-grammatic meaning (cited from Shelestiuk, 2006, pp. 337-338)

a.3) Associations based on the subject of the vehicle

Examples of this type of metaphor classification describe part-function associations of "the human body, animal, bird, flower, etc., according to which metaphors may be anthropomorphic, zoomorphic, vegetative, etc." (Shelestiuk, 2006, p. 338).

a.4) Concreteness-abstractness of the tenor

Based on this type of association, one may distinguish between concrete and abstract metaphors. For example:

Concrete: the company is growing fast;

Abstract: that idea will go a long way.

b) Structural category

This type of classification includes metaphors based on the part-of-speech (nounal, verbal, adjectival, adverbial metaphors) and part-of-sentence (substantive, predicative, attributive, adverbial metaphors) differentiation. This type of classification can be based on:

- formal limitations of metaphor (word-metaphors, phrasal metaphors, and others)
- division into simple and extended metaphors. According to Shelestiuk (2006), in the extended metaphors "one metaphorical statement is followed by another, containing a logical development of the previous metaphor (e.g., This is a day of your golden opportunity. Don't let it turn to brass)" (p. 339).

c) Functional category

The functional category of metaphors refers to the idea that "the semantic content of words is formed in compliance of their role in an utterance" (Shelestiuk, 2006, p. 340). Some of the communicative functions are identification of subjects of speech and predication, introducing their properties and characteristics, and the pragmatic meanings of words. In the process of identification of metaphorical expressions Shelestiuk differentiates between two types of speech signs:

- mono-functional (only identifying, e.g. names and pronouns; or only characterising, e.g. verbs and attributes)
- bi-functional (both the identifying and characterising functions dependent on their specific role in a sentence)

2.4) Grammar of metaphors

In order to understand metaphor we need to analyse its linguistic information. Therefore, defining syntactic rules is one of the fundamental processes of metaphor identification, and in order to define and describe metaphor, we need to discuss grammatical patterns of it. Generally, studies on the grammar of metaphor take different perspectives, however, they provide two important points (Deignan, 2005):

- there seem to be quantitative and qualitative differences in the way that different parts of speech behave metaphorically.
- The target domain use tends to be different from the source domain use, whether at the major level of part of speech or at more detailed level. Metaphorical uses seem to be subject to much more restricted choices than literal uses.

It is important to note that metaphors can take different forms and structures: from single words to the phrases and sentences. On the word-level, we often deal with terminological/conceptual and at the same time metaphorical representations of meaning. Some examples of concepts that are used in OPAALS metaphorically are "Digital Ecosystem", "community building", "knowledge space", and others.

The importance of an analytic approach to metaphor which takes into account the grammatical and syntactical relations is also being prove by the fact that the metaphoric character of single words (noun, verb, adjective, etc.) is closely related to another linguistic case called polysemy. Polysemy involves words that have a number of senses/meanings. According to Kövecses (2002), polysemy is often based on metaphorical relationships between two meanings of the same word. For instance, consider the word *bank*. In the following examples[1], it is used as a physical object and as a metaphor:

*The **bank** was flooded yesterday* (building)

*The **bank** was founded in 1990* (institution)

*A blood **bank**, a memory **bank*** (a place where something is stored)

In the first two sentences the word *bank* refers to a physical object. It is a noun and literal in its use. In the last example the noun version of *bank* changes its original domain and becomes the metaphor "blood bank" or "memory bank".

In case of complex structures of metaphor such as phrases and sentences, different grammatical and syntactical rules of the language determine which sequences of words are syntactically correct. In linguistics, a syntactically correct sequence of words from the vocabulary of the language is called an utterance (Steinhart, 2001). According to Steinhart, an utterance can be defined as a metaphor only if:

a) the words it combines are from distinct fields, and

b) the words it combines are from analogous fields.

To describe characteristics of the grammatical structure of utterances we need to identify the rules. These rules "specify how to transform a syntactical whole into the structure of its parts: a sentence is rewritten as a series of phrases; each phrase – as a series of words" (Steinhart, 2001, p. 30). Such phrases or sentences can either be literal or metaphorical.

An example of the application of syntactic rules to both literal and metaphorical sentences is shown in Table 3.

Syntax rules	***First derivation***	***Second derivation***
S --> NP V NP	S	S
NP --> DET N	(NP1 V NP2)	(NP1 V NP2)
V --> drank, burned	((DET1 N1) V (DET2 N2))	((DET1 N1) V (DET2 N2))
N --> person, car, water, gas	((The person) drank (the	((The car) drank (the gas))

1 Examples are taken from Frath, P. (2001). Polysemy, Homonymy and Reference. *Proceedings of the JASGIL Seminar*, Strasbourg.

DET --> the	water))	

Table 3: Rewrite rules (Steinhart, 2001)

The presented rules reflect the following structure of a sentence:

The sentence (S) contains a noun phrase (NP) followed by a verb (V) and another noun phrase (NP). The noun phrase (NP) is defined as a concatenation of an article and a noun (DET N). Note that the tree structure (Figure 1) of both sentences is the same. However, one sentence is used literally, and the other – metaphorically.

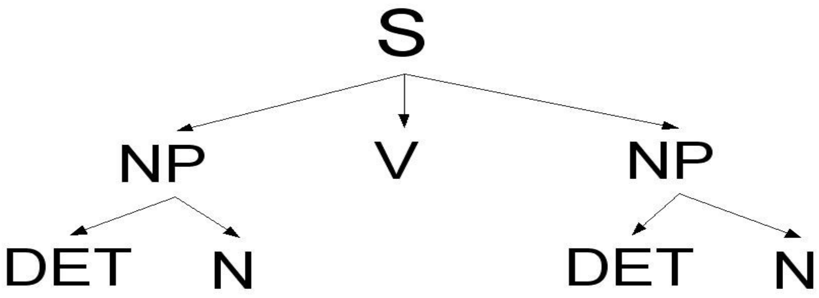

Figure 1: Tree structure of a sentence

The second derivation is the metaphorical transformation, in which a living entity (person) in the sentence is replaced by the non-living entity (car). Important is the fact that this replacement assigns certain specifics (in this case, characteristics of *drinking*) of the living entity to the non-living. This process is an example of metaphor creation in natural language. Here, the first derivation has a non-metaphoric value whereas the second derivation reflects the metaphoric character of the sentence.

In the context of natural language analysis, we need to recognise that the structural representation of metaphors is often more complex and diverse as shown in the previous examples.

Tirell (1991) analysed the grammar of metaphor and distinguished six grammatical types:

(1) Simple identities, of the form "***A is B***". "*Juliet is the sun"*

(2) Pure predications, of the form "***A is F***". "*Juliet is brilliant*"

(3) Sortal predications, of the form "***A is a K***". "*Man is a wolf*"

(4) Substitution metaphors, formed by substituting a term that does not literally apply for one that does; for example, "monster" in the line "Pity this busy monster, manunkind"

(5) Noun-function metaphors, of the form "***The B of A***"; "*A commitment to empiricism lies at the heart of my theory*"

(6) Verb-function metaphors, of the form "***A Vs B***" where V is a verb which A cannot literally do, or which cannot literally be done to B; "*My car drinks gas*"

Steinhart (2001) identifies different grammatical classes (forms) a metaphor can consist of (Table 4). In addition to the syntactic structures (NP – noun phrase, DET – article, ADJ – adjective, GERUND – gerund, INF - infinitive), he provides the distinction of source and target, marked as S(or s) and T(or t) respectively.

Grammatical form	***Example***
NP_**TT** --> {DET} {ADJ_t} NOUN_t	True ideas
NP_**SS** –> {DET} {ADJ_s} NOUN_s	Liveborn babies
NP_**TS** --> {DET} ADJ_t NOUN_s	A male midwife
NP_**TS** --> {DET} ADJ_t GERUND_s	Mentally ovulating
NP_**ST** --> {DET} ADJ_s NOUN_t	Sharp minds
INF_**T** --> TO BE ADJ_t	To be true
INF_**S** --> TO BE ADJ_s	To be liveborn

Table 4: Grammatical classes (Steinhart, 2001, pp. 35)

The grammatical classes are used to generate metaphors. For example, *True ideas* is a noun phrase where both, the noun and adjective, belong to the target domain. The second phrase *Liveborn babies* belongs to the source domain. The metaphorical mapping can be then realised in the sentence "*A true idea is a liveborn baby*".

Another example of grammatical classes is the noun phrase containing an adjective from the target and a noun from the source (*A male midwife*). This class is used, for example, to construct the metaphor "*Socrates is a male midwife*". More generation rules and examples of metaphors can be found in Table 5.

Rule	***Example***
S --> NP_TT BE NP_SS	Socrates is a midwife. ideas are birds. Juliet is the sun.
S –> NP_TT BE NP_TS	Socrates is a male midwife.
S --> NP_TT BE ADJ_s	Her mind is brilliant.

S --> INF_T BE INF_S	To be smart is to be sharp.
S --> NP_TT1 VERB_s {PREP} NP_TT2	John is married to his work.
S --> NP_TT VERB_t {PREP} NP_SS	The geologist married a star.
S --> NP_ST VERB_s NP_TT	Sharp minds cut ideas easily.
S --> NP_ST VERB_s {PREP} NP_ST	Liveborn ideas grow into vigorous theories.

Table 5: Syntactic structure of metaphors (Steinhart, 2001)

As we can see, some of the grammatical forms (e.g. NP_ST = sharp minds) are individual metaphors as well that are then generated into more complex metaphorical expressions (sentences), such as "*Sharp minds cut ideas easily"*.

In the context of digital environments, the extraction of the grammatical masks (templates) and applying them to to the large collections of electronic documents assists not only a linguistic analysis but also computational approaches related to natural language processing, such as metaphor analysis. However, an important question regarding multilingual communities is whether the listener/hearer understands a metaphor's meaning. In this process not only pure linguistic information (grammatical and syntactic structures) is involved. The combination with domain and context specific knowledge provides necessary additional information for a successful metaphor resolution. Furthermore, in the context of digital ecosystems the processes of digital transmission, extraction and interpretation of metaphors are of importance. Hence, we are going to study metaphors not only from linguistic and cognitive but also from social and computational perspectives.

3) Theories and approaches of metaphor research

3.1) Overview

This chapter provides a comprehensive and integrative overview of metaphor (research) for Digital Ecosystems. The main emphasis lies on linguistically inspired approaches to metaphor which build an important basis for an applied and computational framework for understanding, applying, and processing metaphor in Digital Ecosystems. For example, sub-chapter 2.4 depicted a formal and analytic introduction to metaphor research based on linguistic analysis methods such as the grammatical and syntactical dimensions of metaphor. However, the complexity of metaphor in language requires the analysis of a variety of theories from different research domains in order to firstly fully grasp its complexity on the one hand (which could impede communication processes in Digital Ecosystems), and secondly to derive at an integrative approach to metaphor research in Digital Ecosystems. This means, as the name "digital ecosystem" already indicates, that an integrative approach also needs to contain a formal and computational part in order to be applicable to digital environments.

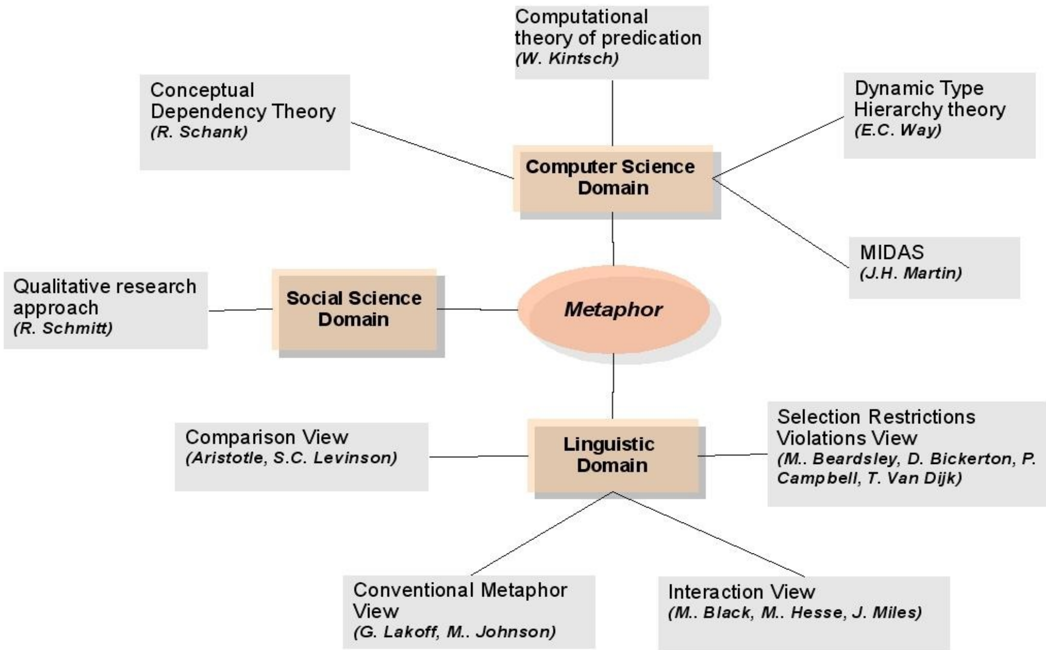

Figure 2: Theories and approaches in metaphor research

The following sub-chapters therefore mainly visualise the complexity of metaphor (research) by introducing shortly the most important theories and methods (Figure 2). Among the four metaphor concepts from linguistics, the Conventional Metaphor View will be the focal point as it represents the current predominant and a very exhaustive conceptual framework. Introductions of metaphor

concepts and approaches from the computer science domain will then lead to the Metaphorological Tool Kit in chapter 4 which provides a strong computational and applied agenda for communication and language analysis in digital environments.

The socio-constructivist analysis will be applied by means of the self-referential loop (see introduction) in the second deliverable of task T6.2, on the basis of a further developed and refined metaphorology together with a comprehensive set of data gathered and analysed by means of the Metaphorological Tool Kit.

3.2) Metaphor concepts from linguistics

3.2.1) The Conventional Metaphor View

The currently most popular view of metaphor was first developed by Georg Lakoff and Mark Johnson in 1980 in their seminal study *Metaphors we live by*. Their concept has become known as the "cognitive linguistic view of metaphor" (or Conventional Metaphor View). Lakoff and Johnson challenged the deeply entrenched view of metaphor by claiming that (Kövecses, 2002, pp. vii—viii):

1. metaphor is a property of concepts, and not of words;
2. the function of metaphor is to better understand certain concepts, and not just some artistic and aesthetic purpose;
3. metaphor is often not based on similarity;
4. metaphor is used effortlessly in everyday life by ordinary people, not just by specifically talented people;
5. metaphor, far from being a superfluous though pleasing linguistic ornament, is an inevitable process of human thought and reasoning.

They define metaphor as:

> [...] metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature. [...] Our concepts structure what we perceive, how we get around in the world, and how we relate to other people. Our conceptual system thus plays a central role in defining our everyday realities. If we are right in suggesting that our conceptual system is largely metaphorical, then the way we think, what we experience, and what we do every day is very much a matter of metaphor. (Lakoff & Johnson, 1980, p. 3)

This statement indicates towards a socio-constructivist notion of language which shapes our

realities. Secondly, it states that our conceptual system is to a high extent metaphorical and therefore, metaphor being much more than only a poetic or rhetorical trope. In fact, metaphors can structure our everyday activities, which sounds on first hearing rather odd. Metaphor being regarded largely as a certain linguistic form that is mainly used in poetic language, could therefore not possibly have such a strong effect on our everyday life. However, the pervasiveness of metaphorical usages in everyday life languages is easily being proved. Lakoff and Johnson give the famous example of “argument is war”, whereas Kövecses (2002) uses in Table 6 the concept of 'love is a journey' as an example for metaphorical richness (please note that “love” is also often used in combination with war metaphors, such as “love is a battlefield”):

“Argument is war”	**“Love is a journey”**
Your claims are *indefensible.*	We're *at a crossroads*.
His criticisms were *right on target*.	We're just *spinning our wheels*.
I *shot down* all of his arguments.	Our marriage is *on the rocks*.
I've never *won* an argument without him.	We'll just have to *go our separate ways*.
You disagree? Okay, *shoot*!	We're *stuck*.
If you use that *strategy*, he'll *wipe you out*.	I don't think this relationship is *going anywhere.*

Table 6: Metaphors “Argument is war” and “Love is journey”

Lakoff and Johnson as well as Kövecses follow the dominant trait in linguistics regarding metaphor research, the conceptual metaphor. The conceptual metaphor is strongly connected to cognitive linguistics and its advantage is that it refers to a clear-cut structural approach, where metaphor concepts are often explained by means of sets of mappings. For example, the “love is a journey” concept can be explained accordingly (cited from Kövecses, 2002, p. 7):

Source: Journey		**Target: Love**
the travelers	->	the lovers
the vehicle	->	the love relationship itself
the journey	->	events in the relationship

the distance covered	->	the progress made
the obstacles encountered	->	the difficulties experienced
decisions about which way to go	->	choices about what to do
the destination of the journey	->	the goal(s) of the relationship

Strongly relevant regarding the development of a systematic approach to metaphor in Digital Ecosystems is the fact that the language of economics (i.e. language of SMEs) is itself a rich source of metaphor usage. Primary example is the language of the stock market, where the performance of stocks and bonds is usually depicted by means of rich metaphoric concepts. More profoundly, the basic notion of 'organisations' and 'companies' is also often transferred by means of concepts from other domains, such as biology or living systems in general. However, especially when dealing with conceptual objects rather then tangible objects, such as “trust”, “security” or “sustainability”, we find that they are often used by other disciplines in a metaphorical way. Another example by Kövecses leads us directly to the usage of metaphor in Digital Ecosystems, 'social organisations are plants', for example:

- He works for a local *branch* of the bank.
- Our organisation is *growing*.
- They had to *prune* the workforce.
- The organisation was *rooted* in the old church.
- There is now a *flourishing* black market in software there.
- His business *blossomed*.
- Employers *reaped* enormous benefits from cheap foreign labour.

Consequently, social organisations are compared to plants or biological forms in general with the following set mappings:

Source: Plant		**Target: Organisation**
the whole plant	->	the entire organisation
a part of the plant	->	a part of the organisation
growth of the plant	->	development of the organisation
removing a part of the plant	->	reducing the organisation

the root of the plant	->	the origin of the organisation
the flowering	->	the best stage, the most successful stage
the fruits or crops	->	the beneficial consequences

Hence, the Digital Ecosystems metaphor seems to be well chosen and aligned with general descriptive frameworks as it represents the framework for social organisations or companies and emphasises the notion of an environment for living systems that is again based on concepts of autopoietic and associative systems.

The set mappings can lead to the assumption that metaphor resolution depends on a dualist model that could be easily adaptable to a computational framework, and that any source could be used to comprehend any target. The second assumption also indicates that we are dealing with completely arbitrary conceptual metaphors, which is – looked at metaphor use in practice – not the case. Therefore, we must assume that there are certain limitations on what can become a conceptual metaphor, so what are the limits that motivate metaphorical links between A and B. Understanding these limitations and rules would of course be essential for processing natural language metaphors in a formal language environment.

Gibbs (1999) argues that there are two ways how metaphor can structure conceptual representations. The first way proposes "that many concepts are not understood via their own representations but by metaphorical connections to knowledge in different domains" (p. 147). The second way suggests that people have well-developed, independent concepts. Some of these concepts are often metaphorically linked to other concepts with similar structure.

Beyond the identification of the relations between concepts and their metaphorical mappings, it is important to understand "how conceptual metaphors coalesce into a system of cultural meanings that inform all sorts of representations, symbols, rituals, and activities" (Danesi, 2006, p. 187). These processes are defined as idealised cognitive models. Such models are useful for investigating semantic structures of language and so-called "cultural groupthink" (ibid, p. 188).

Danesi defines three basic types of the idealised cognitive models:

1. Clustering

This type of model crystallises "from layers of source domains that are formed by implication to provide different perspectives of the same concept or target domain" (Danesi, 2006, p. 192). It means that certain source domains implicate the same type of concept and refer to the same target domain. Models based on clustering principle help to explain why we "use different metaphorical vehicles to deliver the same concept" (ibid, p. 199).

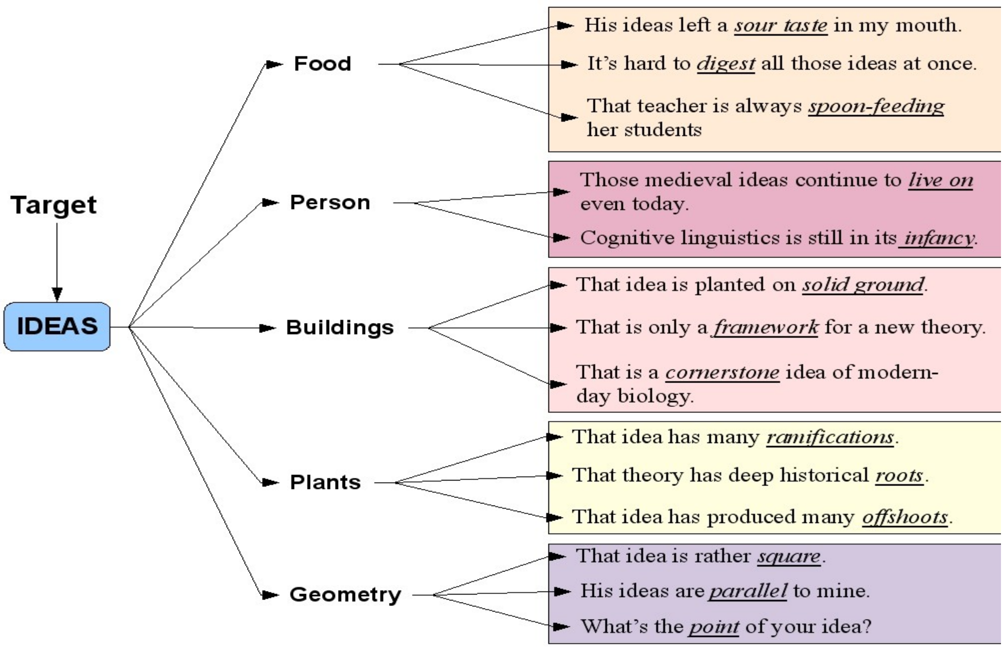

Figure 3: Clustering model

2. Radiation

This model refers to the tendency of "radiation outwards" from a single source domain to different target domains. This means that a single source domain can be mapped to different target domains. Danesi (2006) provides an example of radiation existing between the source "Journey" and the targets "Life", "Love", and "Friendship".

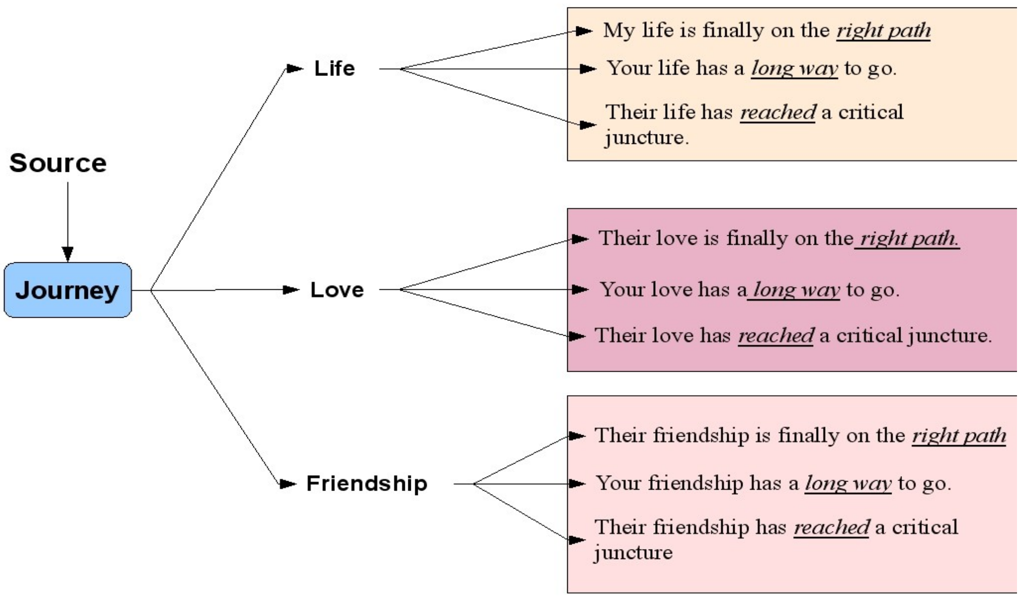

Figure 4: Radiation model

The principle of radiation helps to explain why we talk of seemingly different entities with the same metaphorical vehicles.

3. Ordering

This type of modelling metaphors refers to the formation of different levels of conceptual metaphors. Danesi (2006) differentiates between low-order and higher-order metaphors (see Table 7).

Ordering type		Example
Low-order	Seeing is believing	I have a different point of view
Higher-order	The think up construction	Where did you think that idea up?

Table 7: Low-order and high-order metaphors

According to Danesi (2006), "the lower the order of the conceptual metaphor, the more universal it tends to be; the higher its order the more culture-specific it is apt to be" (p. 195). In the lower-order conceptual metaphors the source domains implicate abstractions such as thought, knowledge, understanding, etc.).

Regarding a categorisation of the different types of conceptual metaphor, Lackoff and Johnson (1980) differentiate three types: structural, orientational, and ontological metaphors.

Structural metaphors are presented by a concepts that is metaphorically structured in terms of another. One example of structural metaphors is "Time is money":

You are wasting my time

This gadget will save you hours

Orientational metaphor organises concepts by giving them a spatial orientation. These representations of orientational metaphors reflects the structure of our bodies and our environment. A lot of them have to do with spatial orientation such as up-down, in-out, front-back, central-peripherial, etc. One example of orientational metaphor is "Happy is up":

I'm feeling up

Thinking about her always gives me a lift

According to Lackoff & Johnson (1980), such metaphors have their roots in our physical and cultural experiences.

Ontological metaphors reflect different way of viewing intangible concepts, such as feelings, activities, ideas and. The abstraction of these entities is represented as something concrete, for example a person, an object, or a container.

The sentences below are examples of the container metaphors (Lakoff & Johnson, 1980, p. 51):

I've had a full life

Life is empty for him

Her life is crammed with activities

Get the most out of life

3.2.2) The Comparison View of metaphor

This traditional concept roots in the Aristotelian view of metaphor and can be briefly characterised by pointing out five of its most commonly accepted features (Kovecses, 2002, pp. vii-viii):

- ➔ metaphor is a property of words; it is a linguistic phenomenon.
- ➔ metaphor is used for some artistic and rhetorical purpose.
- ➔ metaphor is based on a resemblance between the two entities that are compared and identified.
- ➔ metaphor is a conscious and deliberate use of words, and you must have a special talent to be able to do it and do it well.
- ➔ metaphor is a figure of speech that we can do without; we use it for special effects, and it is not an inevitable part of everyday human communication.

According to the comparison view,

> [...] a metaphor is a comparison in which one term (the tenor or subject of the comparison) is asserted to bear a partial resemblance (the ground of the comparison) to something else (the vehicle), the resemblance being insufficient to sustain a literal comparison. As with any comparison, there is always some residual dissimilarity (the tension) between the terms involved in the comparison, but comparison theorists tend not to emphasize this dissimilarity. (Tourangeau & Sternberg, 1982, p. 205)

Despite the fact that this theory survived for centuries and provided important aspects of metaphor research, only single aspects and ideas (e.g. linguistic representation, analogies and metaphors) can be adapted and applied to the areas of Digital ecosystems and Semantic Web.

3.2.3) The Interaction View

Interaction theorists differentiate between tenor and vehicle of metaphor. Max Black (1962) calls them frame and focus of metaphor. The focus of a metaphor refers to the word in the metaphorical expression used metaphorically. The frame refers to the words in the metaphorical expression that are not used metaphorically. According to Black, while using a metaphor "we have two thoughts of

different things active together and supported by a single word or phrase, whose meaning is a resultant of their interaction" (Black, 1962, p. 38). Black introduced the notion of "system of associated commonplaces", areas in which all those qualities that one thinks of about focus and frame come together to give the focus a new meaning. For example, consider the following metaphor: "Man is a wolf." According to Black's description, "Man is" is the frame of the metaphor, and "a wolf" is the focus of the metaphor. This interactionist schema Man-as-Wolf is not just a simple combination, it brings components conceptually closer together (Figure 5).

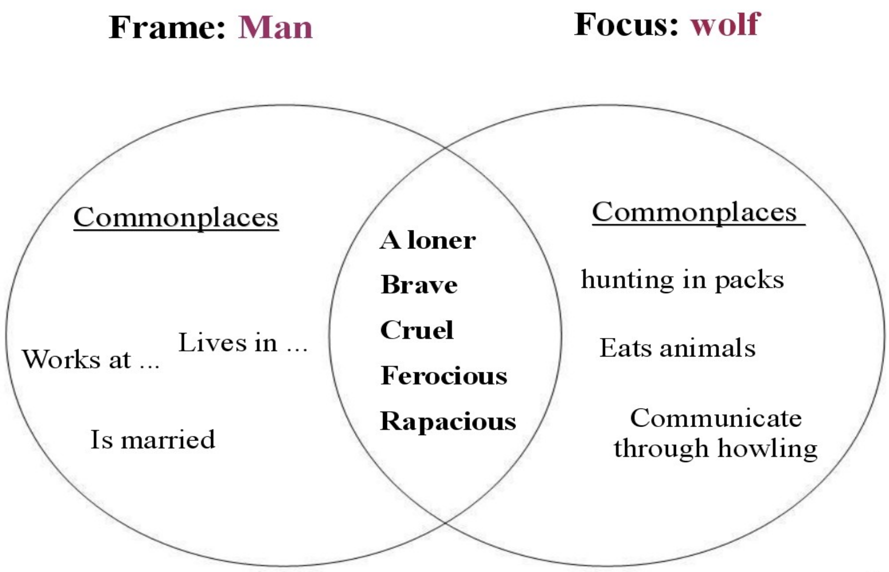

Figure 5: Interactionist view of metaphor "Man is a wolf"

Tourangeau and Sternberg (1982) refer to the "tenor-vehicle" designation and present the vehicle of a metaphor as a template for seeing the tenor in a new way. "This reorganisation of the tenor is necessary, because the characteristics or features of the vehicle cannot be applied directly to the tenor; the features they 'share' are often only shared metaphorically" (ibid, pp. 212-213).

3.2.4) The Selection Restrictions Violations View

The selection restriction violation view known also as "the semantic deviance view" (Fass, 1991) or "the anomaly view" (Tourangeau & Sternberg, 1982) defines metaphor as a violation of specific rules within a given context. One of the ideas behind the anomaly view is the dissimilarity of the semantic features of topics (subjects) and vehicles (predicates). Usually, "selection restrictions are said to be violated when predicates do not fall into the exclusive category ranges determined by their subjects. The result is that such sentences are deemed anomalous or, at best, deviant" (Marschark et al., 1983, p. 18). Example of the violation rules can be found in the metaphor *"The car drinks gasoline"*.

3.3) Metaphor in social science

From the social science point of view, systematic metaphor analysis can be used for reconstruction models of thought, language, and action. For this purpose the qualitative research approach seems to be an effective method that "allows a systematic reflection of the metaphors in which, and through which, we perceive, speak, think, and act" (Schmitt, 2005, p. 368).

Applying a qualitative approach, researchers usually search for and collect metaphors in a wide range of materials containing references to the topic being investigated (encyclopaedias, journals, specialist books written for the general public, etc.). Schmitt (2005) identifies three stages of metaphor use: individual, sub-cultural and cultural use of metaphor.

In the socio-scientific research of the OPAALS project community as well as Digital Ecosystem, metaphor analysis can be used for different purposes:

(1) Description of the results of qualitative research

Qualitative research in general "yields a multitude of heterogeneous pieces of information, which contain complex meaningful structures" (Schmitt, 2005, p. 360). Metaphors are seen as successful filters that reduce this complexity to clearly structured patterns.

(2) Description of the qualitative research process

The research process itself contains complex structures. In this case, metaphors can provide some "orientation for the researchers in their endeavour and in its presentation" (p. 361).

(3) The search for specified metaphors in the material

Some researchers "attempt to orient themselves along metaphors considered central by a specific philosophy" (p. 362).

(4) Self-reflection process of researchers – or – Metaphors we research by

The fact that metaphors are used in unconscious way can be applied not only to the subject of research. The researchers themselves use metaphors in their work, e.g. in the presentation of results, description and analytical representation of specific processes. This viewpoint of metaphor is closely related to the *action research* (see Deliverable 6.1) that introduces the self-reflective aspects to the research processes.

(5) Reconstruction of research participants' metaphorical points of view and of cultural phenomena

In our work we are focusing on the metaphors as mechanisms for knowledge representation and sharing as well as steering mechanisms of interaction and communication processes. We examine metaphors of the OPAALS community through the lens of the community's language development (common conceptual surface).

3.4) Computational theories of metaphor

Our language consists of thousands of concrete and abstract concepts. A lot of them are used in metaphorical sense. The research of conceptual metaphors in Digital Ecosystems is concerned with the creation of a strong and stable structure for a complex knowledge and language system. According to Kövecses, a majority of metaphors captures "three interrelated features of complex systems – their creation, their structure, and the stability of their structure" (Kövecses, 2002, p. 109).

In Digital Ecosystems metaphors can be used in dual way. On the one hand, metaphors are representative as semantic objects for humans. The mapping between source and target domains allows a semantic transformation, where the meaning of a word, phrase or sentence changes. On the other hand, metaphors in digital environments are structured entities where the semantic part is not recognisable or understandable by computers. However, the computer-assisted extraction of different natural language features is an attractive approach due its efficiency and ability to deal with the large amounts of textual information. To bridge the gap between semantic meaning and structural representation of metaphors is one of the difficult tasks in natural language processing and computational linguistics. One of the goals is to recognise and extract metaphorical expressions as a part of information retrieval and knowledge representation in Digital Ecosystems. Different computational theories and approaches dealing with metaphor extraction, recognition, learning, etc. are presented in this chapter.

3.4.1) Computational theory of predication

The Computational Theory of Predication was introduced by Kintsch (2000) and contains two basic components:

- a model of human knowledge structure provided by Latent Semantic Analysis (LSA), and
- a model of text comprehension, the construction-integration (CI) model

The LSA approach refers to the mapping of text onto high-dimensional vectors and is based on the assumption that the similarity of vectors indicates similarity of meaning/sense. It "provides an automatic method for comparing units of textual information to each other in order to determine their semantic relatedness" (Martin, 2004, p. 1). This approach is used as a method for extracting and representing the meaning of words based on the statistical analysis of text corpora. Originally, LSA was designed to improve the quality of information retrieval methods and techniques by focusing on the "semantic" content of words and not on the direct word matching in a query (Deerwester et al., 1990). Kintsch (2000) showed that LSA is not only a tool for automatically computing semantic similarities between words or texts but also a powerful model of word acquisition and metaphor analysis.

The Construction-Integration Model is a "theory of discourse comprehension that is being used as

the foundation for a new Integrated Intelligent Architecture" (Wharton & Kintsch, 1991, p. 169). This model holds different combinations of symbolic features and connectionist techniques and can be divided in two parts: knowledge construction and knowledge integration.

There are several limitations of this approach based on the computational theory of predication (Kintsch, 2000):

- LSA successfully captures only the semantic distance among words. LSA is a model of meaning, not the meaning itself.
- LSA is an incomplete model. "It models only those aspects of meaning that are coded verbally; human meaning is derived from perception and action as well as words" (Kintsch, 2000, p. 6).
- Centroid rule failures in metaphorical predication. According to Kintsch (2000), the centroid rule that "says that the vector representing a set of words is the centroid of the individual word vectors" is inadequate in the case of metaphorical predication.

Despite the aforementioned disadvantages, LSA seems to be successful in computing quantitative measures of the relatedness between terms (Kintsch, 2000).

3.4.2) Metaphor interpretation, denotation, and acquisition system approach

This approach was introduced by Martin (1992) who claims that "the interpretation of conventional metaphoric language should proceed through the direct application of specific knowledge about the metaphors in the language" (Martin, 1992, p. 233). MIDAS (Metaphor Interpretation, Denotation, and Acquisition System) consists of a set of computer programmes that perform different tasks. It contains a complex system of hierarchically organised knowledge about metaphors. According to the author, this system can be used to "represent knowledge about conventional metaphors, interpret metaphoric language by applying this knowledge, and dynamically learn new metaphors as they are encountered during normal processing" (Martin, 1992, p. 233). Basically, MIDAS can be divided into three main parts (Martin, 1992):

Representation

The author identifies different components of metaphors: a source component, a target component, and a set of conventional associations from the source to target. "The target consists of the concepts to which the words are actually referring. The source refers to the concepts in terms in which the intended target concepts are being viewed" (Martin, 1992, p. 238). According to Martin, MIDAS uses directional relations which means that they define the source and target domains as well as the concepts belonging to one of these domains.

Interpretation

At this level, the system checks the input constraints against all the possible interpretations that can be conventionally associated with the input. The interpretation process includes the mapping process from the concepts in the source domain to the appropriate concepts in the target domain.

Learning

This level reflects the process of the dynamic acquisition of new knowledge about metaphors. At this stage an analogical reasoning approach is used. The approach of understanding unknown metaphors is called the Metaphor Extension Approach. According to Martin (1992), through this approach an unknown metaphor "can best be understood by extending an existing well-understood metaphor or combining several known metaphors in a systematic fashion" (p. 243). Therefore, the processes of understanding and learning depend on systematic knowledge about existing metaphors.

3.4.3) Conceptual Dependency Theory

Conceptual Dependency is a theory of representation developed by Roger Schank, and designed to represent the meanings of sentences based on a small set of primitives (Way, 1991). According to Schank (1972), conceptual dependency diagrams are independent of the language in which the sentence they represent was phrased. Schank (1972) proposes six conceptual categories:

- real world objects (PP: Picture Producers),
- real world actions (ACT),
- modifiers of actions (AA: Action Aiders),
- modifiers of PP's (PA: Picture Aiders),
- times (T),
- and locations (LOC);

Additionally, he introduces sixteen conceptual syntax rules, which specify how these categories can relate to each other. Conceptual Dependency also uses four conceptual cases: OBJECTIVE, RECIPIENT, DIRECTIVE, AND INSTRUMENTAL (Way, 1991, p. 102).

However, there are some problems with conceptual dependency theory. The decomposition of all knowledge to the low-level primitives leads to the problem that the representation of complex concepts as primitive expressions is very lengthy and computationally inefficient. Furthermore, there is a possibility that some words, concepts, or especially metaphors cannot be expressed by a small set of primitives. Additionally, it is not always possible to have a unique representation of expressions (Way, 1991).

3.4.4) The Dynamic Type Hierarchy theory

The Dynamic Type Hierarchy theory (DTHT) is a computational theory of metaphor based on Black's interaction theory and the notion of fuzzy sets. This approach is based on the hierarchical knowledge representation that is context neutral. The hierarchy can be seen as a simple order of "types and supertypes that reflects some kind of ontological ordering" (Way, 1991, p. 126).

An important role in differentiation between metaphorical and literal expressions play different masks that reflect different views of the semantic hierarchy. "Whether a statement is literal, metaphoric or figurative depends upon what mask comes into play and what connections in the hierarchy are hidden or exposed by it" (Way, 1991, p. 126). In this approach, there is no usual distinction between metaphoric and literal language because they have the same status. The difference lies in different aspects of the hierarchy. Metaphor can be identified through new semantic linkages. An example of metaphor identification is presented in Figure 6.

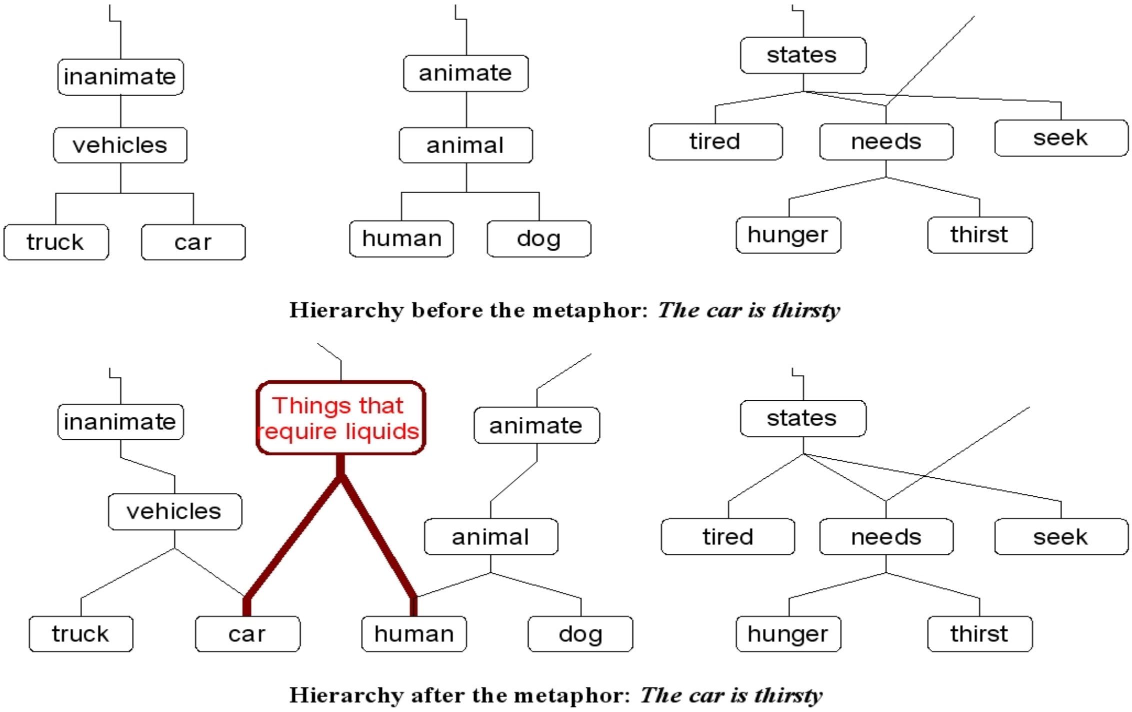

Figure 6: DTHT approach of metaphor analysis (Way, 1991)

As we can see in Figure 6, metaphor "tends to add new connections to the hierarchy, which often results in blurring the distinction between categories" (Way, 1991, p. 128). According to the DTHT approach, the purpose of metaphor is not precision, but creating new levels and connections between concepts.

3.4.5) Related Work

This sub-chapter provides an overview of related work by introducing four methods of computational corpus and metaphor analysis. Generally speaking, computational approaches of language analysis provide powerful techniques that can be applied to metaphor analysis, extraction, learning, and representation in Digital Ecosystems. It is worth to mention that a lot of systems dealing with different aspects of metaphor analysis are predicated on the linguistic preprocessing and/or analysis. The combination of computational and linguistic techniques provide successful

features of natural language processing in general, and the analysis of metaphorical expressions, in particular.

In chapter 5, we will then present the first draft of a Metaphorological Tool Kit that aims to collect, extract, analyse, and represent community's specific metaphors; in our case metaphors of the OPAALS NoE.

CorMet

The CorMet system was introduced by Zachary J. Mason. It is a corpus-based system that deals with discovering metaphorical mappings between different concepts. The focus lies on the verbs because similar tendency of verb selection in the source and target domains has been observed, meaning that "the verbs that select for a concept in the source domain tend to select for its metaphorical equivalent in the target domain" (Mason, 2004, p. 23). The CorMet system does not only detect metaphorical expressions; it also learns the selectional preferences of the characteristic verbs of a particular domain.

The CorMet system starts with detection of the domain-specific verbs. For this purpose, CorMet analyses a large corpus of domain-relevant documents, decomposes them into a bag-of-words representation, and a applies stemming approach by using the Porter (1980) stemmer. Then the ratio of occurrences of each word stem to the total number of stems in the domain corpus is calculated and the frequencies are compared to their frequencies in general English (Mason, 2004).

In a second step, CorMet learns the domain-specific verbs by using a selectional-preference-learning algorithm and finally applies the clustering algorithm to the results. The selectional-preference learning algorithm "takes a set of words observed in a case slot (e.g., the subject of pour or the indirect object of give) and finds the WordNet nodes that best characterize the selectional preferences of that slot" (Mason, 2004, p. 26). Two measures are important at this stage: selectional-preference strength (1) that is based on the concept relative entropy (or Kullback-Leibler divergence) and selectional association (2).

$$S_R(p) = D(P(c|p)||P(c)) = \Sigma_c P(c|p) log \frac{P(c|p)}{P(c)} \tag{1}$$

where $S_R(p)$ is defined as "the relative entropy of the posterior probability $P(c|p)$ and the prior probability $P(c)$ (where $P(c)$ is the a priori probability of the appearance of a WordNet node c , or one of its descendants, and $P(c|p)$ is the probability of that node or one of its descendants appearing in a case slot p) " (Mason, p. 26).

$$\Lambda_R(p, c) = \frac{1}{S_R(p)} P(c|p) log \frac{P(c|p)}{P(c)} \tag{2}$$

$\Lambda_R(p, c)$ is computed based on the distribution over word classes.

CorMet provides a possibility of semi-automatic extraction of metaphoric mappings between concepts. However, this system is not able to recognise metaphoric language in general.

*Met**

This method of metaphor recognition based on selection restriction was introduced by Fass (1991). “Metonymy, metaphor, literalness, and anomaly are recognized by evaluating preferences, which produces four kinds of basic "preference-based" relationship or semantic relation: literal, metonymic, metaphorical, and anomalous” (ibid, p. 59).

Fass (1991) points out that the process of metaphor recognition in the met* method refers to the following aspects:

- a contextual constraint violation (see the selection violation view);
- a set of "correspondences"/ system of commonplaces (see the interaction view);
- a relevant analogy (see the comparison and interaction views);
- analogies that fall into patterns not unlike conceptual metaphors (see the conventional view).

Metaphors and verbal targets

Gedigian, Bryant, Narayanan, and Ciric use a maximum entropy (ME) classifier to identify metaphors. Their focus lies on the verbal target. The authors identify two factors of determining the metaphorical character of verbal targets in a certain utterance (Gedigian et al., 2006):

1. the bias of the verb;
2. the arguments of the verbal target in that utterance.

To determine the metaphoric character of the arguments, the system accesses the PropBank annotations and uses the WordNet lexical database.

Comparison and categorisation models

Utsumi (2006) focuses on the computational solution to the problem of metaphor comprehension. He suggests to use comparison and categorization algorithms “based on word vectors in a multidimensional semantic space constructed by latent semantic analysis” (Utsumi, 2006, p. 2281). In this model, a vector representation $v(s)$ of a sequence s (can be a phrase, sentence, or paragraph) consists of words $w_1, \ldots, w_n$ and can be defined as a function $f(v(w_1), \ldots, v(w_n))$ (Utsumi, 2006). Based on this formalisation metaphor comprehension can be modelled as follows:

$$v(M) = f(v(w_T), v(w_V)) \quad (3)$$

The vector $v(M)$ represents “the meaning of metaphor M with the topic w_T and the vehicle w_V “ (Utsumi, 2006, p. 2282). To improve the target behaviour, the author suggests a comparison and categorization algorithm. One of the important factors in both algorithms is the neighbourhood (n neighbours of a word x). The neighbourhood is based on similarity measure between words x and y , which is computed as the cosine $\cos(v(x), v(y))$ of the angle formed by two word vectors. $N_n(x)$ denotes a set of n neighbours of x .

Comparison algorithm (Utsumi, 2006, p. 2282):

1) Compute a set of k words (i.e., alignments between the topic w_T and the vehicle w_V) as $N_i(w_T) \cap N_i(w_V)$ by incrementing i i by 1 until $|N_i(w_T) \cap N_i(w_V)|$ reaches k.
2) Compute a metaphor vector $v(M)$ as the centroid of $v(w_T)$ and k vectors computed at Step 1.

Categorization algorithm (Utsumi, 2006, p. 2282):

1) Compute $N_m(w_V)$, i.e., m neighbours of the vehicle w_V .
2) Select k words with the highest similarity to the topic w_T from $N_m(w_V)$
3) Compute a vector $v(M)$ as the centroid of $v(w_T)$, $v(w_V)$ and k vectors selected at Step 2. This algorithm is identical to Kintsch’s (2000) predication algorithm.

4) Metaphorological Tool Kit: framework and structure

4.1) Overview

Metaphors introduce the way of describing the connection that exists between two groups of ideas in people's minds. They play an important role in structuring our background conceptual systems as well as our knowledge about the world. Conceptual metaphors enable us to quantify, visualise and generalise abstract concepts because they make use of relationships within source domains that we know well from our concrete experience.

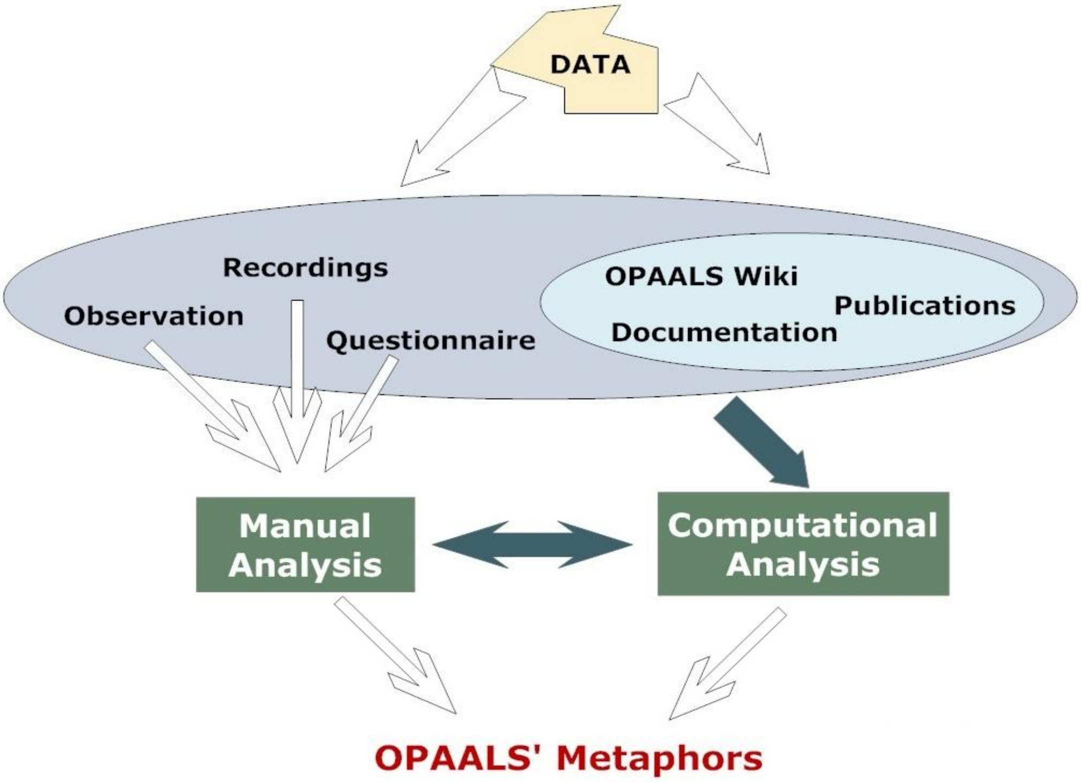

Figure 7: Metaphorological Tool Kit Overview

Based on the theoretical and practical approaches that were introduced in the foregoing chapters, this chapter describes our work regarding the development of a Metaphorological Tool Kit (MTK). The MTK will be refined and tested throughout Phase II. Its prospective users are those that have a sound background in both linguistics and computer science. However, the results of the work conducted by means of the MTK will find many users within the OPAALS NoE, for example researchers in tasks T10.5 "Visualisation of the OKS", T10.11 "Social networks in cross-domain collaboration, T10.12 "Knowledge models and representations", T10.14 "Meta-data tagging tools and environments", T10.15 "Distributed semantically searchable repository", and T6.6

"Evolutionary framework for language". To be more precise, the results produced by the MTK provide two important data sets that can be used for further work and interpretation by other researchers: Firstly, the MTK provides a structured language set which can be used in order to build a common lexicon for DE and in other data repository projects or generally in tasks that deal with formal language environments with natural language input. Secondly, the MTK results provide an important data set for social scientists in terms of mirroring communication patters within the OPAALS community and providing at the same time a structured approach for this.

The MTK consists of two blocks, manual search/analysis and automatic recognition. Manual search introduces the manual recognition and control of metaphor identification and analysis (i.e. carried out by a human being). Automatic recognition refers to the computational part, more precisely to the set of algorithms and tools that will be developed/integrated within the MTK framework in Phase II. The latter part reflects the combination of different techniques from linguistics and computer science that deal with text analysis. Figure 7 depicts a general overview of the MTK which is elaborated in the following sub-chapters by means of introducing and discussing the most relevant tasks and aspects of the MTK: data (e.g. corpus based approaches), term extraction, conceptual surface of metaphors and analysis interface, metaphor extraction including the preprocessing and discovery.

4.2) Data

4.2.1) Qualitative analysis

The process of data collection has already been started in Task 6.1. A questionnaire introduced in Phase I contains a "Language section" where participants are asked to provide associations related to different concepts existing within the OPAALS community such as Digital Ecosystem, Knowledge, Network, Community, Language. The qualitative analysis of the collected data will be used in Phase II as a part of the MTK. Additionally, recordings and observation notes done in the context of the qualitative analysis of epistemic cultures existing within the OPAALS community (e.g. during workshops, meetings, etc.) provide important information for further processes of metaphor identification, collection, and analysis.

4.2.2) Corpus-based approach

In our work we use a corpus-based approach for extracting statistical and linguistic information of terms, concepts, and metaphors relevant in the OPAALS community.

A corpus is "a collection of text or speech material that has been brought together according to a certain set of predetermined criteria" (Megerdoomian, 2003, p. 215).

Biber et. al. (2004) introduce four general characteristics of corpus-based analysis:

- empirical character (analysis of the actual patterns of use in natural texts);
- a corpus, the large collection of natural texts, builds the basis for analysis;
- computer-based approach of analysis by means of using automatic and interactive techniques;
- the analysis of the corpus depends on the quantitative and qualitative techniques.

There are several advantages of the corpus-based approach in terminological studies (Popp, 2001):

a) Using the electronic corpora enables the automatic identification and extraction of the key terms relevant to the subject field.

b) The use of electronic text corpus provides the possibility to determine "linguistic patterns, such as different kinds of syntactic, semantic and pragmatic details".

c) One is able to analyse the context of a particular term or term phrase, for example using collocation, concordance and other linguistic techniques.

d) A corpus-based approach allows the retrieval of conceptual characteristics of terms.

e) Corpus analysis allows to share information of different subject fields and enables them to reuse data.

Megerdoomian (2003) identifies two important criteria for a corpus: representativeness and balance.

Representativeness refers to the construction of the corpus that is "as representative as possible of the domain under study, by including samples from a broad range of material" (Megerdoomian, 2003, p. 216). A representative corpus needs to provide a comprehensive picture of the language population. The text corpus (collection of text documents) we use in our study contains the web sites from the OPAALS wiki pages. In order to make this corpus as representative as possible, we applied a certain set of filtering criteria. Since we are interested in the conceptual and metaphorical knowledge representation of the OPAALS community, the pages containing reports, discussions of research topics, descriptions and progress of the research tasks, etc. were included in the corpus. Contrariwise, web sites containing system-relevant information (for example, help pages), meeting information (place, time, directions, accommodation), and so on have been filtered out and were not included in the corpus.

A *balanced* corpus refers to the collection that "attempts to cover as many textual styles as possible by trying to include samples from various genres" (Megerdoomian, 2003, p. 216). Our text corpus contains different styles of information and knowledge representation, from discussions, agenda and notes to the structured reports and articles related to all three research domains existing within the OPAALS community: social science, computer science and natural science. Working with the electronic corpus is in our case an excellent way to acquire knowledge about the project's subject fields and its specialised languages/metaphors that is influenced by the interdisciplinary and

multilingual character of the community.

However, the dynamic character of the OPAALS wiki-platform requires constant and continual adjustment, expansion through introducing different genres (deliverables, milestones, publications, etc.), and filtering.

4.3) Term extraction and their metaphorical representation

The first step towards the collection of OPAALS metaphors was carried out by applying the Term Extractor tool (Sclano & Velardi, 2007) to the text corpus described in section 5.2.2).

Term extraction is a process of "scanning" the text corpus (text or group of texts) in order to identify relevant terms, concepts and other relevant information (for example, the context where a particular term occurs). It is an important step in our research towards the systematic examination of a commonly understood and used vocabulary, its development and changes occurring through the dynamic communication and community development processes. Term extraction process starts with the term identification that involves the recognition and selection of specific words or word combinations. After candidate terms have been identified in a text, several calculations and filtering processes need to be applied in order to collect terms and concepts that are indeed part of the subject field under study.

As many terminology extraction systems, the TermExtractor tool identifies the relevant terms from a given corpus based on two steps (Sclano & Velardi, 2007):

a) Extraction of terminological structures: A linguistic processor is applied that parses the text corpora and extracts typical terminological structures, for example compounds (digital ecosystem), adjective-noun sequences (local network) and noun preposition noun sequences (network of excellence).

b) Filtering of terminological candidates is based on five values: Domain Relevance, Domain Consensus, Term Cohesion (or Lexical Cohesion), Artificial Frequency and Term Weight (Sclano & Velardi, 2007).

Domain Relevance value is calculated based on two points: the frequency of terms and the measurement of a term with respect to the target domain via comparative analysis across different domains. D_i is the target domain (domain of interest) that is represented by a single document or a set of relevant documents. Comparative domains $D_{1,} \ldots, D_n$ are represented as sets of documents (or terminologies) in different domains, for example, medicine, politics, etc. The more precise definition of the Domain Relevance is given as follows:

$$DR_{D_i}(t) = \frac{P(t|D_i)}{\max\limits_{1 \le j \le n}(P(t|D_j))} \tag{4}$$

The conditional probabilities $P(t|D_i)$ are estimated based on the frequency of term t in the domain D_i .

Domain Consensus, a second relevance indicator, is then assigned to candidate terms. It measures the distributed use of a term in a domain. More precisely, the domain consensus is expressed as follows:

$$DR_{D_i}(t) = -\sum_{d_k \in D_i} norm_{freq}(t, d_k) \log(norm_{freq}(t, d_k)) \quad (5)$$

where d_k is a document in D_i and $norm_{freq}$ is a normalised term frequency. This value of the Domain Consensus is then normalised for each term in the [0,1] interval. In case of an even probability distribution across the documents of the domain, the consensus becomes high (Sclano & Velardi, 2007).

Term Cohesion (or Lexical Cohesion)

The computation of the Lexical Cohesion aims to measure the association of multi-word terms based on co-occurrence frequencies. The mathematical representation of Term Cohesion is expressed as follows:

$$LC_{D_i}(t) = \frac{(n \cdot freq(t, D_i) \cdot \log(freq(t, D_i)))}{(\sum_{w_j} freq(w_j, D_j))} \quad (6)$$

Here, w_j are the words composing the term t . If words composing a particular term occur more frequently within the term rather than alone in texts, lexical cohesion becomes high.

Artificial Frequency

TermExtractor is able to analyse in any type of document different text layouts: bold, italic, title, underlined, capitalized, coloured, small caps, etc. Depending on the layout, it assigns different values to the term. For example, if a term occurs once in bold, the TermExtractor assigns to it a higher value (e.g., an artificial frequency of 5) rather than the low raw frequency (e.g., 1) (Sclano & Velardi, 2007).

Term Weight

Term Weight is a linear combination of Domain Relevance, Domain Consensus, Lexical Cohesion and Artificial Frequency:

$$TW_{D_i}(t) = \alpha * DR + \beta * DC + \gamma * LC \quad (7)$$

This tool successfully identified a lot of terms expressed within the project community. However, not all terms are used metaphorically and not all terms are relevant. Therefore, this process is only the first step towards a MTK development. At this stage, manual control and analysis are required. Thereinafter, the conceptual metaphors and their relationships will be submitted using the MTK interface available at 141.51.104.117/tematres/.

4.4) Conceptual surface of metaphors

An online interface has been created for the purposes of a MTK development. It is based on the TemaTres tool and will be used for submission, search, and representation of metaphors (see Figure 8). Especially regarding the role of ontological and conceptual representation of language in Digital Ecosystems, the hierarchical representation of concepts and terms and their metaphorical characteristics are one of the most important issues.

The existing interface provides possibilities to collect, edit, and search terms and concepts; build different relationships between terms and concepts, build hierarchical representation; export in XML format, duplicates term control, and so on.

In Phase I, the term extraction process has been done to identify the most relevant concepts and terms, analyse their metaphorical representation and use, and to start the process of their submission. In Phase II, some of the extraction techniques will be integrated into the MTK interface. This will require the combination of linguistic and data/text mining techniques. Some of them will be shortly described in the following sections. The final interface design including different algorithms and tools will be provided in Phase II.

The draft structure including possible analytical steps of metaphorical expressions is presented in Figure 9.

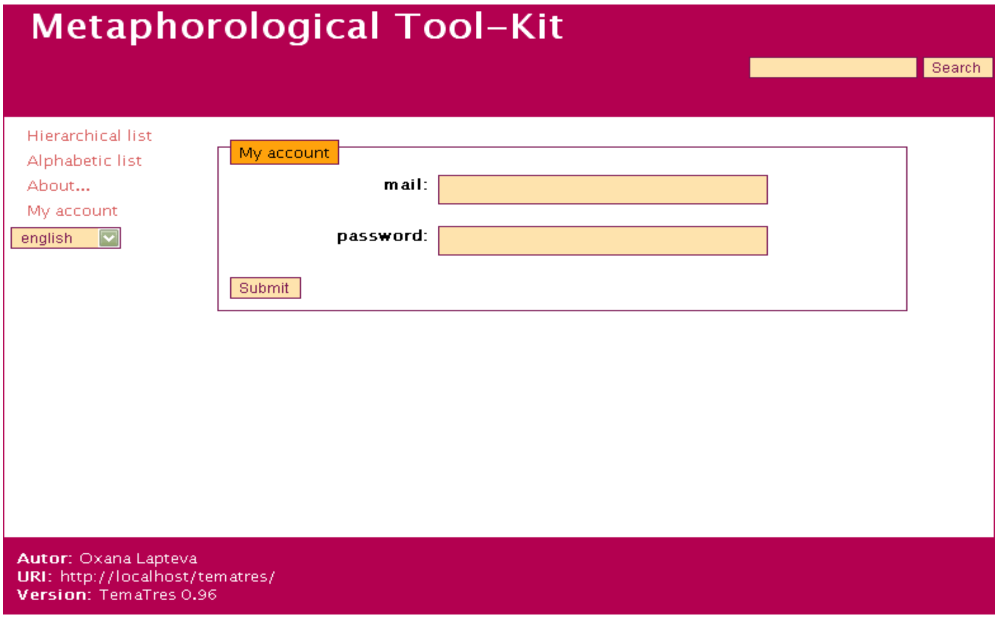

Figure 8: TemaTres interface for the MTK

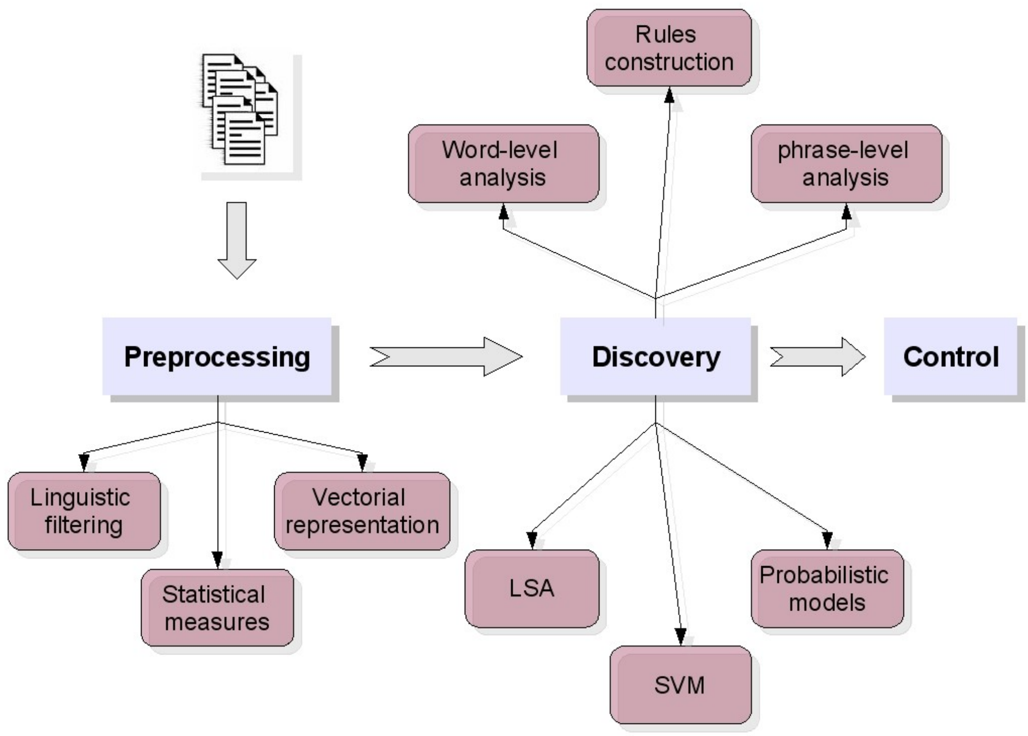

Figure 9: Computational analysis of metaphors, draft

4.5) Metaphor extraction

4.5.1) Preprocessing

Preprocessing different filtering methods and algorithms can be useful, for example, for the removal of stop-words, for stemming, part-of speech tagging, parsing, etc. In this section we will briefly present some of these methods.

Stop-words removal

This is a cleaning process where the stop-words that do not carry an information but have a high frequency of occurrence are removed, for example “and”, “the”, “a”. These words have mainly functional role and usually the filtering process is used to help the further methods of analysis to perform better.

Table 8 shows the results of this filtering process:

Original text	*After the stop-words removal*
The two overarching aims **of** OPAALS are **to** build **an** interdisciplinary research community **in the** emerging area **of** Digital Ecosystems (DE), **and to** develop **an** integrated theoretical foundation **for** Digital Ecosystems research spanning three widely different disciplinary domains: social science, computer science, **and** natural science.	two overarching aims OPAALS are build interdisciplinary research community emerging area Digital Ecosystems (DE), develop integrated theoretical foundation Digital Ecosystems research spanning three widely different disciplinary domains: social science, computer science, natural science.

Table 8: Stop-words filtering

Stemming

In most cases, morphological variants of words have similar semantic interpretations and can be considered as equivalent in a variety of applications. Words with the same meaning appear in various morphological forms. To capture their similarity they are normalised into a common root-form, the stem. For example, the words *laughing*, *laugh*, *laughs* and *laughed* are all stemmed to ***laugh*** (Jurafsky & Martin, 2000). A number of so-called stemming algorithms, or stemmers, have been developed that attempt to reduce a word to its stem or root form. Most widely used is the Porter stemmer that is based on a series of simple cascaded rewrite rules. The following example shows the results after applying Porter stemmer:

<u>Example</u>

Input text:

> The two overarching aims of OPAALS are to build an interdisciplinary research community in the emerging area of Digital Ecosystems (DE), and to develop an integrated theoretical foundation for Digital Ecosystems research spanning three widely different disciplinary domains: social science, computer science, and natural science.

Output ("stop" words and punctuation are automatically removed):

Original Word	*Stemmed Word*
overarching	overarch
aims	aim
opaals	opaal

build	build
interdisciplinary	interdisciplinari
research	research
community	commun
emerging	emerg
area	area
digital	digit
ecosystems	ecosystem
de	de
develop	develop
integrated	integr
theoretical	theoret
foundation	foundat
digital	digit
ecosystems	ecosystem
research	research
spanning	span
widely	wide
different	differ
disciplinary	disciplinari
domains	domain
social	social

science	scienc
computer	comput
science	scienc
natural	natur
science	scienc

Table 9: Stemming

Part-of-Speech tagging and parsing

Part-of-speech (POS) tagging provides the annotation of words. It is the process of assigning a part-of-speech (noun, verb, adjective, etc.) to each word in a text corpus. POS tags assign categories to words based on the role they play in the sentence in which they appear and provide information about the semantic content of a word. The input to a tagging algorithm is a string of words and a specified tagset (for example, Brown/Penn tagsets). The output is a tag for each word.

POS taggers are often integrated in a **parser**. Parsing can be defined as a combination of recognising an input string and assigning some structure to it. The following example shows results of applying POS tagging and parsing (Stanford NLP parser, http://nlp.stanford.edu/software/lex-parser.shtml).

Input:

> The two overarching aims of OPAALS are to build an interdisciplinary research community in the emerging area of Digital Ecosystems (DE) , and to develop an integrated theoretical foundation for Digital Ecosystems research spanning three widely different disciplinary domains: social science , computer science , and natural science .

Tagging:

The/DT two/CD overarching/JJ aims/NNS of/IN OPAALS/NNS are/VBP to/TO build/VB an/DT interdisciplinary/JJ research/NN community/NN in/IN the/DT emerging/VBG area/NN of/IN Digital/NNP Ecosystems/NNPS -LRB-/-LRB- DE/NNP -RRB-/-RRB- ,/, and/CC to/TO develop/VB an/DT integrated/VBN theoretical/JJ foundation/NN for/IN Digital/NNP Ecosystems/NNPS research/NN spanning/VBG three/CD widely/RB different/JJ disciplinary/JJ domains/NNS :/: social/JJ science/NN ,/, computer/NN science/NN ,/, and/CC natural/JJ science/NN ./.

Parse:

```
(ROOT
  (S
    (NP
      (NP (DT The) (CD two) (JJ overarching) (NNS aims))
      (PP (IN of)
        (NP (NNS OPAALS))))
    (VP (VBP are)
      (S
        (VP (TO to)
          (VP (VB build)
            (NP (DT an) (JJ interdisciplinary) (NN research) (NN community))
            (PP (IN in)
              (NP
                (NP (DT the) (VBG emerging) (NN area))
                (PP (IN of)
                  (NP (NNP Digital) (NNPS Ecosystems)))
                (PRN (-LRB- -LRB-)
                  (NP (NNP DE))
                  (-RRB- -RRB-))
                (, ,)
                (CC and)
                (S
                  (VP (TO to)
                    (VP (VB develop)
                              (NP (DT an) (VBN integrated) (JJ theoretical) (NN
foundation))
                      (PP (IN for)
                        (NP
                          (NP (NNP Digital) (NNPS Ecosystems) (NN research))
                          (VP (VBG spanning)
                            (NP
                              (NP (CD three)
                                (ADJP (RB widely) (JJ different))
                                (JJ disciplinary) (NNS domains))
                              (: :)
                              (NP
                                (NP (JJ social) (NN science))
```

```
                                        (, ,)
                                        (NP (NN computer) (NN science))
                                        (, ,)
                                        (CC and)
                                        (NP (JJ natural) (NN science))))))))))))))
    (. .)))
```

Through parsing we are able to identify the structure of a sentence, extract verb and noun phrases. This process would provide important linguistic information that will be used in the analysis of metaphors.

Besides the linguistic filtering techniques, the **vectorial representation** of textual information is a processes that is often involved in metaphor recognition. The vector representation most commonly used is the bag of words vector. A bag of words vector v exists in a vector space Sw. This is a space of very high dimensionality, in which w represents the total number of unique words in the vocabulary, and each word i is represented by a unique coordinate vi.

Text as a multidimensional vector is then used in Latent Semantic Analysis (LSA) and Support-Vector Machines (SVM).

Different statistical measures may help to discover different metaphorical patterns.

C-value

The C-value method is a domain-independent method used to automatically extract multi-word keywords and terms from a given document corpus. “It aims to get more accurate terms than those obtained by the pure frequency of occurrence method, especially terms that may appear as nested within longer terms” (Milios et al., 2003, p. 276). It uses statistical characteristics of the candidate string such as frequency of occurrence in the corpus, frequency as part of other longer candidate terms, the number of these longer candidate terms, the length of the candidates.

Furthermore, this method enhances the common statistical measure (frequency of occurrence) and makes the measure sensitive to a particular type of multi-word terms (the nested terms) (Frantzi et al., 2000). It applies more weight to nested terms. C-value is formally represented as follows:

$$C-value(a)=\left\{\log_2(|a|\cdot f(a))\right\} \quad \text{if } a \text{ is not nested}$$

$$C-value(a)=\left\{\log_2|a|(f(a)-\frac{1}{(P(Ta))}\sum_{b\in Ta} f(b))\right\} \quad \text{, otherwise.} \tag{8}$$

a = candidate term

b = longer candidate terms

$|a|$ = length of the candidate term (number of words)

$f(a)$ = frequency of occurrence of a in the corpus

Ta = set of extracted candidate terms that contain *a*

$P(Ta)$ = number of candidate terms in *Ta*

$f(b)$ = frequency of occurrence of longer candidate term *b* in the corpus.

Frantzi et al. (2000) provide an example of applying C-value measurement to the corpus consisting of medical records. The string *basal cell carcinoma* is analysed after linguistic preprocessing and filtering of the corpus. The process begins with the longest string, in this example *adenoic cystic basal cell carcinoma.* Therefore the first case of the equation (8) is applied, where:

a is *adenoic basal cystic basal cell carcinoma*,

$f(a)$ is 5.

The C-value is:

$$C-value(adenoic\ cystic\ basal\ cell\ carcinoma)=\log_2 5\cdot 5=11.6096$$

In this example, the longest string (*adenoic basal cystic basal cell carcinoma*) contains the nested term *cystic basal cell carcinoma.* Therefore, the second case of the equation (8) occurs, where:

a is *cystic basal cell carcinoma*,

$f(a)$ is 11,

Ta is *adenoic cystic basal cell carcinoma*,

$P(Ta)$ is 1,

$\sum_{b\in Ta} f(b)$ is 5.

The C-value is:

$$C-value(cystic\ basal\ cell\ carcinoma)=\log_2 4\cdot(11-\frac{5}{1})=12$$

An extension of C-value is called NC-value that provides (Frantzi et al., 2000):

1) a method for the extraction of term context words (words that tend to appear with terms);
2) the incorporation of information from term context words to the extraction of terms.

Some experiments show that the C- value/NC-value method performs well on a variety of special

text corpora.

TFIDF

TFIDF (Zhang et al., 2005) is a standard keyword identification method in information retrieval tasks. It asserts preference to words that have high frequency of occurrence in a single document but rarely appear in the whole document collection. The TFIDF measure is often used in the vector space model together with cosine similarity to determine the similarity between two documents. The TFIDF is calculated as follows:

$$w_{ij} = tf_{ij} \cdot \log_2 \left(\frac{N}{n} \right) \tag{9}$$

where

w_{ij} = weight of term T_j in document D_i

tf_{ij} = frequency of term T_j in document D_i

N = number of documents in collection

n = number of documents where term T_j occurs at least once

The described linguistic and statistical methods can be combined to serve the ongoing discovering process of metaphorical expressions. The final configuration of this step will be decided in Phase II.

4.5.2) Discovery

The last item to be introduced and discussed for our MTK is discovery. Discovering conceptual metaphors is a complex task that may involve different linguistic and computational approaches:

Rules construction represents the set of patterns for searching and identifying metaphors in a large text corpus. This process involves the definition of the metaphorical structure, for example metaphoric usages involving nouns. Krishnakumaran & Zhu (2007) categorise three types of metaphoric relationships: the verb-noun relationship in sentences, based on the verb, and the adjective-noun relationship. The automatic identification of different relations depends on the quality of parsing and part-of-speech tagging in the preprocessing stage. Additionally, the differentiation between metaphorical and literal use of words and phrases requires some knowledge referencing system. Krishnakumaran & Zhu (2007) suggest to look up the hyponymic relations existing in WordNet, a large lexical database of English created at the Cognitive Science Laboratory of Princeton University.

N-grams present the way of generating phrases and multi-word terms. N-gram is a sequence of n consecutive words, for example *digital ecosystem*. The parsing results can help us to identify different n-grams, for example we can extract bi- and tri-grams by looking at the noun phrases (NP): *(NP (NNP Digital) (NNPS Ecosystems) (NN research))* or *(NP (JJ social) (NN science))*.

***Collocations**,* defined as arbitrary and recurrent word combinations (Benson, 1990) or institutionalised phrases (Sag et al., 2002), represent a subclass of multi-word expressions that are prevalent in language and constitute a key problem for natural language processing (NLP) and text mining. Generally, a process of collocation extraction comprises two main stages (Sag et al., 2002):

- the *candidate selection*, in which the word expressions that could represent a collocation are extracted from text according to specifically defined collocation patterns (or configurations);
- the *relevance test*, which assigns to each extracted candidate a weight indicating its likelihood to constitute a collocation. The result of a collocation extraction system is a ranked list (usually called *significance list*) of potential collocations, with the most probable collocation on the top.

LSA: The basic idea behind LSA is that semantically similar words are likely to occur in semantically similar paragraphs and semantically similar paragraphs are likely to include semantically similar words. LSA is a "method for automatically constructing a high-dimensional semantic space from the analysis of a large amount of written text" (Kintsch, 2000, p. 4). This approach is based on the occurrence patterns over contexts and does not "represent meaning in terms of co-occurrence frequencies, but as vectors in a semantic space [...] dimensions" (ibid).

SVM: Support Vector Machines are usually used for document classification problems. A SVM classifier represents a "hyperplane in the feature space separating the points that present the positive instances of the category from the points that represent the negative instances" (Feldman & Sanger, 2007, p. 76). This idea can be applied to metaphor identification, namely differentiation between metaphorical and non-metaphorical use of terms and phrases based on the relatively small subset of the training instances (support vectors).

***Probabilistic calculations*:** The aforementioned methods related to discovery of metaphors and their patterns can be expanded through integration of probabilistic calculations, for example the calculation of conditional probability for a noun and a verb.

5) Conclusion

This deliverable aimed to provide a first theoretical background of metaphors as well as descriptions of different methods and approaches that might be useful in the processes of metaphor identification, extraction, representation, and learning. The comprehensive analysis of theories helped to examine metaphors from different points of view and develop first ideas for the Metaphorlogical Tool Kit (MTK). Additionally, the comprehensiveness aimed at providing an introduction to the complexity of metaphor (and, as a representative means, the complexity of natural language in general) for other partners who will also use the results from task T6.2 and do not come from an either linguistic or social science background.

We are aware of the importance of metaphor analysis from the social point of view: metaphors represent an important means to construct in a highly creative way our reality. We use metaphors in order to describe either something new or to add new values, connotations, visions to already existing words/concepts. As we have already stated above, metaphor usage is very frequent in our everyday lives, but also in our professional lives. Hence, we can assume that the OPAALS community of researchers consists of a rich set of metaphors, imported from the different domains that are present in the NoE. Naturally, this surplus or diversity of metaphor (or also called razzmatazz in the introduction) can cause misunderstandings and complicate or even impede communication across different disciplines. For example, the term “social” is being used in many metaphors (e.g. social software, social network, social robots), however it often causes misunderstandings as social scientists interpret “social” quite distinctly compared to, for example, computer scientists. Hence, there are different concepts that are being transmitted (see also the introduction for another example on page 6).

Additionally, metaphor research is one of the most important issues we are facing in OPAALS in connexion with language development and change as well as knowledge management. Metaphors highlight, represent, and reflect the knowledge system of a community. They force the process of (re-)structuring and change of knowledge.

For our further work in Phase II we will rely on the linguistic analysis of textual information as it provides a solid background for applications using computational techniques. The presented possibilities of linguistic analysis are widely used for different applications in general. However, we argue that these methods are important in our work related to metaphor analysis as well, especially at the preprocessing step. They provide effective techniques of filtering and reduction of textual information. The statistical measures (such C-value, TFIDF) can be used for the extraction of terms (concepts) that then will be analysed through the metaphorical lens (i.e. differentiation between metaphorical and literal usage of a term or concept). Based on the linguistic preprocessing, combination of rules construction (for metaphors) and n-grams approaches can be applied in order to identify the most important patterns of metaphorical expressions.

The results of the linguistic analysis can then be further expanded by applying LSA (Latent Semantic Analysis) and SVM (Support Vector Machines) approaches to train the system to identify metaphorical expressions. The existence of knowledge system(s) is necessary in the whole process of metaphor extraction. WordNet will serve as the main lexical reference. Already identified and categorised metaphorical expressions (through the manual analysis based on qualitative data gained through the social research methods) will be stored in the MTK and can serve as an additional reference.

Our ongoing work will also include:

- evaluation and manual analysis of data gained through qualitative work carried out in Phase I (T6.1)
- finalising the structure of MTK including the modification of the interface and methods of analysis (linguistic, statistical and computational)
- implementation of the MTK
- integration of WordNet
- enhancement of the text corpora
- testing and evaluation of the MTK
- applying the results from the MTK to socio-pragmatic questions regarding community building, and to the development of an evolutionary framework for language

Finally, we need to stress that the MTK is not designed to become a fashionable “gadget” for all OPAALS members. It is rather its results that will be used and hopefully be helpful for other tasks and researchers than the tool kit itself. As we pointed out, the MTK includes qualitative, linguistic analysis which is therefore the reason why the tool itself is not so much designed as a mass “medium”. However, the results certainly may influence the “masses” or the OPAALS and DE community as the MTK analyses natural language. And it is language that constructs our communities and by means of which we can depict, share, and manage knowledge.

References

Aristoteles (1995). *Poetics*. Cambridge, Mass. Harvard University Press.

Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography, 3*(1), 23—35.

Biber, D., Conrad, S., & Reppen, R. (2004). *Corpus linguistics: investigating language structure and use.* Cambridge: Cambridge University Press.

Black, M. (1962). *Models and metaphor*. Ithaca: Cornell University Press.

Brooke-Rose, C. (1970). *A grammar of metaphor*. London: Seeker and Warburg.

Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Basingstoke, Hampshire: Palgrave Macmillan.

Danesi, M. (2006). The sense implication hypothesis and idealized cognitive models. *Semiotica, 161*(1/4), 185–198.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391-407.

Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Derrida, J. (1982). White Mythology. In J. Derrida, *Margins of Philosophy*, Brigthon: Harvester.

Encyclopaedia Britannica. (1976). Chicago: Encyclopaedia Britannica Inc.

Farghaly, A. (Ed.). (2003). *Handbook for language engineers.* Stanford, CA: SCLI Publications.

Fass, D. (1991). Met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics, 17*(1), 49—90.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analysing unstructured data*. Cambridge: Cambridge University Press.

Frantzi, K., Ananiadou, S. & Mima, H. (2000). Automatic recognition of multi-word terms. *International Journal of Digital Libraries, 3*(2), 117–132.

Frath, P. (2001). Polysemy, homonymy and reference. *Proceedings of the JASGIL Seminar,* Strasbourg.

Gedigian, M., Bryant, J., Narayanan, S., & Ciric, B. (2006). Catching metaphors. In *Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, New York City, 41-48.

Gibbs, R. W. (1999). Researching metaphor. In L. Cameron & G. Low (Eds.), *Researching and applying metaphor* (pp. 29—47). Cambridge: Cambridge University Press.

Grady, J. E. (1999). A typology of motivation for conceptual metaphor: correlation vs. resemblance. In R. W. Gibbs & G. Steen (Eds.), *Metaphor in cognitive linguistics* (pp. 79–100). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Indurkhya, B. (1992). *Metaphor and cognition: an interactionist approach.* Boston: Kluwer Academic Publishers.

Johnson, G. J. (1994). Of metaphor and the difficulty of computer discourse. *Communications of the ACM, 37*(12), 97—102.

Jurafsky, D. & Martin, J.H. (2000). *Speech and Language Processing: an introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, New Jersey: Prentice-Hall Inc.

Kintsch, W. (2000). Metaphor comprehension: a computational theory. *Psychonomic Bulletin & Review, 7,* 257—266.

Knowles, M. & Moon, R. (2006). *Introducing metaphor*. New York: Routledge.

Kövecses, Z. (2002). *Metaphor: a practical introduction*. Oxford: Oxford University Press.

Krishnakumaran, S. & Zhu, X. (2007). Hunting elusive metaphors using lexical resources. In *Workshop on Computational Approaches to Figurative Language (NAACL 2007),* Rochester, New York: Association for Computational Linguistics, 13-20.

Lakoff, G., & Johnson, M. (1992). *Metaphors we live by.* University of Chicago Press.

Lotman, Y. (1990). *Universe of the mind: a semiotic theory of culture*. Bloomington: Indiana University Press.

Luckmann, T. (2005). On the communicative construction of reality. *Lecture to the LSE Department of Information Systems 2nd February 2005.*

Marschark, M., Katz, A.N., & Paivio, A. (1983). Dimensions of metaphor. *Journal of Psycholinguistic Research, 12*(l), 17-40.

Martin, E. F. (2004). Assessment outcome coherence using LSA scoring. *Academic Exchange Quarterly.*

Martin, J. H. (1992). Computer understanding of conventional metaphoric language. *Cognitive Science, 16*, 233—270.

Mason, Z. J. (2004). Cormet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics, 30*(1), 23—44.

Megerdoomian, K. (2003). Text mining, corpus building, and testing. In A. Farghaly (Ed), *Handbook for language engineers* (pp. 213—268). Stanford, CA: CSLI Publications.

Milios, E., Zhang, Y., He, B., & Dong, L. (2003). Automatic term extraction and document similarity in special text corpora. *Proceedings of the 6th Conference of the Pacific Association for Computational Linguistics (PACLing'03)*, Halifax, Nova Scotia, Canada, 275-284.

Morris, M. (2000). Metaphor and philosophy: an encounter with Derrida. *Philosophy, 75*, 225-244.

Nöth, W. (2006). Yuri Lotman on metaphors and culture as self-referential semiospheres. *Semiotica 161*(1/4), 249-263.

Popp, C. (2001). Nomenclatures and computer-aided terminography. In F. Mayer (Ed.), *Language for special purposes: perspectives for the new millennium* (195—201). Tübingen: Gunter Narr Verlag.

Porter, M.F. (1980). An algorithm for suffix stripping, *Program, 14*(3), 130−137.

Ricoeur, P. (1975). *The rule of metaphor: multidisciplinary studies of the creation of meaning in language* (R. Czerny, K. McLaughlin & J. S. J. Costello, Trans.). London: Routledge and Kegan Paul.

Sag, I.A., Baldwin, T., Bond, F., Copestake, A. & Flickinger, D. (2002). Multiword expressions: a pain in the neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, Mexico City, 1—15.

Schank, R.C. (1972). Conceptual Dependency: a theory of natural language understanding. *Cognitive Psychology, 3,* 552-631.

Schmitt, R. (2005). Systematic metaphor analysis as a method of qualitative research. *The Qualitative Report, 10*(2), 358—394.

Sclano, F. & P. Velardi (2007). TermExtractor: a web application to learn the shared terminology of emergent web communities. To appear in *Proc. of the 3rd International Conference on Interoperability for Enterprise Software and Applications (I-ESA 2007)*, Funchal, Portugal.

Shelestiuk, H. V. (2006). Approaches to metaphor: structure, classifications, cognate phenomena. *Semiotica: Perspectives on Metaphor, 161*(1/4), 333–343.

Simpson, J., & Weiner, E. (Eds.). (1989). *The Oxford English Dictionary* (2nd ed.) Oxford, USA: Oxford University Press.

Steinhart, E. C. (2001). *The logic of metaphor: analogous parts of possible worlds*. Dordrecht: Kluwer Academic Publishers.

TemaTres: Open source thesaurus management, http://www.r020.com.ar/tematres/index.en.html

Tirrell, L. (1991). Reductive and nonreductive simile theories of metaphor. *Journal of Philosophy, 88*(7), 337-358.

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11*, 203–244.

Utsumi, A. (2006). Computational exploration of metaphor comprehension processes. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society (CogSci2006)*, 2281-2286.

Veale, T. (1995). Metaphor, memory and meaning: symbolic and connectionist issues in metaphor interpretation. *PhD dissertation.* Trinity College, Dublin, Ireland.

Wharton, C. & Kintsch, W. (1991). An overview of construction-integration model: a theory of comprehension as a foundation for a new cognitive architecture. *ACM SIGART Bulletin, 2*(4)*,* 169 – 173.

Way, E. C. (1991). *Knowledge representation and metaphor* (Vol. 7). Boston: Kluwer Academic Publishers.

Zhang, Y., Zincir-Heywood, N. & Milios, E. (2005). Narrative text classification for automatic key phrase extraction in web document corpora. In *Proceedings of the Seventh ACM International Workshop on Web Information and Data Management (WIDM'05)*, Bremen, Germany, 51- 58.