Differentiae Specificae in EuroWordNet and SIM*u*LLDA

Maarten Janssen

UiL-OTS, Utrecht University
Trans 10, 3512 ED Utrecht, The Netherlands
m.janssen@let.uu.nl

Abstract

(Euro)WordNet, like all other semantic network based formalisms, does not contain differentiae specificae. In this article, I will argue that this lack of differentiae specificae leads to a number of unsurmountable problems, not only from a monolingual point of view, but also in a multilingual setting. As an alternative, I will present the framework proposed in my thesis: SIM*u*LLDA. The SIM*u*LLDA set-up not just contains differentiae specificae (called definitional attributes), but differentiae specificae form the building blocks of the system: the relations between meanings are derived from the application of Formal Concept Analysis to the set of definitional attributes.

1. Introduction

Given the many shortcomings of systems based on semantic primitives, WordNet, like many other lexical databases and knowledge bases, is based on semantic networks (see for instance Miller (1998)). In semantic networks, there is no need for anything like semantic markers or, as you would call them from a lexicographers point of view, differentiae specificae, since all information is formulated in terms of relations between (in the case of WordNet) synsets. In this article, I will argue that this lack of differentiae specificae leads to a number of insurmountable problems, not only from a monolingual point of view, but also in a multilingual setting.

As an alternative, I will present the framework proposed in my thesis (Janssen, 2002): SIM*u*LLDA, a Structured Interlingua MultiLingual Lexical Database Application. The SIM*u*LLDA set-up not just contains differentiae specificae (which are called definitional attributes in the system), but differentiae specificae form the building blocks of the system: the relations between meanings are derived from the application of a logical formalism called Formal Concept Analysis (FCA) to the set of definitional attributes.

After the presentation of the framework, I will indicate why definitional attributes do not give these traditional problems by showing that the resulting framework should not be viewed as an ontological hierarchy, nor as a knowledge base, but as a modest lexical database.

In this article, the following notational conventions will be used: meaning-units, in the case of WordNet the synsets, will be typeset in SMALL-CAPS, word-forms are set in sans serif, differentiae specificae, as well as the relations in WordNet, in **bold-face**.

2. The Need for Differentiae Specificae

One of the main aspects of the WordNet system is its ontological hierarchy, provided by the **is_a** links. Although not de facto a separate system (the **is_a** link is just a link as any other), the hierarchy is often presented that way, and many applications of the WordNet database only make use of this ontology. So for the moment I will consider the (ontological) hierarchy of WordNet as a system on its own.

The **is_a** relation links a synset to its *genus proximum* (to use the lexicographer's term), hence strongly characterising the meaning of the synset by indicating what kind of meaning it is. But on its own, the **is_a** link does not fully characterise the meaning of the synset: it fails to distinguish the various hyponyms of the same synset. From the point of view of the hierarchy we also need *differentiae specificae* to keep the meanings/synsets within the same genus apart.

In the WordNet approach, this differentiation is done by means of the other links. As an example, one could define the synset ACTRESS by means of an **is_a** relation to ACTOR, and a **female** relation the other way around (or alternatively a **is** relation to FEMALE). But although the other links in WordNet do provide additional information about the synset, they are not designed to provide differentiae specificae. This shows in two ways: firstly, the other links give information independent of the **is_a** link, which means that they are independent of the information already provided by the **is_a** link. So they cannot structurally supplement the information lacking from the **is_a** link.

Secondly, not all differentiating information can be modelled by means of these other links. Consider for instance the word *millpond*, which **is_a** AREA OF WATER. But a millpond is not just any area of water, it is specifically one *used for driving the wheel of a watermill* (according to LDOCE). And there are no WordNet links for this type of differentiating information.

So differentiae specificae as such do not exist in WordNet, even though in some (or many) cases the differentiating information will be present or can be provided somehow. This absence of a structural modelling of differentiae specificae leads to serious problems. Let me illustrate this using three examples.

The first example is that, according to Vossen & Copestake (1993), (Euro)WordNet has problems dealing with verb nominalisations: SMOKER is a hyponym of PERSON, but so are RUNNER, SLEEPER, JOGGER, etc. The point here is not so much that distinguishing these nominalisations is impossible in WordNet: in principle, these can be distinguished by means of the **involved_agent** relation. So we can express that the involved agent for SMOKE is SMOKER, and hence by means of backward search say that a smoker is a person *who smokes*. The point is that for synsets with large numbers of hyponyms, there is no structural way of telling them apart: WordNet in many cases depends on the ontological hierarchy, so the less layered it is, the less informative it is.

The second example makes a similar point: because of the high dependence on hierarchy, WordNet is forced to accept as layered a structure as possible: to indicate the relation between ENEMY and MURDERER, WordNet has to introduce a synset for BAD PERSON, even though there are no words related to that synset. This introduction of 'empty synsets' is not really incorrect, but at least conceptually unattractive.

The lack of differentiae specificae is most disturbing when considered in a multilingual setting. As a third example, consider the Spanish word DEDO. It is a (translational) hyperonym of both the English FINGER, and the English TOE, since a finger is a *dedo del mano*, and a toe is a *dedo del pie*. The way this is modelled in EuroWordNet is as follows: the Spanish DEDO has an **eq_synonym** relation to an InterLingual Item (ILI) DEDO, and both the English FINGER and TOE are related to this same ILI with a relation **eq_has_hyperonym**[1]. In this way, the words finger and toe are correctly modelled as translational hyponyms of dedo.

But in this cross-linguistic linking, there is nothing keeping the two translational hyponyms *finger* and *toe* apart. That is to say, language internally, FINGER will have a **part_of** relation to HAND, and TOE to FOOT, but this information is not (directly) related to the cross-linguistic link to DEDO. Furthermore, if we would use these **part_of** relations to tell the translational hyponyms apart, they would be used as differentiae specificae. And there are other examples in which such differentiae specificae are not available. For instance, the French BIEF will be linked as a translational hyponym of CANAL, but the reason why bief is more specific (namely that it is a canal *bringing water from a stream to a hydraulic installation*) would not be modelled, because WordNet has no links to provide for it.

Such examples show that in a lexical database, there is a definite need for a structural modelling of differentiae specificae, especially in a multilingual setting. Although in this section, the criticism is specifically aimed at (Euro)WordNet, any hierarchy based system without a structural modelling of differentiae specificae will encounter the same problem, though they might show up in a different guise. Let me now turn to the system proposed in my thesis which does use differentiae specificae.

3. SIM*u*LLDA

In my thesis, I describe a multilingual lexical database set-up called SIM*u*LLDA, in which differentiae specificae play a crucial role. The differentiae specificae are modelled within the system by means of entities called *definitional attributes*. The SIM*u*LLDA system is designed to be a multilingual lexical database system from which bilingual definitions between arbitrary pairs of languages in the system can be derived.

The SIM*u*LLDA set-up consists of a number of steps: the data from monolingual dictionaries are reduced to sets of definitional attributes. These sets of definitional attributes are turned into a lattice structure by means of a logical formalism called Formal Concept Analysis (FCA). The result is a lattice structure, which can serve as a structured interlingua, connecting words from different languages. Let me show how this works using a simple example: the words for horses in English. This explanation is very brief; for a more complete explanation I refer to my thesis (Janssen, 2002).

3.1. Creating Sets of Definitional Attributes

The hierarchical set-up of the SIM*u*LLDA system is best shown using a small and simple lexical field, such as the words for male, female, young, and adult horses in English. The SIM*u*LLDA system aims at modelling lexicographic data, so takes the definitions of these words as found in a monolingual dictionary as a starting point. The relevant definitions are given in table 1 (these are cleaned-up version of the definitions in the Longman Dictionary of Contemporary English, henceforth LDOCE).

colt a young male horse
fil·ly a young female horse
foal[1] a young horse
mare a fully-grown female horse
stal·lion a fully-grown male horse

Table 1: Definitions of Words for Horses

The definitions in table 1 are analysed in the SIM*u*LLDA set-up as relating English words to defining aspects of the meanings expressed by these words. These defining attributes are called *definitional attributes*. As an example, the first definition in table 1 relates the word colt to the definitional attributes **male** and **young**. On top of these definitional attributes, colt is related to a sense of horse. But this meaning of horse is itself also related in the dictionary to definitional attributes and a further meaning of animal, etc. This will go on until the genus term is what you might call an *empty genus term*. The claim is that *thing* in a definition reading *a thing which* ... is just there because a lexical definition without a genus term is hard to formulate (in some cases). In this way, all lexical definition can be 'unravelled' into sets of definitional attributes. For simplicity, I will here ignore the relation of the words in table 1 to the word horse, and treat **horse** as if it were a definitional attribute. This leads to a situation in which the definitions in table 1 are analysed as in table 2.

	horse	**male**	**female**	**adult**	**young**
HORSE	×				
STALLION	×	×		×	
MARE	×		×	×	
FOAL	×				×
FILLY	×		×		×
COLT	×	×			×

Table 2: Definitional Attributes for Horses

So in the SIM*u*LLDA set-up, every word expresses a number of meanings, and these meanings are analysed in terms of sets of definitional attributes. And these defini-

[1] This situation is symmetrical in EuroWordNet: DEDO and FINGER are also related via the ILI FINGER. But that has no impact on the example.

tional attributes are nothing more than the accumulated differentiae specificae from their lexical definitions in monolingual dictionaries.

3.2. Formal Concept Analysis

The data in table 2 are organised within the SIM*u*LLDA set-up by means of a logical framework called Formal Concept Analysis (henceforth FCA). FCA was developed by Ganter and Wille in Darmstadt (Ganter and Wille, 1996). It is an attempt to give a formal definition of the notion of a ‘concept’, within the boundaries of a model-theoretic framework. The idea behind FCA is the following: in a model, those objects that share a common set of attributes belong together; they form the extension of a concept, the intention of which is the set of attributes that they share.

The formal representation of FCA is follows. Take a set of objects *G*, a set of attributes *M*, and a relation *I* relating the objects to the attributes. We define the set of formal concepts $\mathfrak{B}$ over a context (G, M, I) in the following way:

$$B^{\downarrow} = \{g \in G \mid \forall b \in B \,.\, (g, b) \in I\} \tag{1}$$

$$A^{\uparrow} = \{m \in M \mid \forall a \in A \,.\, (a, m) \in I\} \tag{2}$$

$$\mathfrak{B}(G, M, I) = \{\langle A, B\rangle \mid A = B^{\downarrow} \wedge B = A^{\uparrow}\} \tag{3}$$

The way FCA is applied in SIM*u*LLDA is as follows: the meanings in table 2 are taken as formal objects (the elements of *G*), and the definitional attributes relation to them are taken as formal attributes (the elements of *M*). This lead to a set $\mathfrak{B}$ of formal concepts consisting of pairs of sets of meanings and sets of definitional attributes. There are ten such formal concepts in total, which are listed in table 3.

⟨{HORSE, COLT, STALLION, MARE, FOAL, FILLY}, {**horse**}⟩
⟨{MARE, FILLY}, {**horse, female**}⟩
⟨{MARE}, {**horse, female, adult**}⟩
⟨{STALLION, COLT}, {**horse, male**}⟩
⟨{STALLION, MARE}, {**horse, adult**}⟩
⟨{STALLION}, {**horse, male, adult**}⟩
⟨{FOAL, COLT, FILLY}, {**horse, young**}⟩
⟨{COLT}, {**horse, male, young**}⟩
⟨{FILLY}, {**horse, female, young**}⟩
⟨∅, {**horse, female, young, male, adult**}⟩

Table 3: Formal Concepts for Horses

The formal concepts in $\mathfrak{B}$ have a natural order: formal concepts with more defining attributes are more specific those with less defining attributes. And also, all those objects that belong to a subconcept also belong to its superconcept. So we define an order relation $\leq$ over $\mathfrak{B}$ as follows:

$$\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_2 \subseteq B_1 \tag{4}$$

The relation $\leq$ orders the formal concepts in table 3 into a lattice structure, which can be displayed in a Hasse-diagram as in figure 1. The nodes in this lattice represent the formal concepts, where the related sets of meanings and attributes can be found as follows: all formal concept below the node above which the definitional attribute **young** is placed have **young** in their set of definitional attributes, and conversely, all nodes above COLT have COLT in their set of meanings (i.e. a definitional attributes **a** is put above $\langle \mathbf{a}^{\downarrow}, \mathbf{a}^{\downarrow\uparrow}\rangle$, and a meaning A is depicted under $\langle \text{A}^{\uparrow\downarrow}, \text{A}^{\uparrow}\rangle$).

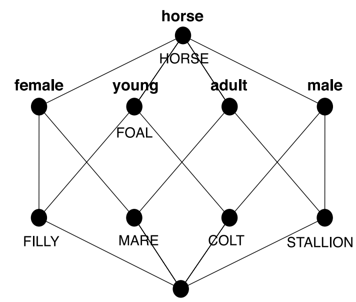

Figure 1: Concept Lattice for Horses

The construction of a concept lattice from a tabular representation of a context can be done automatically on-line by means of Java Applet written as part of my thesis. The Java-Applet is called JaLaBA (a Java Lattice Building Application). JaLaBA gives ask for a set of formal objects and a set of definitional attributes, and a relation between them, gives the related set of formal concepts, and then displays a 3D rotatable model of the corresponding Hasse diagram. JaLaBA can be found on the web-site of my thesis: `http://maarten.janssenweb.net/simullda`.

3.3. Interlingual Concept Lattice

The meanings in SIM*u*LLDA are abstracted from monolingual dictionaries. So the meanings STALLION in table 2 is derived from LDOCE. But the meaning STALLION as such is not an English meaning: the same meaning can be expressed by the French word étalon. Therefore the formal objects in SIM*u*LLDA are not taken to be language dependent meanings, but rather *interlingual meanings*, which can be expressed by words in various languages. It is clear that the definitional attributes defining these interlingual meanings cannot be language specific themselves. So also definitional attributes in SIM*u*LLDA are interlingual entities: **female** is a language independent definitional attribute, that can be lexicalised in English by the expression *female*, but also in French by the expression *femelle*, or in Dutch by the expression *mannelijk*.

Since the lattice in figure 1 thus contains only language independent entities, it can be taken as an interlingual structure, to which words of various languages can be related. This gives the situation as depicted in figure 2. Some notational conventions related to this figure: every interlingual meaning *y* has a (possibly empty) set of words lexicalising it in every language *X*, denoted by $\text{wrd}_X(y)$, and every word *x* of every language has a set of interlingual meanings *Y* it expresses, denoted by $\text{mng}(x)$.

In the set-up depicted in figure 2, it is possible to find translational synonyms: *x* is a translational synonym of *y*, iff $\text{wrd}_Y(\text{mng}(x)) \supseteq y$. To give an example:

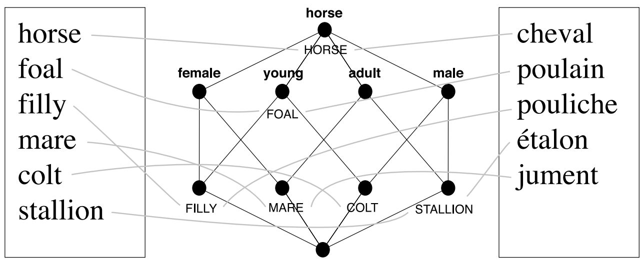

Figure 2: Concept Lattice with Words

mng(stallion) $\supseteq$ STALLION, and wrd$_{French}$(STALLION) $\supseteq$ étalon, so étalon is a translational synonym of stallion. In other words, just following the lines gives you translational synonyms.

More interesting is the situation when there is a lexical gap. In the SIM*u*LLDA set-up, there is a lexical gap iff $\mathrm{wrd}(\mathrm{mng}(x)) = \emptyset$. An example of a lexical gap in figure 2 is that there is no French translational synonym for colt. There only is the more general translational hyperonym poulain.

To find a translational hyperonym for a word *x*, first take mng*x*), and look up the lattice to find the first superconcept which has an interlingual meaning depicted under it for which there is a lexicalisation in the target language. So for colt, this interlingual meaning would be FOAL, and the fact hat poulain is a translational hyperonym of colt is modelled by the fact that COLT $\subseteq$ mng(colt), the related formal concept $\langle \mathrm{COLT}^{\uparrow\downarrow}, \mathrm{COLT}^{\uparrow} \rangle$ (I will use **COLT** as a name for this formal concept) is a subconcept of **FOAL**, and wrd$_{French}$(FOAL) $\supseteq$ poulain.

As claimed in the previous section, the things keeping colt and poulain apart should be the differentiae specificae. And differentiae specificae are implicitly present in the SIM*u*LLDA set-up: if we consider the formal concepts **COLT** and **FOAL**, then by the simple fact that **COLT** $\leq$ **FOAL**, we know that **COLT** has more definitional attributes than **FOAL**. If we define a function *ext* to give the set of definitional attributes of a formal concept ($\mathrm{ext}(\langle A, B \rangle) = B$), then this *definitional surplus* will be ext(**COLT**)$\backslash$ext(**FOAL**) $=$ **male**. So **male** is the differentiam specificam distinguishing COLT from other hyponyms of FOAL such as FILLY.

The differentiae specificae, as well as the genus proximum, are hence modelled at the interlingual level. Within the interlingua, you could say that 'COLT = FOAL + **male**'. The language specific differentiae specificae are obtained by taking the lexicalisation in the desired language of this definitional surplus. We get the translation of our lexical gap by lexicalising both parts of the right-hand side of this equation in the target language. Since the French lexicalisation of **male** is *mâle*, we can conclude that colt in French is *poulain mâle*. This process of generating a translation for a lexical gap is called *lexical gap filling*. Notice that the lexical gap filling procedure renders what Zgusta (1971) calls an *explanatory equivalent*, and not a *translational equivalent*.

We could also have opted to lexicalise all elements of the above equation within the same language, hence in English relating the word colt to the description *male foal*. In this way, also lexical definitions can be retrieved from the system. Notice that this lexical definition *male foal* is not the same definition as the one that formed the starting point of the analysis (see table 1): LDOCE does in fact not give the genus proximum, but a more remote genus term. But firstly, the rendered definition is nevertheless correct, and secondly, the LDOCE definition can also be rendered in the same way: we also have that **COLT** $\leq$ **HORSE**, with a larger definitional surplus: {**young, male**}. This leads to the original definition of colt as *young male horse*. The claim is that the generation of lexical definitions, as well as the lexical gap filling procedure, does not give a unique result, but does give only correct results.

Let me conclude this section by observing that not all definitional attributes are as 'simple' as the ones in this example. For instance, the Petit Robert definition of bief is *canal qui conduit les eaux d'un cours d'eau vers une machine hydraulique*[2]. There is no translational synonym in English for bief, but given an analysis of the data in SIM*u*LLDA, we would have that 'BIEF = CANAL + **qcledcvumh**', where the lexicalisation in English of CANAL would be canal, and the English lexicalisation of **qcledcvumh** would be *bringing water from a stream to a hydraulic installation*. So any differentiam specificam can be captured by a definitional attribute.

4. Definitional Attributes

As I have tried to show in the previous two sections, there is a definite need for differentiae specificae in a lexical database, especially in a multilingual one. That it is possible to set up a system using such differentiae specificae such as in the SIM*u*LLDA set-up. And that such a set-up leads to a correct modelling of lexical relations even in such problematic cases as lexical gaps. But of course the differentiae specificae introduced in a system, such as the definitional attributes in the case of SIM*u*LLDA, are at least reminiscent of the very thing WordNet reacted against: Katz & Fodor style semantics primitives (Katz and Fodor, 1963). So naturally, from the perspective of semantic network theories, there is a reluctance to introduce differentiae specificae.

In the theory of Katz & Fodor, semantic markers are supposed to provide the foundation of knowledge, by their being innate building blocks to which all concepts can be reduced. But the presence of semantic primitives does not necessarily entail such a strongly reductionistic theory of meaning; there are more modest versions of semantic primitives, such as for instance in the French tradition of *sémantique interpretative*, as advocated by Rastier (1987), Pottier (1980) and others. The semantic primitives in this theory are called *sèmes*, which constitute meaning units calles *sémèmes*. Rastier explicitly discusses that sèmes do not have any of the strong properties semantic markers are supposed to have: they are not innate, not universal, not (interestingly) indivisible, they are not (necessarily) small in

[2]It actually is *canal de dérivation qui* ..., but I want to avoid here the for this point irrelevant question whether *canal de dérivation* should be taken as a complex genus term, or whether *de dérivation* counts as a differentiam specificam.

number, and they are not qualities of a referent or part of a concept. Especially in its description by Messelaar (1990), sèmes have a striking resemblance to definitional attributes.

I do not want to give here an elaborate description of sèmes, their relation to semantic markers or a comparison to the SIM*u*LLDA set-up: definitional attributes are not sèmes either. But it is important to observe that the introduction of definitional attributes does not entail a strong theory of meaning. Definitional attributes are meant to be little more than what they are: theoretical entities that help to distinguish hyponyms of the same genus, and that make it possible to generate bilingual lexical definitions even for non-corresponding meanings. In my thesis, I give a lengthy discussion of the nature of the basic element of the SIM*u*LLDA set-up: words, word-forms, languages, interlingual meanings, and definitional attributes. For the moment, I will merely mention three properties definitional attributes are explicitly *not* supposed to have.

Firstly, definitional attributes do not form a special closed set of indivisible, innate semantic primitives. This should be clear from the example in section 2: the differentiam specificam *used for driving the wheel of a water-mill* will constitute a definitional attribute, even though it has a clear internal structure. As a definitional attribute, it will count as an atomic entity, disregarding its internal structure[3]. So it is not an interestingly indivisible definitional attributes. And it would clearly be absurd to suppose that such a definitional attribute is in any way innate. New concepts arise every day, and new concepts can entail new definitional attributes, so there is not even a closed set of definitional attributes: new definitional attributes are introduced when need arises.

Secondly, sets of definitional attributes do not constitute a complete description of the concept related to the word that expresses the interlingual meaning in question. That is to say, interlingual meanings in the SIM*u*LLDA set-up are in a way defined in terms of sets of definitional attributes. But that does not result in saying that all information related to the word expressing that interlingual meaning is captured by the definitional attributes. For instance, stylistic information and other language-internal characteristics of the word are not modelled by the interlingual meaning, but handled at the level of the individual languages. Also, prototypes play an important role in the information/concept related to a word. But prototypes cannot be interlingual since, as shown by for instance Putnam (1988), prototypes do not translate[4]. So the SIM*u*LLDA set-up is not supposed to provide a knowledge base: it is a lexical database, containing some aspects of word-meaning. In particular those aspects necessary for producing the kind of bilingual definition found in bilingual dictionaries.

[3]In my thesis, I discuss some cases in which adopting a certain internal structure for definitional attributes proves beneficial, and also discuss order sets of definitional attributes, but in general, definitional attributes are atomic.

[4]Putnam goes on to claim that *perceptual prototypes may be psychologically important, but they just aren't* meanings – *not even "narrow" ones (op.cit. p.46)..* Although I am not unsympathetic with this point, it is not this strong claim I am aiming at here.

Thirdly, definitional attributes are not denotational in nature. Definitional attributes are aspects of word meanings, not of (the) objects denoted by those words. And the interlingual meaning and/or the related set of definitional attributes are not supposed to fix the denotation of the word. Denotational semantics is very problematic, and it is even very dubious if every word(meaning) can be said to have a fixed denotation at any given moment. Furthermore, denotational semantics can never give a complete picture of word meaning. For instance, words can be metaphorically attributed to objects, where the meaning of the word is applied without the claim that the object to which it is attributed falls under the denotation of the word. So the fact that within the SIM*u*LLDA set-up, **COLT** is a subconcept of **FOAL** is not intended to express the ontological inclusion of the class of colts in the class of foals[5]: SIM*u*LLDA provides a lexical hierarchy, which should not be taken as an ontological hierarchy.

This last point is independent of the presence of differentiae specificae: also hierarchical systems without differentiae specificae, such as WordNet, should be taken as providing a lexical hierarchy, and not an ontological hierarchy. It is even dubious whether there really is an ontological ordering on the world. This is not to say that SIM*u*LLDA is not an ontology in the sense often used in computer science. For instance, the set-up is in many ways comparable to the *ontology clustering* set-up proposed by Visser & Tamma (1999), which has a shared ontology and attributes over the concepts in it. Also in their set-up, a translation for a lexical gap is created after "*the attributes of the concept in the source ontology are compared with the attributes of the hypernym [found in the shared ontology] to select the distinguishing features.*" The point is that SIM*u*LLDA does not provide an ontological hierarchy in the philosophical sense.

Given the modest nature of definitional attributes, it will be clear that there are no strong claims concerning the meanings in the SIM*u*LLDA set-up. This is not surprising if you consider that SIM*u*LLDA aims at modelling lexicographic definitions, and lexicographic definitions do not really 'give' a description of the meanings of a word; they rely on knowledge of related words to 'hint at' the meaning of the word. A nice example of this is given by Hanks (2000), who shows that a lexicographic definition of a chinaman (say *a left-hander's googly*) is only useful if you know about googlies, leg breaks, off-breaks and related cricket terms. Given the elusive nature of words, any theory that makes strong(er) claims is likely to runs into grave problems.

5. Conclusion

In this article, I hope to have shown the need for a structural modelling of differentiae specificae in a (multilingual) lexical database, and the advantages of the SIM*u*LLDA set-up which has such differentiae specificae by means of its definitional attributes. As already said, the criticism in this article was mainly directed at the EuroWordNet set-up, but

[5]This independently of the questions whether all colts are in fact foals.

applies equally to other hierarchical systems without differentiae specificiae. For instance, as far as I can tell, the SIMPLE framework, which in a way is a succesor of EuroWordNet, does not add structure to overcome the problems described in section 2.

Of course, the question whether SIM*u*LLDA could really provide a better alternative for a system like EuroWordNet is an (at least partly) empirical question: lexical databases and knowledge bases are designed for practical applicability. The SIM*u*LLDA approach is, however, a theoretical feasibility study, performed as a PhD-project, and the SIM*u*LLDA system has not (yet) been implemented or tested at large scale.

This is not to say that there is no empirical evidence for the applicability of the system: in my thesis, there is an empirical test whether the around 50 words for bodies of water from 6 different languages (English, French, Dutch, German, Italian, and Russian) can be correctly handled within the SIM*u*LLDA set-up. Describing the results of this test here would be too lengthy, and the test did bring forward some problems (or weaknesses) of the set-up. But the claim is that all the problems that have a solutions could be solved to satisfaction within the system. Although this does not provide a large-scale test, it does show that within an actual domain of lexical definitions, the systems works properly. The lexical field was not arbitrarily chosen, but was taken because it is a lexical field that is often quoted as problematic, both in terms of definability, as in terms of cross-linguistic differences, such as the often cited case of *river* and *fleuve*. So it is intended to provide some empirical evidence for the practical applicability of the system. But the only way to really test it is of course to build an application and fill it with data.

6. References

Bernhard Ganter and Rudolf Wille. 1996. *Formale Begriffsanalyse: mathematische grundlagen*. Springer Verlag, Berlin.

Patrick Hanks. 2000. Contributions of lexicography and corpus linguistics to a theory of language performance. In *Proceedings of the Ninth Euralex International Congress*, Stuttgart.

Maarten Janssen. 2002. *SIM*u*LLDA: a Multilingual Lexical Database Application using a Structured Interlingua*. Ph.D. thesis, Universiteit Utrecht, Utrecht.

Jerrold J. Katz and Jerry A. Fodor. 1963. The structure of a semantic theory. *Language*, vol. 39:170 – 210.

P.A. Messelaar. 1990. *La Confection du Dictionnaire Générale Bilingue*. Peeters, Leuven.

George A. Miller. 1998. Foreword. In Christiane Fellbaum, editor, *Wordnet: an Electronic Lexical Database*. MIT Press, Cambridge. 1998.

Bernard Pottier. 1980. Sémantique et noémique. *Annuario de Estudios filológicos*, vol. 3:169 – 177.

Hillary Putnam. 1988. Fodor and block on “narrow content”. In *Representation and Reality*. MIT Press, Cambridge.

François Rastier. 1987. *Sémantique Interprétative*. Presses Universitaires de France, Paris.

Pepijn R.S. Visser and Valentina A.M. Tamma. 1999. An experience with ontology-based agent clustering. In Benjamins, Chandrasekaran, Gomez-Perez, Guarino, and Uschold, editors, *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5)*, Stockholm.

Piek Vossen and Ann Copestake. 1993. Untangling definition structure into knowledge representation. In Biscoe, de Paiva, and Copestake, editors, *Inheritance, Defaults, and the Lexicon*. Cambridge University Press, Cambridge. 1993.

Ladislav Zgusta. 1971. *Manual of Lexicography*. Mouton, Den Haag.