Department of Applied Mathematics
Faculty of EEMCS

University of Twente
The Netherlands

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl
www.math.utwente.nl/publications

Memorandum No. 1750

Structuralistic linguistics: The case of knowledge graph theory

C. Hoede

February, 2005

ISSN 0169-2690

Structuralistic Linguistics: The case of knowledge graph theory

C. Hoede

Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
P.O.Box 217
7500 AE Enschede, The Netherlands

Abstract

The theory of knowledge graphs is discussed with respect to various points of view in structuralistic linguistics. In particular the essential problem of translation is approached in the context of the theory.

Key words: structuralism, linguistics, knowledge graphs, translation.

AMS Subject Classifications 2000: 05C99, 68F99.

1 Introduction

Natural language has been studied from many different points of view. One of these points of view is given by knowledge graph theory, a theory developed at the University of Twente since 1982. A *graph* G=G(V,E) consists of a set V of *vertices* and a set E of *edges*, pairs of vertices from V. The reader is referred to any textbook on graph theory for the, in most cases self-evident, terminology. G(V,E) is an *undirected* graph. If the pairs are ordered the set of edges is replaced by a set A of *arcs* and G(V,A) is then called a *directed* graph. The edges or arcs may carry labels and the graph is then called a *labeled* graph. We will also consider pairs that can be ordered or not and will then speak of a *mixed* graph.

The labels considered in the theory are modeling relationship types of pairs. The choice of these types stands central in the discussion of the theory as structuralistic theory of language. We will come back to this choice of eight types of binary relationships. First we must mention that there are also four types of *frames*. Suppose we have a mixed labeled graph, which we will call a knowledge graph, for reasons that will become clear soon. Then we may consider any *subgraph*, not necessarily *induced*, and see this as a unit. We will say that the subgraph has been *framed*. The subgraph can be represented by a vertex as well. Its elements, vertices and labeled edges or/and arcs, are put in relationship with that vertex by arcs of the type FPAR, for "Frame PARt". Note that here edges and arcs are seen as single elements, not as pairs of vertices. In the socalled *total* graph representation they can be represented by vertices.

Frames may be labeled too. The most important type of labelling is attachment of a *word*, a *name* for the frame. As this will play a central role in our discussion, throughout this paper, let us already mention that, given a certain knowledge graph, two different subgraphs might be framed and given the same name. The two subgraphs may then be called *homonymic structures.* In case one specific subgraph is framed the naming may be by two, or more, different words, that may then be called *synonymic words.* This already hints at where the theory is aiming at.

The line of reasoning in knowledge graph theory is the following. People, and probably also animals, process perceptions, in the sense that neural networks in the brain determine types of relationships between them. A natural modeling is by representing the perception units by vertices and the relationships by labeled edges or arcs. The "content" of the mind is thus modeled by a *mind graph.*
In the beginning of the knowledge graph project, texts were extracted and the *knowledge* in the texts was represented by what was called a *knowledge graph.* Only three types of relationship were used, a "causal-", a "part of-" and a "kind of-" relationship, see Bakker [1], de Vries [24] and Smit [21]. Most of the texts was simply left out of consideration. In a second phase of the project all elements of texts were considered and the graph structure was investigated on its ability to represent both natural language and logic. For logic the formalism had to be extended by four types of frames. A proposition p is a linguistic expression that may be represented by a knowledge graph, that is seen as a frame labeled p. Van den Berg [2] showed that, by introducing a NEG-frame, for negation, a POS-frame, for possibility, and a NEC-frame, for necessity, the logical expressions $\neg p$, $\Diamond p$ and $\Box p$ can be represented and, together with an AND-frame, therewith the various logical systems. These are the types of frames we mentioned before.

Our discussion will focus on natural language modeling by knowledge graphs, see Willems [27]. The set of types in the *ontology* had to be extended considerably. We will only mention the final result, at the moment of writing. EQU, SUB, ALI and DIS are types, due to the four ways two sets A and B can be related. Equality; $A = B$, being a subset; $A \subset B$, alikeness; $A \cap B \neq \emptyset$ and disparateness; $A \cap B = \emptyset$, determined four types of the relationships in the ontology of knowledge graph theory. The assumption is that brains have developed ways to recognize these four types, due basically to the granular structure of the world.
Two more types are CAU and ORD, for causality and ordering respectively. Possibly the philosopher Hume [20] is right and the ORD-relationship is more basic than the CAU-relationship. But in modeling the physical world causality is extremely important, reason to keep this type in the ontology. Both types refer to the space-time aspects of the world. The seventh and eighth type included, PAR and SKO, are particularly interesting in the context of this paper because they refer to "active" processing by the mind, whereas the six other types of relationships model a more "passive" processing of perceptions. Here

we should immediately say that in the process of thought these types too may play an important role. We will come back to this point later.
The "activity" is best explained by two examples. Suppose a speaker says "this is a nice pitbull", then the words "nice" and "pitbull" are put in relationship with each other in the mind of the speaker. The, rather subjective, statement boils down to linking something named "nice" to something named "pitbull". Unlike in "this is a white pitbull", where both colour and dog are due to perceptions, the word "nice" does not refer to a perception. We typically have to do with an *attribution* by the speaker. "Nice" is an *attribute of* the pitbull, not a *part of* the pitbull, like its tail, which should be modeled by a SUB-link, or a *property* of the pitbull, which should be modeled by an FPAR-link. Note that we have made a precise distinction between three *merological* relationships here. The common use of the preposition "of" has led to the choice of the label PAR. ATT, for attribution, might have been an alternative, but also the FPAR-relationship, between a named frame and its constituents, the properties of the something with that name, is due to an "active" process in the mind. The mind determines what is to be understood under a given name. Hence, PAR and FPAR are similar in nature.
Our second example should explain the SKO-link. The label is deduced from the name Skolem, a logician. It expresses that something is seen as *informationally dependent* on something else. In mathematics a function gives a good example. The value of the function, a number, depends on the number that is mapped by the prescription. Such a mapping is not perceived by the mind, but is actively established by the mind.

These two examples should make our distinction between active and passive processing by the mind clear. Observing some ordering may lead to a "picture" in the mind, that is a mind graph, in which the ORD-*arc* occurs, as the ORD-link, like the SUB-link,the CAU-link, the PAR-link, the SKO-link and the FPAR-link, is oriented. The EQU-link, the ALI-link and the DIS-link are not oriented and are represented by edges. The mind can be seen as just registering the relationships, passively. However, in the process of designing some apparatus, the mind might, actively, use the ORD-link, constructing a part of its mind graph describing the design.

The main point of the theory summarized sofar is the fact that the processing by the mind is seen as an essential part of the theory of language. A language might be seen as a set of words that can be combined in certain linear orderings according to syntactical rules. The users of the language are then left out of the theory. This is very much in line with the way one is dealing with a subject in physics or mathematics. However, that may be the proper way to deal with formal languages, as considered in computer science, it is **not** the proper way to deal with natural language. **The speaker and the listener, and their minds, are essential for a proper account of what natural language is about**.
This comes forward especially with respect to semantics. Two unfortunate things influenced the theory building. The first thing was the restriction to see language as a set of sentences, ruled by a generative grammar, thus focusing on syntax mainly, and restricting the set of sentences to so-called well-formed formulae. This development was strongly in-

fluenced by the development of formal computer languages. The focus on syntax was so strong that Chomsky even suggested that people had an innate ability to generate sentences. Though understandable, this was an unfortunate suggestion. If it were so, the generative grammars of the about 6000 languages in the world should show much more similarity. The ordering of subject S, verb V and object O in a sentence may be as in English, which is an SVO-language. However, all six orderings occur, in Japanese e.g. the verb comes at the end. It is an SOV-language. An innate ability to generate sentences could not be expected to allow this diversity and, in fact, the suggestion was gradually withdrawn.
The focus on syntax led to considerable difficulty in dealing with semantics. The second unfortunate thing was the solution for this coming from the side of logicians, who were inclined to deal with natural language in the same way as with formal languages. Formal languages are important for correctness proofs of algorithms. Does the algorithm do what it is supposed to do? That is the central question posed. Giving its answer led to model-theoretical semantics and truth-conditional semantics. That was an important development, *for formal languages*! The unfortunate thing was making the assumption that the same ideas should be applicable to natural language. A well-known example should make this point clear. Suppose that in a closed room a speaker says :"it is raining outside". The listeners are in control of English and are asked whether they *understand* what has been said. They will probably all say that they do. The fact, whether it is really raining outside or not, is completely irrelevant! Yet semanticists in the mathematical-logical tradition will say something like: "the semantics have not been properly defined". Within the framework of their thinking about language, in general(!), they are right, but only by the assumption that natural language should indeed be dealt with as formal languages are dealt with! The most unfortunate thing is that they form a majority. The number of theorists of language that focus on semantics from the beginning and do not bother so much about syntax, truth values and formal models is considerably smaller. One reference to this "camp" in linguistics is [5]. The theory of knowledge graphs also belongs to this camp.

Let us summarize the basic elements of this theory now. People are considered to have a structured representation in their mind in the form of a mind graph. This mind graph is determined by processing of perception units or by autogenous generation, as the mind can be creative. Our distinction between "passive" and "active" processing was only made to make the position of the PAR-link and the SKO-link clear. In principle there is only active processing.
The vertices of the mind graph are the basic elements of the ontology, called *tokens*. All of them are of type "something", also if they are frames. In first instance no words are involved. The links between the tokens are of eight types. The labels used for these types are not words either. They are used on the meta-level of discussion, as the word token is. Together with the four frame types, the token and the eight types of links form the full ontology of knowledge graph theory.

Words come in as names of frames, by a process of *framing and naming* by the mind.

It is this process that might be suggested to be innate instead of the generation process for sentences. We may use different words for a particular frame, but the process of doing that seems language independent! Another important remark is that *thinking is linking somethings.* This implies that a *thought* is a subgraph of the mind graph and that e.g. designing is an active creation of a subgraph.
The frame that is named by a word is a subgraph called *word graph*. A sentence has a corresponding *sentence graph*. Note that here we should immediately make the distinction between such graphs as we may discuss in our theory and such graphs as "are actually in the mind" of a speaker or listener and of which the former are models.
The most important statement explaining the basic idea of the theory is that *the structure is the meaning*. This expresses that the meaning of a word is the word graph for that word and the meaning of a sentence is the sentence graph for that sentence. In the rest of this paper, we will try to show how these basic ideas compare with other theories of semantics. Here we have chosen as reference the Chapters 8 and 9 of the classical book of Lyons on semantics [16]. Although already dated, this choice avoids the "mer à boire" on semantic theory and allows for easy consultation and comparison by the reader.

For completeness of our account, the third phase of the knowledge graph project should be mentioned. Liu [15] and Zhang [28] wrote theses on the description of the language of Chinese. The work of Willems was extended in this direction with the main goal to prove that knowledge graph theory was indeed capable of dealing with specific features of languages rather different from English. This line of research started with an investigation of word graphs for nouns, verbs and prepositions by Hoede and Li [11]. A second set of word graphs, for adverbs, adjectives and Chinese classifiers, gathered under the name *adwords*, were investigated by Hoede and Liu [12]. A third set of word types, the socalled logic words, was investigated by Hoede and Zhang [13]. In both theses, of Liu and Zhang, the problem of translation was considered, next to the problem of information extraction of texts.

The problem of translation is answered as follows. Given a sentence in language L_1 a sentence graph is constructed. This is called *structural parsing*, which was introduced in Zhang [28]. For this two types of word graphs were introduced for each word, a *syntactic word graph*, describing the way words can be combined, e.g. as adjective and noun, and a *semantic word graph*, describing the meaning of the word. Note that both types of word graph are expressed in terms of the ontology, that contains links on the sub-word level. This makes this way of representation **language independent**.
After structural parsing the sentence graph in L_1 has to be *transformed* into a sentence graph in L_2. This is the main problem for translation and will be discussed further on in this paper. The third phase in translation is *uttering*. This too has been discussed by Zhang [28].

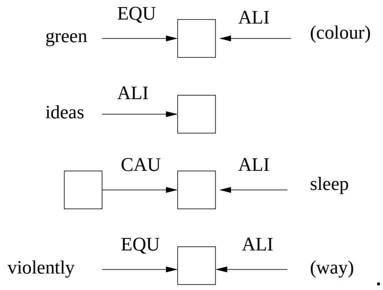
For all other aspects of knowledge graph theory we refer to the references [14], [27], [9], [15] and [28], where it is attempted to give a complete survey, at the moments of writing.

2 The terminological chaos

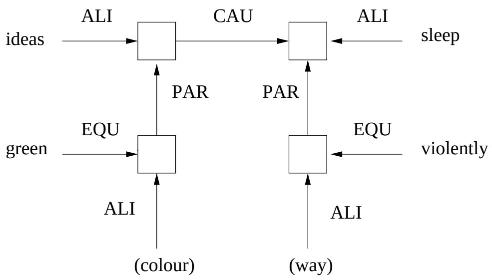
It is remarkable that in a field like linguistics there are papers with a title like: ”Sense, Denotation, Reference: A terminological/philosophical chaos”, by Materna [17]. However, such a title is not completely unjustified. There is unclear use of words, as will become clear in the attempt to defend the knowledge graph position on terminology in the field of semantics.

Historically 1892 can be seen as the year in which the confusion started, with Frege's paper on ”Sinn” and ”Bedeutung”, [6]. A German-English dictionary [4] gives ”sense, organ of perception, faculty, mind” for ”Sinn” and ”meaning” as well as ”sense (of a word)” for ”Bedeutung”, and the problem comes forward. As we have stressed that the minds of speakers and listeners should be taken up into the theory and that minds can be ”active”, we start with looking at the verb ”deuten”, for which we find ”explain, expound, interpret”. In Webster's dictionary [25], we read under ”meaning” a.o. ”SENSE applies especially to words or utterances and may denote one out of several meanings of any one word” and ”MEANING is the general term used of anything (as word, poem, action) requiring or allowing interpretation”.
The first interesting things are that the desciptions involve an active mind, that interprets and gives meaning to something, usually a word or a sentence. Moreover, ”sense” is described as ”denoting one out of several meanings of any one word”. Here we can start defending our theory. Hearing or using a word, a person has an interpretation of the word in mind. This interpretation is, in our theory, a semantic word graph. That word graph is the meaning given to the word by that person. This is not only a subjective process, it can also be done in several ways, i.e., a more or less extended word graph can exist in the mind. All these word graphs are meanings that can be given to the word, there is abundant *homonymy*.
If the word is part of a sentence, likewise various sentence graphs, i.e., various interpretations, can correspond to the sentence. This extends the set of meanings of a word considerably, as now the sentence graph embedding the concept (we will come back to this word) is enlarged. Each knowledge graph containing a specific word is a *meaning* of the word. If two words are contained in the graph this implies that both formally have the same meaning. A distinction can be made by considering *foreground knowledge* and *background knowledge* in a knowledge graph. The subgraph induced by the concept and its direct neighbours is the foreground knowledge about the concept. The subgraph on all vertices but that for the concept is the background knowledge about the concept. Two words in the same knowledge graph have different *foreground meaning* and different *background meaning*.

For our interpretation of ”sense” we discuss a famous example sentence: ”green ideas sleep violently”. The four words may have, depending on the word graph dictionary, representations as word graphs like



A word related to a token by a directed ALI-link gives the type of the token, and when related to the token by a directed EQU-link gives an instantiation of the type. This occurs with "green" and "violently". The intransitive verb "sleep" is having a subject token related to the token for "sleep" by a CAU-link. Transitive verbs also have a CAU-link from the verb-token to the object-token. In the syntactic word graph for a verb the CAU-links are replaced by SKO-links, see Zhang [28]. "Ideas" are just typing a token here, but the token may represent a large frame describing the meaning of "ideas".
Structural parsing leads to the sentence graph



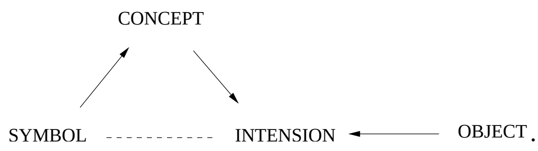
The parentheses around "colour" and "way" indicate that these words are not mentioned in the sentence. The sentence graph is constructed using the syntactic word graphs, that contain a.o. the PAR-links, that have come in. That is why this graph is called a meaning of the sentence, although one may pose the question: "does the sentence make sense?". This is where the semantic word graphs come in. The four graphs given do not forbid the construction of the sentence graph. However, all four have a much larger subgraph of the mind graph as meaning. For linking "green", i.e. a "colour", to something like "ideas" the larger word graph for "ideas" should contain the concept colour as an element that can be attributed to ideas. Likewise the CAU-link between "ideas" and "sleep" may not be present in the extended word graph for "ideas". The given sentence graph, that would describe much more about the four words, would then not be constructible for reason that a PAR-link and a CAU-link cannot be justified. So, *if no sentence graph can be constructed, the sentence has no sense.* If a sentence graph can be constructed, then that sentence graph is a meaning of the sentence and gives a sense to the sentence. An interpretation constructed by a mind is nothing but such a sentence graph and the mind understands the sentence *in that sense.*

We have let the word "denote" slip in when saying that "SENSE may denote one of several meanings of any one word", quoted from Webster's dictionary. For "denote" we read "DENOTE and CONNOTE, when used of words, together equal "mean". DENOTE implies all that strictly belongs to the definition of the word. CONNOTE implies all the ideas or emotions suggested by the word.". This is in line with our theory. Given a subgraph G of the mind graph that a mind associates with a word, then the foreground meaning might be considered by the mind as a definition, although any subgraph G^- of the considered subgraph G, that is associated with the word, might also be taken as a definition. It is this subgraph G^- that is then the *denotation* of the word. All elements of G that do not belong to G^- form the *connotation* of the word. The denotation is thus seen as one of the meanings of the word, depending on the definition chosen by the mind. The addition of "by the mind" is essential for our discussion. As remarked before, homonymy is abundant. A word is given a meaning by "taking together" parts of the mind graph. It is therefore that its meaning, this framed structure, is also called *concept*.

Sofar we have only considered the relationship between name and frame, i.e., between word and concept denoted by the word. The various ways a concept can be thought of, i.e., be linked with other concepts, are *senses* in which the word can be understood. Each sense is a meaning, a structure, namely a subgraph of the mind graph.
It may seem that we try to strictly follow the philosopher Berkeley, known for his thesis that the only thing existing was Berkeley's mind! This is not entirely the case. We do develop our theory from consideration of processes in the mind. However, next to Descartes' "cogito ergo sum", we can put "cogitas ergo es", as argument against Berkeley's thesis. Note that "cogitare" in Latin means "bringing together", but is usually translated by "thinking", in line with our theory. The observation of the terminological chaos hints at the existence of other minds. "You think so you are" is the argument for the assumption that there are other minds, a world outside the mind!

This brings us to the *meaning triangle*, introduced in 1923 by Ogden and Richards [18], and to the term (3) "referent", next to (1) sign, symbol or word and (2) concept, intension, thought, idea , meaning, definition or sense, as other two vertices of a triangle, of which we sofar discussed the relationship and tried to make clear the subtle differences, in particular for the meaning structures.
The *intension* of a word is sometimes seen as synonymous with its *definition*. This way different definitions can be compared and, in the line of "normal" science, a unique definition can be determined. The same, of course, is attempted for word meanings, but dictionaries prove that there is no consensus, as does the title of this section! There is the possibility to reserve the word intension for the whole subgraph of a mind graph, activated by the word. But not only makes this the intension different in the minds of different people, the activation clearly comes "from the outside". Intension is intimately related with *extension*, the set of all existing things to which the word applies.

Before making choices, let us remark that a thought or idea is a subgraph of the mind graph, a frame, that need not be named. The other way around a, strange, word need not have a meaning for the mind. Interesting is the word "unicorn", that does have a meaning and can be defined, say as "white horse with a horn". Let us focus now on the descripton of perceptions. There are many pictures, paintings or tapestries showing unicorns in the world. These pictures are perceived and are represented in the mind as well! This is an important remark. Berkeley is right in stating that basically we have only the mental representation of the outside world at our disposal. The perceived picture is what can be compared with the word graph at our disposal, when trying to give a description. The word intension may be reserved for this substructure of the mind graph, a rather large structure if all details are included. A word to describe it will necessarily have a word graph that is a substructure of the perceived structure. This substructure is then *typing* the perceived structure. The triangle can then be given as

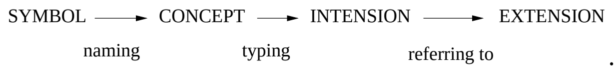


Let the symbol be "unicorn" and let, of all concepts that may carry that name, "living horse with horn" be chosen as unique definition of "unicorn". The perceived structure does not include the substructure to which "living" corresponds and the conclusion is that the concept unicorn does not type the intension. Had the definition been "picture of a horse with a horn", then this concept does type the intension. In the first case the extension is an empty set, as far as we know, in the second case the extension is a rather large set.

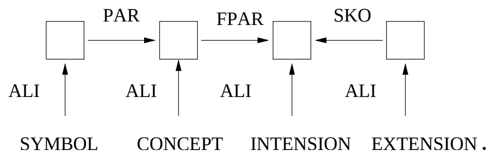
Now the word *reference* can be given a place in our theory. Consider the word "cow". This is the name of a frame, a subgraph of the mind graph of a person. That is what the word directly refers to, and that is its meaning. In the original meaning triangle the top vertex was also described as "reference". Let us now consider the intension of the word "cow". If a cow is perceived a large structure is created in the mind. Perceiving more cows gradually reduces this structure to an image that describes, for one mind, what may be taken as **the** meaning of "cow". However, the definition of "cow" may be considerably simpler, because the perceived common features of the cows seen may be too many to be *brought under words*. The reference of the word "cow" may be, as a substructure, present in the structure that we call intension, distilled from the cows perceived. This reference, as a substructure of the intension, types the perceived objects as cows.

Confusion arises if the word "referent", as that what is referred to, is used for the third vertex of the meaning triangle, where we have sofar considered "intension". For the third vertex not only "referent" or "object", but also "extension" is mentioned in literature. Clearly the real cows are meant! But is a specific real cow a referent for the word "cow"? Our position is that this is not the case. The word "cow" refers directly to a concept, not to an object. Even descriptions like "cow, black and white skin, called Clara, number shield "2" in ear" may not uniquely determine an object and can have a set of cows as objects the intension of which contains that description, in the form of a graph as a substructure. In the extreme case, that there is indeed only one such cow, the description refers to a substructure of the perception, which is then equal to the intension of that, one-element, set!

Instead of speaking of "referent" of the word or description, it seems better to use the word "extension" for the objects that are described, and say that their "intension" refers to them. Then the extension is the referent, but not of the word or description, but of the intension, which is not the same. Usually the first and third vertex of the meaning triangle are depicted as connected by a dotted line, suggesting that the relationship is not that clear. As an alternative we propose a *meaning path* of four vertices in the form



It is interesting to represent this meaning path as a knowledge graph:



All three types of links, that we discussed as describing active processing in the mind, occur, the attribution (PAR), the typing(FPAR) and the informational dependency (SKO).

3 An analysis of sense relations

In this section we intend to show how various distinctions among sense relations holding within sets of lexemes, as discussed in Chapter 9 of Lyons' book [16], are described in knowledge graph theory. We follow his division into subsections. Chapter 8, on semantic fields, will likewise be discussed in Section 4. The translation problem will be discussed separately in Section 5.

3.1 Opposition

Oppositeness of meaning between lexemes is called *antonymy*. Given a word like "white", one might consider the word "non-white". In knowledge graph theory this can be expressed by using the NEG-frame. The problem is whether there is another, less constructed, antonym. For "white" this is not the case, although many people would suggest "black". As the meaning of a word is seen as the, subjective, semantic word graph associated with the word, people may indeed see white and black as antonyms, simply because they do not see a third concept as belonging to the same category as these two concepts. We are forced here to analyze the concepts on an intersubjective, i.e., an objective, level, as is usual in natural science. We recall that in knowledge graph theory the processes in the mind are included in the object of study, in fact are seen as essential for the understanding of natural language.

As for white and black, physics teaches that visible light has a spectrum of wavelengths. Sunlight is seen as composed of photons of all physically possible wavelengths. If an object absorbs photons of certain wavelengths, the object's *colour* is determined by the reflected photons. An object is white when all photons are reflected and black when all photons are absorbed. In this sense white and black are, extreme, instances of colours. In nature reflected light is probably never "monochromatic", i.e., consisting of photons of precisely one wavelength. The important point is, of course, that, next to white and black, there are many other colours. "Non-white" does not imply "black".

The distinction between *gradable* and *ungradable* opposites asks for a discussion too. In case of the presence of some scale, for colours this poses a nice modeling problem that we will not discuss here, the two extremes of the scale may be seen as opposites. Such a pair, however, differs essentially from a pair for which no third comparable can be indicated. The pair "male " and "female", ungradable opposites, gives an example of true antonyms, unless "hermaphrodite" is pointed at.
Gradation involves an ordering. This means that word graphs differing with respect to some arc, in particular to an ORD-link, may correspond with words that are opposites. In its most basic form we find the pair "from" and "to", with word graphs that correspond to the ORD-arc with accent on the begin vertex respectively accent on the end vertex, see Hoede and Li [11]. The same oppositeness is present in pairs like "front" and "back", "up" and "down", etc. Also with respect to other types of directed links, in the ontology of knowledge graph theory, such opposites can be indicated.
With respect to the CAU-arc we mention "cause" and "effect" as well as what Lyons calls *converseness*. A sentence "John killed Bill" has the same sentence graph as the sentence "Bill was killed by John". The sentence graph includes two consecutive CAU-arcs, from "John" to "kill" and from "kill" to "Bill". Uttering can be by an utterance path, see Zhang [28], starting from the token representing "John" or from the token representing "Bill", traversing the same sentence graph against the direction of the CAU-arcs.
The merological relationship types, PAR-link, FPAR-link and SUB-link lead to converse opposites like "with" and "of". For the SUB-arc also the pair "in" and "out" should be mentioned, but here "out", or "ouside", is essentially "not in".
If "antonyms" is used only for pairs of type "W" and "non-W", "with" and "of" as pair differ from "with" and "non-with", i.e., from "with" and "without". The latter pair are antonyms, the former pair are not. They are converse opposites.

3.2 Directional opposition

In discussing the opposition induced by directed arcs of the ontology of knowledge graph theory we already saw that in particular the ORD-link is involved in lexical opposition. Lyons distinguishes between antonymy (narrowly defined in terms of gradability) , complementarity, converseness and directional opposition. We will have to formulate our position here. This also holds for terms like contradictories, contraries, polarity and contrast. The first two of these four terms concern propositions and are therefore not considered here, as we want to compare word meanings.
Polarity refers to distinction of opposites, say "good" and "bad", into a positive and a

negative one. Lyons mentions a preference in stating both words as a criterion for determining their polarity. If people prefer saying "high and low" above saying "low and high", an aspect is coming in that is not an objective meaning aspect. This is reason not to consider this term either. The fourth term, *contrast*, is seen by Lyons as genus of both antonymy and opposition, and we do not consider the term in our attempt to get clear distinctions.
Converseness is exemplified by pairs like "husband" and "wife". Lyons mentions the "converse" propositions: "X is the husband of Y" and Y is the wife of X". Like for the example given in Section 3.1, "John killed Bill" and "Bill was killed by John", we have one sentence graph, now with meaning "male X married to female Y", that can be uttered in two converse ways. This type of contrast was also met with the pair of prepositions "from" and "to", both with word graphs that are subgraphs of the graph consisting of a single ORD-arc. If the graph consists of two vertices, one for tail T and one for dog D, connected by a SUB-arc, we may say "tail T of dog D" or, conversely, "dog D with tail T".

All these relationships between two words, expressed by a knowledge graph of which both are part may be called converse relationships. *Directional opposition* is essentially not of this type, as exemplified in "forward" and "backward", where in a knowledge graph for "moving forward" an ORD-link will occur. That graph does not contain a word graph for "backward", because for "moving backward" the ORD-link has to be reversed, which is essentially different from the situation where the graph is not changed as is the case for converseness.
Lyons writes: "Opposition will be restricted to dichotomous, or binary, contrasts; and antonymy will be restricted still further, to gradable opposites, such as "big" and "small", "high " and "low", etc.". We can agree with opposition as binary contrast. However, the given examples are analogous to our "white" and "black" example. Grading on a scale almost always involves a third element on the scale. We repeat our suggestion to use the word antonyms for pairs of type "W" and "non-W", representable by a frame and the same frame within a NEG-frame respectively.

3.3 Hyponymy

We will not discuss non-binary contrasts, as they are not relevant in the context of knowledge graph theory. *Hyponymy* is relevant. The relationship between "cow" and "animal" in the sentence "a cow ISA animal" has been subject to much discussion in the literature. The question is what ISA means in this sentence. The answer from knowledge graph theory is the following. The meaning of "cow" and "animal" is given by word graphs that may be as large as the intensions of both words, see Section 2. What makes a cow an animal is that the word graph for animal can be found back in the word graph for cow as a substructure. That substructure can be seen as the *genus*, the rest of the word graph for cow as the *differentiae* in the classical description of the type hierarchy in the vocabulary. In knowledge graph terminology the relationship is described by the FPAR-link. The con-

cept "animal" is frame part of the concept "cow". The FPAR-link is therefore inverse to the ISA-link. Whereas the intension of "animal" is frame part of the intension of "cow", the extension of "animal", of course, includes the extension of "cow".
The token in knowledge graph theory is said to be "something", which is a concept of which all other concepts are hyponyms. It is the top of the type hierarchy. An interesting mathematical theory is *formal concept analysis*, developed by Wille *et al.* [8], in which a concept is defined as the pair (intension,extension), i.e., the last two vertices in the meaning path. Knowledge graph theory focuses on the first two vertices of the meaning path.

3.4 Part-whole relations

Lyons discusses lexical gaps, which we do not consider here, and marking of terms, which he calls an extremely important concept in structural linguistics. Formal *marks* are suffixes, like -ess, or prefixes like un-, in- or dis-. They occur rather often. From the point of view of knowledge graph theory these markings correspond to substructures of word graphs. The mentioned prefixes e.g. correspond to the occurrence of a NEG-frame. This is the only thing we would like to mention here.

Very interesting, from the ontological point of view, are the part-whole relations, also called *merological* relationships. In the ontology there are three merological links, the SUB-link, the FPAR-link and the PAR-link, for part-of, property-of and attribute-of respectively. Let us first quote a sentence from Lyons' section on componential analysis: "This approach to the description of the meaning of words and phrases rests upon the thesis that the sense of every lexeme can be analyzed in terms of a set of more general sense-components (or semantic features), some or all of which will be common to several different lexemes in the vocabulary.". This is precisely the thesis of knowledge graph theory, where the ontology forms the set of sense-components.

In particular the three mentioned merological relationships should be enough to represent all part-of relationships. In principle this can only be shown by giving the explicit knowledge graphs in which both elements of the relationship occur. We will just go through a few of the examples mentioned by Lyons.
"Arm " and "body" are linked by a SUB-arc, as is clear from considering the sets of molecules of both. The problem of *transitivity*, in the mathematical sense, can be answered clearly. SUB-relationship and FPAR-relationship are both transitive. The PAR-relationship is not. For that reason the subset-set relationship is represented by a SUB-link, but the element-set relationship is represented by a PAR-link. The transitive FPAR-relationship was considered in particular for hyponyms. If a cow is a mammal and a mammal is an animal, then a cow is an animal. If "animal" is a frame part of "mammal", that is a frame part of "cow", then "animal" is a frame part of "cow".

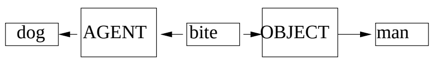
4 Structural semantics

We will now desribe our position with respect to the contents of Chapter 8 of Lyons' book, where socalled semantic fields are discussed.

4.1 Structuralism

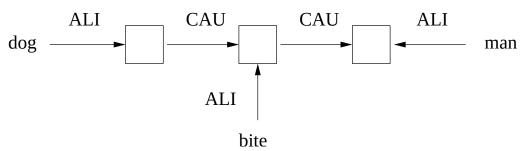
We start with the following quotation: "Linguistic units are but points in a system, or network, of relations; they are the terminals of these relations, and they have no prior and independent existence.". This is completely in line with the paradigm of knowledge graph theory. We recall in particular that in that theory the meaning of a linguistic unit is defined as the knowledge graph to which it belongs, implying that there may be many senses of the unit.

Lyons mentions de Saussure as founder of modern structrural linguistics, whose main work appeared in 1916. Here we should mention the position of Peirce, who introduced *existential graphs*, in 1885, for representing logic. If p and q are propositions, then writing p and q on paper, the *sheet of assertion*, and drawing a frame around them can be seen as representing $p \wedge q$, the proposition "p AND q"" in logic. In knowledge graph theory, developed from 1982 on, the same choice was made for representing the word "and" by a word graph. This happened independently. In fact, two other theories can be mentioned that evolved independently of each other. In 1984 Sowa [22] introduced the theory of *conceptual graphs*, taking the work of Peirce as a starting point. However, the formalism essentially differs from that of knowledge graphs. More importantly, the ways of dealing with semantics differ completely. Conceptual graph theory tries to get in line with model-theoretic semantics, whereas in knowledge graph theory the idea of structure as meaning, like for the AND-frame, is put central. A very simple example should make this important difference clear. The sentence "dog bites man" is represented in conceptual graph theory as



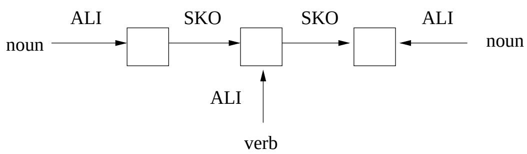
.

In knowledge graph theory the sentence graph is given as



.

This is the semantic knowledge graph. The syntactic knowledge graph is given as



.

Let us now compare these representations. In the conceptual graph we must read: "bite" has AGENT "dog" and has OBJECT "man". Here we have to know what AGENT and OBJECT are. The meanings of these two words are presupposed and not described. In the semantic knowledge graph these words are not mentioned. "Dog" acts like an AGENT, *because of the CAU-link* to the verb "bite". Likewise "man" acts like a PATIENT, *because of the CAU-link* from the verb "bite". The word graphs of these two words are present as substructures. In the syntactic knowledge graph the corresponding substructures are word graphs too, but now for the words SUBJECT and OBJECT, which are syntactical terms! "Agent", "patient", "subject" and "object" are words, with word graphs that determine their meaning. The ontology of knowledge graph theory simply lies on a deeper level than that of conceptual graph theory. More comparable with it is the structuralistic theory of Ebeling [7], of 1974. The author only recently became aware of that theory and made a comparison [10]. The two theories show remarkable parallels.

4.2 The Saussurean dichotomies

We have postponed the discussion of translation to Section 5. Some remarks are due on the terminology of de Saussure.
"Langue" and "parole" have been discussed by many. In line with our focus on mind processes we distinguish on one hand the set of words and sentences of a language (langue), i.e., as observed objectively, *from the outside* so to say. Here we consider purely the grammatical structure. On the other hand we have the understanding, by construction of a mind graph, of what is heard and the bringing under words of what a person has in mind, i.e., the uttering process. That is our position with respect to the term "parole", which is interpreted as language as it is subjectively functioning *on the inside.*

"Substance" and "structure" (form) is another distinction. Clearly we will equate structure with a knowledge graph, in which substructures are word graphs of words in terms of which the structure can be brought under words. For *substance* we would suggest the word graph lexicon at the disposal of the users of a certain language.

The third dichotomy of de Saussure is that of *paradigmatic* versus *syntagmatic* relationships between units in a language-system. The description, given by Lyons, is starting from the term "syntagm", or construction. Meant by "syntagmatic" is that units can be combined in an, also semantically, allowed way. But this we met before, when considering the sentence "green ideas sleep violently". The construction is not allowed semantically. "Green" and "sleep" are not syntagmatically related, *in this sentence*. However, in "green frogs sleep too" the two units are syntagmatically related. The relationship clearly depends on the syntagm considered.
Paradigmatic relationships occur for pairs that may be substituted for each other in a syntagm. Here we have a problem when comparing "big elephants fly" with "big eagles fly". "Elephants" cannot be said to be paradigmatically related to "eagles" as the first syntagm is not an allowed one. We would like to demand that the comparison will be based on allowed structures.

Like Lyons, we will not say much on the *synchronic/diachronic* dichotomy. The terms apply to investigation of languages, at some moment in time respectively in the course of time. For knowledge graph theory only the latter seems interesting. The meaning of words changes in the course of time and therewith the word graphs.

4.3 Semantic fields

We can be short on *relativism* and *functionalism*. The first term refers to the hypothesis that "the actualization of particular distinctions in different language-systems is completely arbitrary". We would say that the choice of word graphs is completely arbitrary and agree with this hypothesis, due to Whorf [26]. At the same time *universalism* is accepted too, due to our thesis that the knowledge graph ontology is universal.
Functionalism, as described by Lyons, "is the view that the structure of every language-system is determined by the particular functions that it has to perform". We will not go into this.

We have to take position to what Lyons says about *semantic fields*. This subject was studied a.o. by Trier [23] (1934), in a time where in physics various "fields" were studied intensively. Possibly the term was introduced in analogy. Like a charged particle derives its behaviour from the electromagnetic field of other charges, linguistic units are considered to be subject to the influence of the whole vocabulary of a language. Lyons mentions the following passage from Trier's work: "Fields are living realities intermediate between individual words and the totality of the vocabulary; as parts of a whole they share with words the property of being integrated in a larger structure (sich ergliederen) and with the vocabulary the property of being structured in terms of smaller units (sich ausgliedern).". The resemblance with knowledge graph theory is striking. The total mind graph of a mind contains all the word graphs at its disposal. The meaning of a word can be any substructure of the mind graph that contains the word graph. The field concept corresponds with such a substructure. Like the word graphs it is part of the whole mind graph and like them it can be expressed in the knowledge graph ontology.

This concludes our remarks on Chapter 8 of Lyons' book.

5 The essential problems of translation

In this section we will discuss translation, a subject dealt with here and there in the book of Lyons.

In knowledge graph theory translation takes place according to:

(1)	Sentence in L1	STRUCTURAL PARSING →	Sentence graph in L1
(2)	Sentence graph in L1	TRANSFORMATION →	Sentence graph in L2
(3)	Sentence graph in L2	UTTERING →	Sentence in L2

A sentence in L_1 is mapped on a knowledge graph called *sentence graph*. This process is called *structural parsing* (1), see Zhang [28]. A sentence graph in L_2 can be *uttered* (3) in L_2, usually in more than one way, see Zhang [28]. The main problem is (2), the *transformation* of the sentence graph in L_1 into a sentence graph in L_2. Liu [15] has investigated the translation of a Chinese sentence into an English sentence as performed by the system of Yao, a Chinese linguist, and compared this with translation according to the knowledge graph approach.
Let us assume that a sentence in L_1 can be structurally parsed and mapped on a sentence graph in L_1. We have to map this knowledge graph on a knowledge graph in L_2 with the same meaning. This is easy when we can apply word by word substitution maintaining the structure. We immediately get the required sentence graph and we can utter this graph by the rules of L_2, which is a process we also assume can be carried out. So where do the real problems come in? We will follow the remarks and examples of Lyons and will also discuss some points, typical for knowledge graph theory.

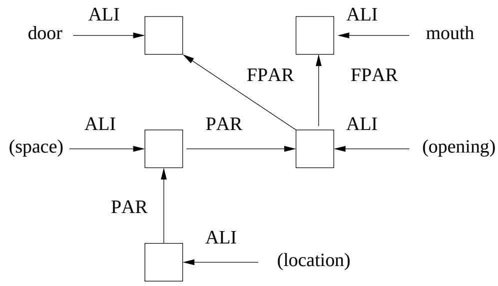
The first problem is that a word has many meanings, its senses. We phrased this as homonymy being abundant. Many word graphs can be associated with a word, each giving a different meaning to the word. Lyons too speaks of *nuclear* or *central* meaning of a word. By using definitions a single word graph can be *expanded* into a larger word graph as the words used in the definition have word graphs of their own. In a sentence graph, by expansion some tokens may be replaced by the frame contents. In this way a sentence graph becomes expanded, in fact in many different directions. Not only in the mind different mind graphs may be generated this way. Also in the case of given descriptions of word graphs we can have a more or less extensive structure, as everyone knows who has ever compared different dictionaries. The sentence graph obtained by structural parsing depends on the quality of the word graph lexicon used. We assume one lexicon that is agreed upon in language L_1 and one in language L_2. But then too expansion can yield different sentence graphs. This opens possibilities for translation as a more explicit sentence graph may allow more ways to transform the graph.

The second problem is that for a word in L_1 no direct counterpart in L_2 is present. For transforming the sentence graph two approaches can be followed. The sentence graph is covered by structures for which in L_2 words exist, and possibly not in L_1. To achieve this covering it might be necessary to expand the sentence graph in L_1. But even then the covering might on one hand be not complete, whereas on the other hand more structure, meaning, might be introduced. For following the first approach a word graph lexicon in L_2 is necessary. A dictionary, explaining a word in L_1 in terms of words of L_2, is necessary in the second approach. These explanations form then structures, syntagms, in L_2 and substitution in the sentence graph in L_1 then gives a sentence graph in L_2. Again that graph may have less meaning than the corresponding sentence graph in L_1, as well as more meaning.

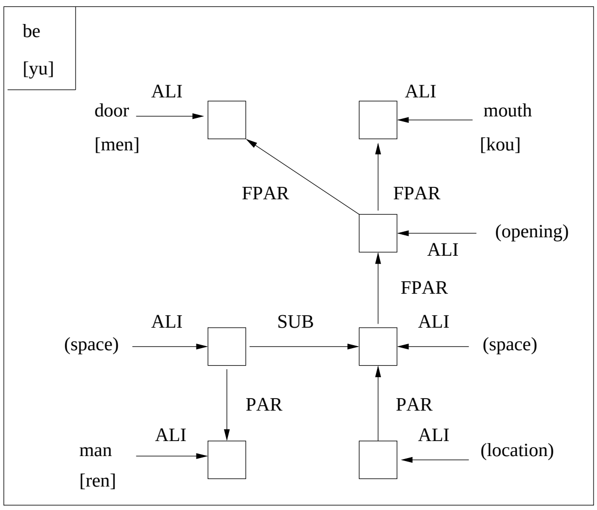
These two problems for an automated translation based on knowledge graph theory form the bottleneck. They make clear that it may be impossible to translate certain sentences in principle! We will now consider a few examples.

Lyons gives the example of translating into French the sentence "The cat sat on the mat". He refrains, as we do, from consideration of tense and some other aspect like rhyme. How to translate "the cat"? In French we have "le chat" and "la chatte". The first cat is male, or its sex is left undetermined. The second cat is female. Anyhow, "cat" and "chat" are denotationally non-equivalent, their word graphs are essentially different. The subject of the sentence can occur in the sentence graph as token of type "cat", where we do not indicate the determiner "the". The expansion of "cat" to "male or female cat" would make the covering possible by using the word graph for "chat" for transforming this part. Likewise information that the cat was female, which might be clear from the context, may make the use of the word graph for "chatte" possible. "Sit" and "sit down" correspond to "être assis" and "s'asséoir", so here the choice is easy. Lyons focuses on the translation of "the mat".
"Mat", "rug" and "carpet", etc. on one hand and "tapis", "paillason", "carpette", etc. on the other hand show no denotational equivalency, due to subtle differences. Lyons' choice to accept this and use the most similar counterpart is the easy way out. It is an example of the second type of problem we mentioned. The word "mat" can be analyzed in French by using a dictionary, replacing that part of the sentence graph by the found structure and possibly covering the resulting structure by other words from the French word graph lexicon. Any further information on "mat", from context, or from an English dictionary, i.e., by expanding first, may make the transformation more easy, as most denotationally equivalent pairs of words will describe less complex concepts, which have a high probability to have a name in both languages. Webster's dictionary [25], gives as one of the meanings of "mat": "a piece of coarse fabric made of rushes, straw or wool". This expansion gives the sentence "The cat sat on the piece of coarse fabric, etc.". The words occurring now may be more easily translated into French than "mat" is.

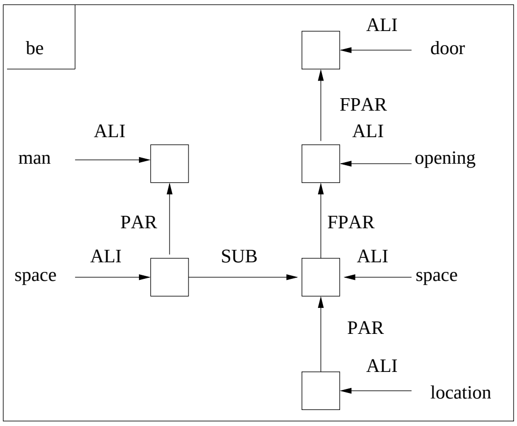
Our second example is the Chinese sentence "Men kou yu ren", literally this reads "Door mouth have man". The English translation given by a translater will be something like "There is a man standing in the opening of the door". We assume that "ren" can be simply replaced by "man". The first problem is one of structural parsing. "Men kou" is a fixed combination in Chinese, but we will consider the combination door+mouth. Only by expansion can the two words be combined. Both "door" and "mouth" have an opening which is a part of space, at a certain location. The combined uttering here focuses on the common element. The knowledge graph is:



"Opening" is connected to "door" and "mouth" by an FPAR-link as we discovered the link between "door" and "mouth" by discovering the common element in the expansions. "Location" is, like "time", attributed to concepts. The parentheses stress that "location", "space" and "opening", do not occur in the (Chinese) sentence.
The second problem too concerns structural parsing. "Have" is read as "be with" in knowledge graph theory. "With" can be represented by any of the three merological relationship types: PAR-, SUB- or FPAR-link. Now the expansion of "man" is needed to link this concept to the knowledge graph we gave before. The fact that "man" occupies space, which can be a part of the space of "opening" suggests the SUB-link, so that we get the following sentence graph.

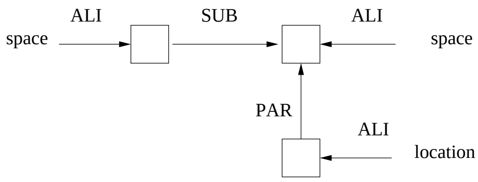


The sentence graph may be read (in Chinese) as "door mouth be with man". But now the transformation of this graph has to take place. Let us assume that all occurring (Chinese) words have direct counterparts in English, which are given in the graph. Then the same graph also gives the sentence graph in English. The second step of the translation therefore seems easy once we have chosen the right expansions. Just *bring* this graph *under words* and then follow the rules for uttering, the third step. However, this will not be what a translater would do. First of all, the word "mouth" would not be used. The opening of the door is essential. The sentence graph may be reduced, without essentially changing what is meant to be said, to



.

We have omitted the parentheses as the words occurring might all be used, *or words that can be present as substructures.* The substructure:



may be present in the English word graph lexicon as word graph for "at". If space is seen as property of door (by the transitivity of the FPAR-link) the concept "opening" might be skipped, like "mouth", and the uttering could be "man at door", which would be extended, by uttering rules, to "(there) is (a) man) at (the) door", which is a perfectly English sounding translation.
Another way of restructuring the sentence graph would be to frame the SUB-link, which is a word graph for "in". Framing the FPAR-link between "opening"-token and "door"-token, which contains a word graph for "of" we may utter the sentence graph as "be man in opening of door", or, extended, "(there) is (a) man in (the) opening of (the) door".
Note that both resulting translations do not cover the whole given sentence graph, in the sense that the sentence graphs of the given translations possibly are not completely identical with the given one. However, by expanding the sentence graphs of these two translations, so now structurally parsing these translations, the given sentence graph in English, or a graph containing it, may be recovered.

These two examples should illustrate the problems automatic translation by means of knowledge graphs is facing. It should also illustrate that a natural translater interprets a given sentence and can choose from several possible transformed and restructured sentence graphs in the target language to achieve an "economic" translation, i.e., grasping the essential meaning in such a way that the given translation is understandable in that way.

References

[1] Bakker, R.R., *Knowledge Graphs : Representation and Structuring of Scientific Knowledge*, PhD thesis, University of Twente, Enschede, The Netherlands, ISBN 90-9001963-4, (1987).

[2] Berg, H. van den, *Knowledge Graphs and Logic : One of Two Kinds*, PhD thesis, University of Twente, Enschede, The Netherlands, ISBN 90-3651835-0, (1994).

[3] Bondy J.A. and U.S.R. Murty, *Graph Theory with Applications*, The MacMillan Press, London and Basingstoke, SBN 333-17791-6, (1976).

[4] *Cassell's German-English Dictionary*, MacMillan, ISBN 0-02-522920-6, (1978).

[5] Eco, U., M. Santambroglio, P. Violi, eds., *Meaning and Mental Representation*, Indiana University Press, ISBN 0-253-33724, (1988).

[6] Frege, G., "Uber Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25-50 (1892).

[7] Ebeling, C.L., *Een inleiding tot de syntaxis*, Summary (J.Wiedenhof en R.Smeets, eds.), The Hakuchi Press, Leiden, ISBN 90-711760-8, (1994).

[8] Ganter, B. und R. Wille, *Formale Begriffsanalyse Mathematische Grundlagen*, Springer, ISBN 3-540-60868-0, (1996).

[9] Hoede, C., *On the Ontology of Knowledge Graphs*. In: Conceptual Structures: Applications, Implementation and Theory (G. Ellis, R. Levinson. W. Rich and J.F. Sowa, eds.), Proceedings of the Third International Conference on Conceptual Structures, ICCS'95, Santa Cruz, Lecture Notes in Artificial Intelligence Nr. 954, Springer, ISBN 3-540-60161-9, 308-322 (1995).

[10] Hoede, C., *Syntax and semantics: A comparison of the structuralistic language theory of Ebeling with knowledge graph theory*, Memorandum No. 1710, Faculty of Mathematical Sciences, University of twente, Enschede, The Netherlands, ISSN 0169-2690, (2004).

[11] Hoede, C. and X.Li, *Word Graphs: The First Set*. In: Conceptual Structures. Knowledge Representation as Interlingua, Aux. Proc. of the Fourth International Conference on Conceptual Structures, (P.W.Eklund, G.Ellis and G.Mann, eds.), Bondi beach, Sydney, Australia, 81-93, (1996).

[12] Hoede, C. and X. Liu, *Word Graphs: The Second Set*. In: Conceptual Structures: Theory, Tools and Applications, Proc. of the 6th International Conference on Conceptual Structures, (M-L. Mugnier and M. Chein, eds.), Montpellier, Springer Lecture Notes in Artificial Intelligence Nr. 1453, 375-389, (1998).

[13] Hoede, C. and L. Zhang, *Word Graphs: The Third Set*. In: Conceptual Structures: Broadening the Base, Proc. of the 9th International Conference on Conceptual Structures, (H.S. Delugach and G. Stumme, eds.), CA, USA, Lecture Notes in Artificial Intelligence Nr. 2010, 15-28 (2001).

[14] James, P., *Knowledge graphs*. In: Linguistic Instruments in Knowledge Engineering, (R.P. van de Riet and R.A. Meersman, eds.), North-Holland, ISBN 0-444-88394-0, 97-117 (1992).

[15] Liu, X., *The Chemistry of Chinese Language*, PhD thesis, University of Twente, Enschede, The Netherlands, ISBN 90-365-1834-2, (2002).

[16] Lyons, J., *Semantics, Volume I*, Cambridge University Press, ISBN 0-521-21743, 230-335 (1977).

[17] Materna, P., *Sense, Denotation, Reference : A terminological/philosophical chaos*, http://www.muni.cz/ materna/sense.html.

[18] Ogden, C.K. and I.A. Richards, *The Meaning of Meaning*, Routledge & Kegan Paul, (1923).

[19] Peirce, C.S., On the Algebra of Logic. *American Journal of Mathematics, 7*, 180-202, (1885).

[20] Russell, B., *History of Western Philosophy, Book Three, Chapter XVII*, Allen & Unwin, London, ISBN 0-04-109009-8, (1961).

[21] Smit, H.J., *Consistency and Robustness of Knowledge Graphs*, PhD thesis, University of Twente, Enschede, The Netherlands, ISBN , (1991).

[22] Sowa, J.F., *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, ISBN 0-201-14472-7, (1984).

[23] Trier, J., Das sprachliche Feld. Eine Auseinandersetzung, *Neue Jahrbücher für Wissenschaft und Jugendbildung* 10, 428-449 (1934).

[24] Vries, P.H. de, *Representation of Science Texts in Knowledge Graphs*, PhD thesis, University of Groningen, Groningen, The Netherlands, ISBN , (1989).

[25] *Webster's New Encyclopedic Dictionary*, Black Dog & Leventhal, ISBN 1-884822-25-8, (1996).

[26] Whorf, B.L., *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*, edited by J.B. Carroll, Wiley, (1956).

[27] Willems, M., *Chemistry of Language*, PhD thesis, University of Twente, Enschede, The Netherlands, ISBN 90-9005672-6, (1993).

[28] Zhang, L., *Knowledge Graph Theory and Structural Parsing*, PhD thesis, University of Twente, Enschede, The Netherlands, ISBN 90-3651835-0, (2002).