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. 1710

Syntax and semantics: A comparison of the structuralistic language theory of Ebeling with knowledge graph theory

C. Hoede

January, 2004

ISSN 0169-2690

Syntax and Semantics: A comparison of the structuralistic language theory of Ebeling with knowledge graph theory

Cornelis Hoede

Faculty of Mathematical Sciences
University of Twente
P.O.Box 217
7500AE Enschede, The Netherlands

Abstract

This paper describes a comparison of two structuralistic theories of syntax and semantics of language. Both theories were developed in The Netherlands. One was developed by C.L.Ebeling at the University of Amsterdam about 30 years ago, the other by C.Hoede and students at the University of Twente in the last 20 years.

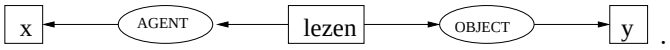
Key words: Syntax, semantics, structuralism.
AMS Subject Classification: 05C99

1 Introduction

The American philosopher C.S.Peirce [9] developed a theory of *existential graphs* at the end of the nineteenth century. His goal was to develop a notational system for logic. The basic principle was very simple. Let p and q be propositions. Introduce two types of *frames* around them, an AND-frame and a NEG(ATION)-frame, that can also be put around a single proposition. We can now represent the two logical connectives $\wedge$ (AND) and $\neg$ (NOT). $p \wedge q$ is represented by writing p and q on paper and drawing an AND-frame around them, $\neg p$ is represented by writing p on paper (Peirce calls this the *sheet of assertion*) and drawing a NEG-frame around it. As the set of connectives $\{\wedge, \neg\}$ is a functionally complete set of connectives, all formulae of first order logic can now be expressed.

In the eighties of the twentieth century two types of graphs for representing logic and language were introduced. Sowa [10] introduced *conceptual graphs*, that

were clearly inspired by Peirce's existential graphs. Every year an International Conference on Conceptual Structures is being held. Only one example should be given here. The sentence "x leest y" ("x reads y") is represented as



We give the sentence in Dutch, because of the comparison we are going to make with a text written in Dutch. Where necessary we will give a translation into English. However, we want to focus on representational aspects mainly. In the figure two types of vertices occur in the directed graph, rectangles and ellipses. The graph should be read as 'lezen' has AGENT 'x' and 'lezen' has OBJECT 'y'. So AGENT and OBJECT are words used to describe how the semantic units 'x' , 'lezen' and 'y' are related, and can therefore also be seen as typing the arcs between these units. Note that these words have a meaning themselves, expressing basically syntactic information. The linking of the semantic units is *on the word level*.

A second representational system was developed, from 1982 on, by the discrete mathematics group in Twente, see the paper of James [8]. *Knowledge graphs* are related to both existential graphs and conceptual graphs, but show some special features. This is mainly due to the way they were developed. The original goal was to represent sociological and medical texts in the form of an expert system. This meant focusing on causal relationships (CAU) between semantic units, and turning them into "if-then"-rules. Soon it was recognized that the CAU-link was not sufficient and that other types of relationship had to be introduced, representing "part of"- and "isa"-relationships. So the types of the links were chosen for structuring the semantic contents of a text in a very restricted way. The three types of links were of very basic nature and can be said to be *on the sub-word level*.

Till 1990 the group in Twente was unaware of the two other representational systems, that were focusing on logic and language respectively. A natural continuation of the study of knowledge graphs was therefore to see whether logic and language could be represented as well. To cut a long story short, van den Berg [2] showed that Peirce's program could be carried out within the knowledge graph formalism by introducing four types of frames in the style of Peirce. These frames were a BE-frame, a NEGATION-frame, a POSSIBILITY-frame and a NECESSITY-frame. In fact van den Berg proved that all known logical systems, including the modal logic systems, could be represented. Moreover, and this is important for our story about language, the frames could be used for knowledge graphs that do not necessarily represent a proposition. Willems [11] started the investigation of the potential of the knowledge graph formalism for the representation of language. It turned out that the number of types of links had to be

extended. By now there are eight binary relationships in the knowledge graph ontology, see Hoede [4] , next to the four types of frame-relationships that express the link of elements of a frame with the frame itself. The choice of the four types of frames was easily defended as the existing logical connectives gave the examples. The important change was the introduction of the BE-frame, as this gave the opportunity to describe concepts, not necessarily propositions. We will come back to this in a moment.

The idea behind knowledge graphs is that perceptions and awarenesses are representable by a token, an unlabeled vertex, called "something". It is the highest vertex in any taxonomy. These tokens are related to each other by binary relationships of such types, as can be recognized by the neural networks in the brain. As neurophysiology cannot give us a list of possible relationship types yet, we have to make a guess. Our brains will, in the course of time, have developed in reflection to sets. This is due to the granular structure of the world, which is a consequence of its quantummechanical nature. Two sets can be related in four ways. If A and B are sets, then we have:

$$\begin{array}{lcl} A = B & , & \text{EQUality of A and B} \\ A \subset B & , & \text{A is a SUBset of B} \\ A \cap B \neq \emptyset & , & \text{A and B are ALIke} \\ A \cap B = \emptyset & , & \text{A and B are DISjoint.} \end{array}$$

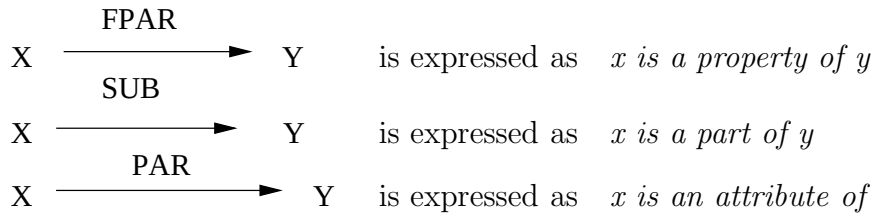
For this reason we choose the following four links in our representational ontology; the EQU-link, the SUB-link, the ALI-link and the DIS-link. Note that the triples of letters are not words but are only used on the meta-level, as types, supposed to be recognizable by the brain.

Another aspect of the world to which the brain has responded in its development is space-time. The basic type of relationship here is the ORDering relationship. We also take up the CAUsal relationship, in spite of the doubt cast upon it by the philosopher Hume. This extends our ontology with an ORD-link and a CAU-link.

Finally, we were forced to introduce two more types, that can be seen as due to the presence of other *minds* in the world. A subjective attribution as in "my nice dog", may not be shared by the listener and led to the introduction of the PAR-link. The last type was introduced by van den Berg and Willems and is called a SKO-link after the logician Skolem. It describes informational dependency between somethings, as e.g. in the mathematical concept of function.

The mind is considered to have a representation in the form of a graph structure. For graphtheoretical terminology we refer to any of the many textbooks on graph theory. The vertices are the tokens, linked by any of the eight types introduced in the ontology. The graph is called a *mind graph*. The slogan is "

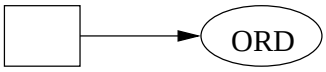
thinking is linking somethings". Now we arrive at a very important point. The second slogan is " the structure is the meaning". A subgraph of the mind graph can be " framed and named". This means that words come in! The various substructures of the mind graph are "brought under words". These substructures can be seen as being put in a BE-frame, that is then labeled by a word. The contents of the frame may contain substructures that have already been baptized. The relationship between elements in the frame with the frame itself is called an FPAR-relationship. Together with the SUB-link and the PAR-link it forms the three merological relationships in the ontology.



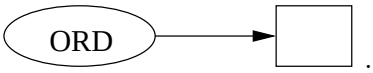
As a very simple example of framing and naming we consider the graph consisting of a single ORD-link, where now the socalled total graph form is chosen, in which the labeled arc is now also represented by a, labeled, vertex. Consider the graph

.

This graph has two subgraphs:



and

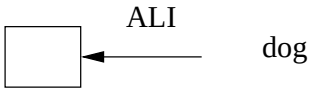


.

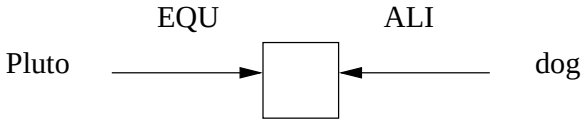
These two very simple graphs can be framed and named FROM and TO. I refer to the paper of Hoede and Li [5] for an account of representing prepositions by knowledge graphs, that for obvious reason are called *word graphs*. A start was made with determining word graphs for the first hundreds of most frequent words in English. Given a sentence and a lexicon of word graphs the next problem is to determine a *sentence graph*. That process was studied by Hoede and Zhang [6] and is called *structural parsing*, see also Zhang [12]. The process was carried out

by determining *chunks*, as introduced by Abney [1], determining *chunk graphs* and connecting these chunk graphs into the sentence graph.

The, *semantic*, word graphs are not sufficient for this process. The way word types, like verb, noun, adverb, preposition etc. may be combined can be expressed by socalled *syntactic* word graphs that are essentially semantic word graphs for word types. So the same ontology is used to express the syntactic functions of words. Before considering an example, we should say that frames can occur as tokens again, which can be typed. So

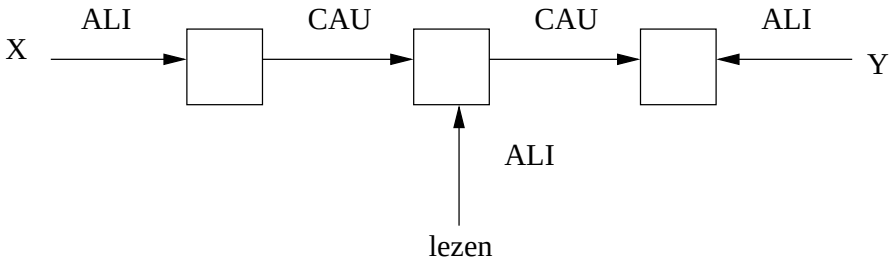


is "something like a dog". An instantiation or valuation is indicated by an EQU-arc to the token. So



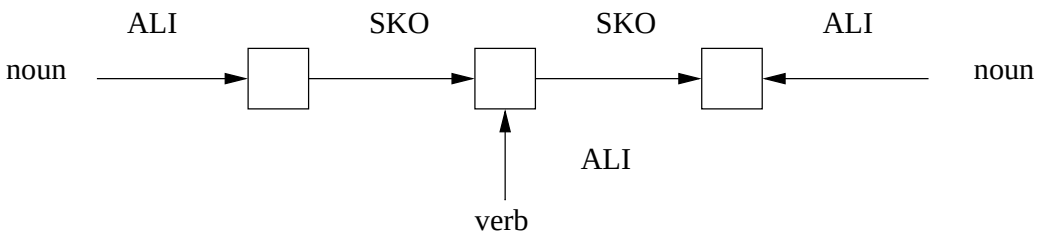
is "dog Pluto".

We consider the sentence "x leest y" again. In knowledge graph theory verbs are connected semantically to semantic units by CAU-arcs and syntactically to word types of those units by SKO-arcs. Think of x as a man and of y as a book, both nouns. Semantically we have the sentence graph



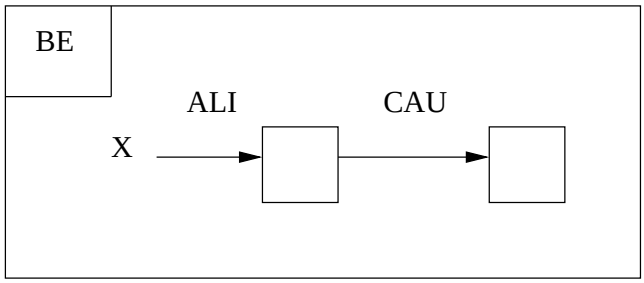
,

whereas syntactically we have

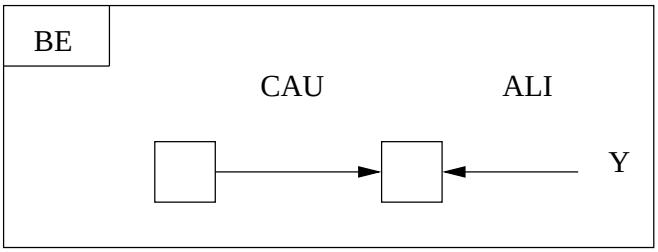


,

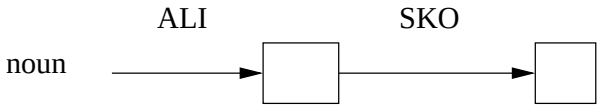
for a transitive verb. The reader should note the precision of this way of representing a sentence.



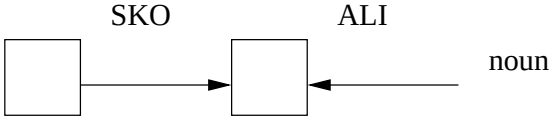
represents "x is the agent of the verb", as only verbs are linked by means of CAU-arcs. Likewise



represents "y is the patient of the verb". In conceptual graph theory this is described by an OBJECT-relationship. However, that is a syntactic term, like SUBJECT is a syntactic term. As both "object" and "subject" are words, there should be word graphs for them. And, indeed, the graphs



and



are graphs that are knowledge graphs for "noun being subject" and "noun being object", the word "being" expressing the BE-frame that can be thought around each of the two graphs.

We now come to the main goal of this paper. The author asked F.Kortlandt of the University of Leiden for a linguistic problem that might be attacked by knowledge graph theory. He pointed out that C.L.Ebeling had developed a theory of syntax and semantics that can also be said to be a structuralistic theory and he suggested to compare both theories. Although the material at my disposal is a summary of Ebeling's theory, I see this as an advantage, as this may keep this paper short. The material was produced under the responsibility of J.Wiedenhof and R.Smeets [3]. My aim will be to discuss Ebeling's theory against the background of knowledge graph theory. It seems important to point out similarities and differences. One similarity should be mentioned immediately: Both methods claim to be useful for the description of *any* language.

2 Comparison of the general approaches

I have decided to start the comparison of Ebeling's booklet (henceforth denoted by E) with knowledge graph theory (henceforth denoted by KGT) at the end; §16.GRAMMAR. For that §2.INVARIANTS should also be considered. A *word form* tree has a set of *appropriate referents*, APR's, being all somethings that can be indicated as realisation of the word. This is usually called the *extension* of the word tree. From the set of APR's of tree a *meaning* "tree" is postulated (§2). The combination (tree,"tree") is called a *sign*. Forms, meanings and signs are invariants, constituents of a language.

In KGT we meet

as " something of type tree". The word tree in this figure is like the form tree in E. The token is a frame containing all features something must have to be of type tree, the *intension* of tree. This structured collection of features corresponds to the meaning "tree". This is a first striking similarity.
Ebeling pays attention to *homonymy*, the same word form being used for more meanings. In KGT this has an extreme form. When people give definitions of a concept, usually several different structures are produced. However, homonymy plays an even much greater role than just being an indeterminacy between two well-defined meanings; E mentions the Dutch word form schimmel as describing plant as well as horse. Any word in KGT has a meaning that is considered to be equal to the graph to which it belongs. This means that in *another context another graph describes the meaning.* Any use of a word puts the word in a specific graph context. All possible graph contexts are meanings of the word. So

there is abundant homonymy. This position of KGT is completely in line with Wittgenstein's view of meaning being determined by use.

Let us now consider a quotation from §16 in E, IC being short for Immediate Constituent and SP for Semantic Particle (§4).

"The generative history of a sentence can be imagined in the following way. First there is a COMMUNICANDUM, i.e. a thought that one wants to transfer. This communicandum can be compared with the meaning of the sentence that in the end will be formed, although differing from it by not being structured in the way prescribed by the language."

The similarity with KGT is striking. The communicandum clearly is the subgraph of the mind graph that is to be "brought under words". This subgraph is the meaning of the sentence that is to be uttered. The uttering process is language dependent, see Zhang and Hoede [7].

"For this the communicandum must be programmed. The first step is to separate it into two parts, related to each other like the IC's of the sentence that will finally be formed, but differing from these IC's by not being structured according to the language. The next steps are further splittings, always in correspondence with the IC-organisation of the sentence. This splitting goes on until each resulting IC can be identified with an SP that can occur in the wanted position according to the law of the language.....Usually certain aspects of the communicandum get lost, whereas other aspects have to be taken up, in order to obtain a well-formed semantic construction."

Again there is similarity with KGT. Given a sentence graph that is to be COMMUNICATED, i.e. uttered in a certain language, then the graph is to be covered by graphs at the disposal of the utterer. In this process the lexicon of graphs, for which the utterer has words, may not be sufficient to cover the whole graph, corresponding to E's loss of aspects. Possibly the covering can be done by a set of word graphs, but the uttered sentence has a sentence graph larger than the original one, corresponding to E's necessarily added other aspects.

In making a sentence graph starting from a sentence, this is called *structural parsing* in KGT, see Zhang [12], the sentence is first cut into chunks. These chunks are then represented by chunk graphs that are combined into a sentence graph. The other way around, in uttering, the splitting up of the sentence graph, i.e. the explicitly structured communicandum, might be done searching for dissection procedures, but is in fact done by constructing an *utterance graph* on the tokens labeled with words at the disposal of the utterer, see Hoede and Zhang [7].

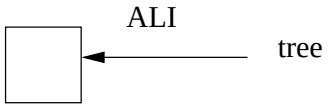
I will not discuss the very short discussion of E on grammar. The comparison with semantic particles is subject of the next section

3 Semantic particles

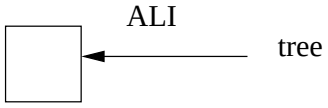
We now consider §4 and §5 of E; SEMANTIC PARTICLES and ABSTRACTIONS.

"Semantic constructions consist of a number of semantic particles (SP's) connected by relation symbols. 'tree' indicates an SP, whereas "a tree", "the high tree" indicates that not necessarily all elements are SP's. SP's are bundles of properties. These meaning properties are distinctive properties of APR's. They are those properties without which an entity would not be an APR of the given SP. Properties are indicated by "|...|", as e.g. in "|human|"."

In KGT the semantic construction is a subgraph of the mind graph. The relation symbols are the types, of the relationships, present in the ontology. The links are between tokens , or "somethings". Sofar we have a parallel situation in E and KGT. However, what about the notations in E? The graph



in KGT indicates "something of type tree", where we use double apostrophes for indicating a linguistic utterance. Let us remember here that the token stands for a concept, a frame around a, possibly large, subgraph of the mind graph. In E "tree" indicates the meaning of the word form tree, but that is the type of the token in KGT. In E (tree,"tree") is said to from an "inseparable combination", called sign. In KGT the graph



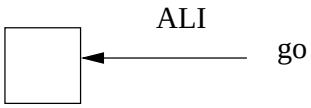
is comparable with this. The word tree, used as a label, is the form in E, the contents of the frame, indicated by the token, forms the meaning "tree" in E. The subgraph itself contains subgraphs that are properties of the concept. If non-human is such a property then in E "|non-human|" would be used to indicate it. But then what corresponds to the SP 'tree' in E? The solution I see is that **the combination of token and label in KGT is what corresponds to the semantic particle**. However, we do meet a first difference here. In KGT

every word has a semantic word graph. The three forms in "the high tree", in the notation of E, all have meanings; "the", "high" and "tree", from the point of view of KGT. Therefore all three correspond to semantic particles. If we use a word, in KGT there must be a corresponding SP.

It may be clarifying to analyze Ebeling's example of the form *went*.

"A *sememe* is an SP (or group of SP's) correlated with something in the form. Correlated with (went,"went") is one sememe, but the latter consists of two SP's, namely 'going' and 'PAST', because these two parts of the notation for a meaning complex will occur in two different places."

Here we clearly meet a deviation of the two theories. In KGT went is a word, so there must be a word graph for it. That graph will contain



as an SP. Note: not 'going', as that word contains the ending -ing, which must be expressable by an extension of the given graph. 'Going' has more meaning than 'go'! Also the SP 'PAST' uses the form past, which word also has a word graph, which may be constructed, using old ideas of Reichenbach on tenses. KGT simply gives a much more precise description than E does. A similar finding can be mentioned for E's §5.ABSTRACTIONS, where the SP is discussed further.

" The properties* constituting an SP are projections of distinctive properties of entities (things) in the world, e.g. the SP 'city' in the sign (city,'city') consists of the property* "|city|", which is a projection of the property |city|, i.e. the property an entity must have to be an APR of 'city'."

This is a difficult sentence in E. Note that I used "property" and "property*" in English translation. E uses "eigenschap" and "kenmerk" respectively, although in §4 E writes:

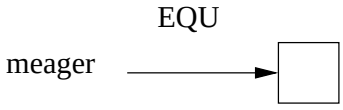
"SP's zijn in de betekenis wat fonemen in de vorm zijn: bundels **eigenschappen (kenmerken)**."

Let us try to disentangle this. E uses the word property for the entities (things) in the world. The projection refers to the mapping to the elements of the mental image made of those entities. These elements are the properties* (kenmerken) of the SP. There seems to be a mistake in E when the sign (city,'city') is mentioned,

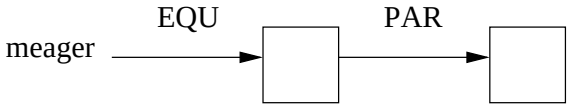
as signs do not have that format. Probably (city,"city") was meant. The SP 'city' corresponding to the sign (city,"city") consists of the property* "|city|", which is a projection of the property |city|. This makes sense. |city| is a property of the APR, and "|city|" is the property* of the mental image to which it is mapped. The properties* of this mental image form the SP 'city'. But this means that, as we have seen before, the semantic particle of E is essentially the word graph of KGT! If this is indeed so, the two theories are remarkably alike. A final remark in this section concerns the following citation.

" ... it is quite usual that in the correlating form of a meaning there is something indicating that an SP, though defined as a projection of a carrier of properties, yet must be understood as a projection of these properties themselves. Something like that in the form is for example the suffix -ness that has this effect in *meagerness*. We symbolize the meaning of *meagerness* as '$<$ *meager* $>$' (sothat -ness correlates with "<>") and speak of an ABSTRACTION."

This can be interpreted in KGT as follows. The carrier of the property is perceived and the property*, as part of the image, is seen as a semantic particle. As any subgraph of a word graph may itself be a word graph, this supports the thesis that SP and word graph are basically the same concept. The example of *meagerness* is illuminating. In KGT "meagerness" is a word and must have a corresponding word graph. That word graph must differ from that for "meager". It is for this reason that the notation <> is introduced in E. The solution in KGT is different, as the representational power of KGT is greater;



could be chosen for the word "meager" and



for the word "meagerness". The EQU-link is chosen as "meager" is essentially a valuation. The first graph expresses "meager" sec, the second graph expresses "meager being attribute of something", i.e. it expresses "meagerness". In the rest of the paper we will now discuss the various concepts in E and compare them with the KGT-formalism.

4 Detailed comparison

4.1 Valencies

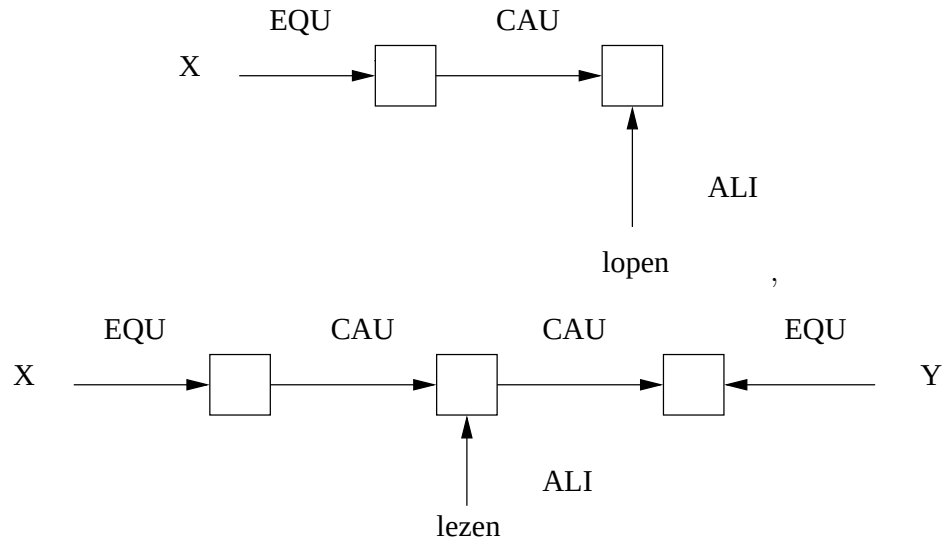
The thesis that in E '...' stands for semantic particle and corresponds to a knowledge graph in KGT is supported in §6 VALENCIES. "x leest y" is a Dutch sentence that in E is dealt with as follows;

‘ [lezend] ’ or also ‘ [X ; X leest Y] ’
[gelezen] [Y ; X leest Y] .

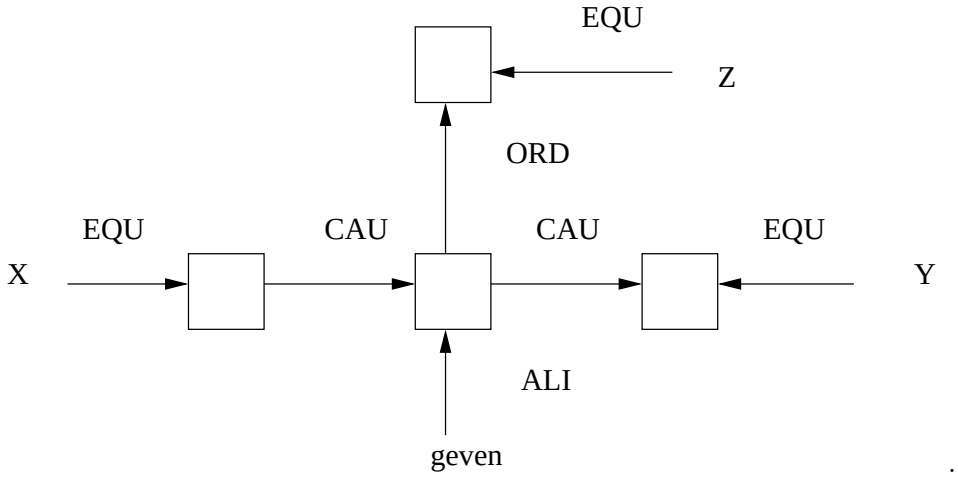
The two *valencies* of one SP '[lezend]' are written in a column. It is said that "lezen" is BIVALENT, whereas "geven" is TRIVALENT:

‘ [X ; X geeft Y aan Z]
[Y ; X geeft Y aan Z]
[Z ; X geeft Y aan Z] ’ .

From the point of view of KGT various critical remarks are to be made about this notation in E. The word valency is used in various textbooks on graph theory, although "degree" is used more often for the number of neighbours, -in or -out, of a vertex. The example "x loopt" is said to pose no problem in §6. '[x loopt]' would be the notation for that SP. Let us consider the three sentence graphs in KGT:

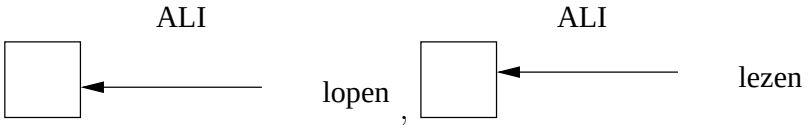


and



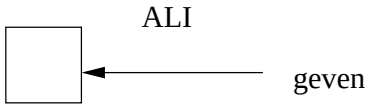
.

The graphs



,

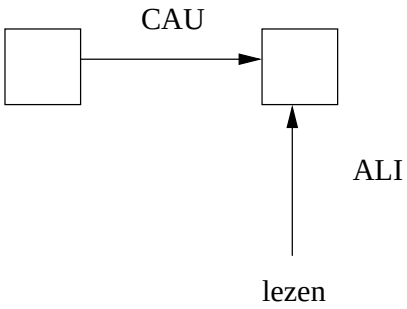
and



are the word graphs for “lopen”, “lezen” and “geven”. Note that these are subgraphs of the given sentence graphs. Let us focus on the example with “lezen”. Clearly the token of type “lezen” is bivalent, i.e. is connected to two other tokens, whereas the tokens for “lopen” respectively “geven” are monovalent respectively trivalent. The criticism on E’s notation comes in when the SP ‘lezend’ is considered and then can be replaced by

$$\begin{matrix} \text{‘} & [\ \text{lezend}\ ] \\ & [\ \text{gelezen}\ ] \end{matrix}\ \text{’},$$

or even a notation in which x and y are mentioned.

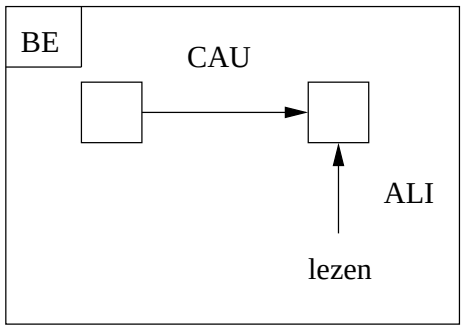


is a word graph, but it is the word graph of “lezer”. Including the EQU-arc from x it is the graph for “lezer x”. The SP ‘lezend’ is written as

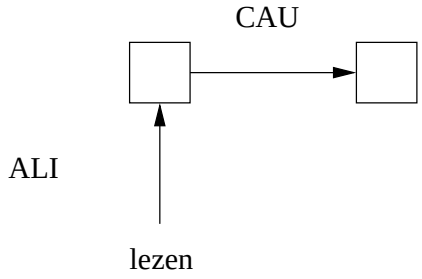
‘ [X ; X leest Y]

[Y ; X leest Y] ’,

but this is strange, because it is suggested that ‘gelezen’, which is also an SP, is of secondary nature, whereas they are called COMPLEMENTARY. In KGT “lezend” has a word graph , as any word, which is the subgraph



Note the use of the BE-frame. This subtly differs from the word graph for “lezer”. The focus is not on the token that is doing the reading, but the graph expresses that “something IS reading” rather than “something who/that reads”. Similarly



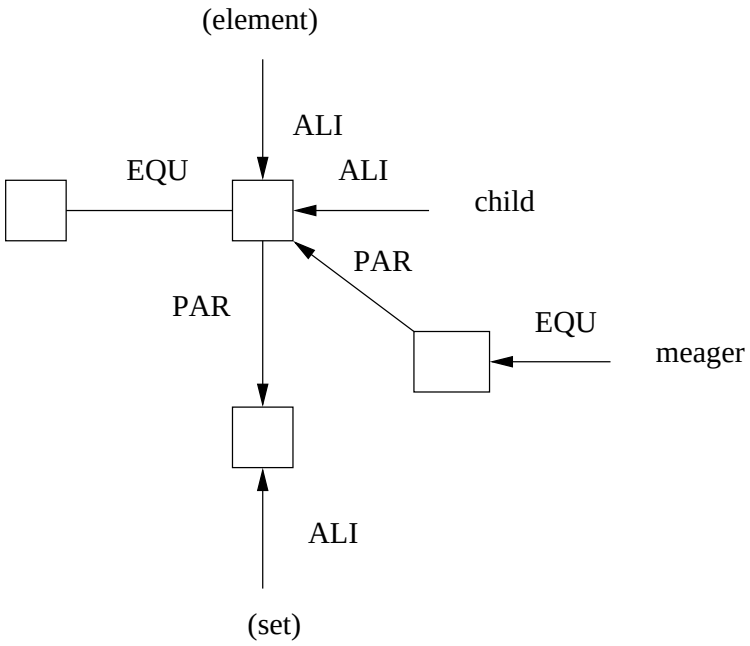
expresses “something read” (“iets gelezens”). With the BE-frame the meaning is “something BEING read” (“iets gelezen wordende”). Note that inclusion of x and y changes the meaning, a distinction not made in E. It seems clear that the whole sentence graph is considered to be the semantic particle.

4.2 Convergence

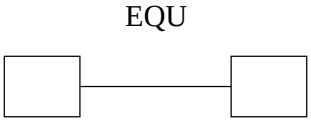
In §7.CONVERGENCE we meet example (a); the partial sentence *het magere kind*. In E the notation is: ‘kind-mager/SING-DE’, and we read that four SP’s are found: ‘mager’, ‘kind’,‘DE’ and ‘SING’. Note that the notation ‘...’, used to indicate an SP, is also used for the whole partial sentence. This again supports our thesis, for which the following citation is important too.

“ If we return to *the meager child*, then we find for this four SP’s: ‘meager’, ‘child’, ‘THE’, ‘SING’. Each of them is a projection of an entity with certain properties*; ‘THE’ (correlating with any particle being used) is short for ‘something that, given the situation of speech and the context, is most probably the thing meant by the speaker’; ‘SING’=‘set consisting of precisely one element’. But not only these four SP’s determine the meaning of the whole: also the fact that in every APR the carrier of each of the four in an SP projected (groups of) properties is always one and the same entity. We express this by saying that ‘meager’, ‘child’, ‘THE’ and ‘SING’ in the meaning of *the meager child* are CONVERGENT.(The negation of convergent is DIVERGENT).”

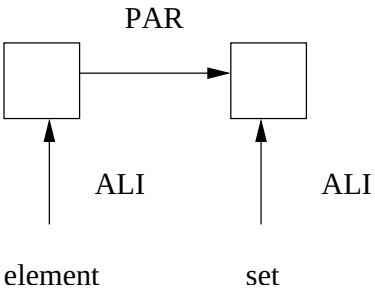
The vital remark is “the four SP’s IN THE MEANING of *the meager child*”. As in KGT the meaning is the knowledge graph considered to represent the partial sentence, this remark shows that in E the four SP’s are seen as constituents of the partial sentence, that is represented as a single SP. But this is completely in line with KGT. The representing knowledge graph is:



The subgraph



is used for reference and is the basic structure representing the particle 'THE'. 'SING' is represented in KGT by the construction



The word child is typing the token for element. If the SP 'children' would have occurred the word children would type the token for set, and the SP 'PLUR' would have been expressed by this knowledge graph. Some remarks about the notation in E should be made. Introduced are " - " and "/". Moreover, the convergency is expressed by placing the elements in the same horizontal line, called LAYER. In KGT this is not necessary at all, the graph may be drawn in

any way, it is its structure that counts. Let us also consider example (b) in §7;

‘ ... child = [X ; X reads Y]

[Y ; X reads Y] ; boek ... ’ .

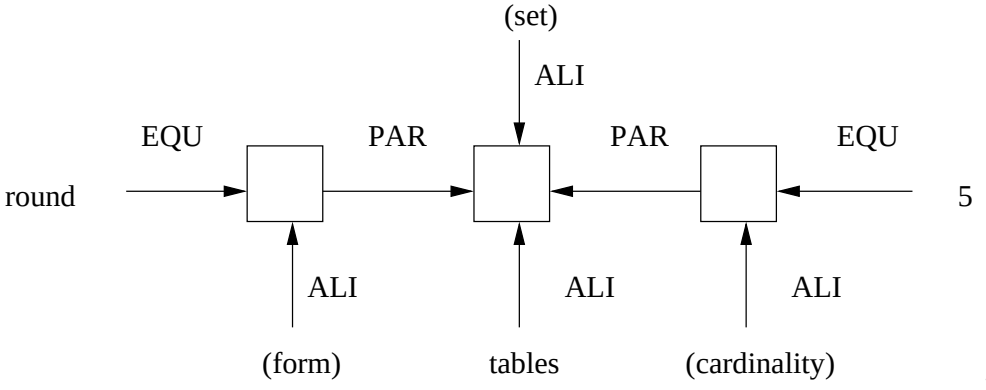
The triple dots indicate that the “semantic construction” (a nice description for a knowledge graph!) is not complete. In fact the SP’s ‘the’, ‘a’ and ‘SING’ (twice) are missing. What makes the discussion somewhat difficult is that the notations in E are not defined at the place where they are introduced. The second example contains the “ = ”-sign as well as the “ ; ”-sign, so in all we have to discuss four notations.
The STRATIFICATION-symbol “/” is explained in §10.STRATIFICATION. We cite:

“ Often language maps a part W of the world by mapping one or more other parts W′ and at the same time indicating what the relationship between W and W′ is. W′ can, as in the following example, be one or more parts of W:

‘ five round tables : table − round / 5 ’

..... assuming that the three SP’s are images of carriers of the properties⋆ |table|, |round| and |set of five elements|, and that W is the five round tables, we conclude that W is only carrier of the third of the mentioned properties⋆: W, seen as a whole, is neither round, nor table. The SP ‘5’ therefore maps W directly, whereas ‘table’ and ‘round’ map an auxiliary piece W′, a constituent of W. The STRATIFICATION-symbol “/” indicates how the notation should be read, where in the definition of ‘5’ must be thought to be included that in the “stratum” placed on its left, the elements of the set are mapped.”

Here the first serious divergence of the two theories occurs. The KGT-representation of this example is:



.

The central token of this knowledge graph is that of type tables. That is a set and therefore the set concept is indeed central, but neither the form nor the number of the tables plays a central role. *39 square tables* is another instantiation of a set of tables of some form. We should now list the semantic particles that occur behind a stratification sign. We find '/SING', '/THE' (§7), '/PLUR', '/NON' (§8),'/5', 'TT' (Dutch for Present Tense) (§9). The example '/TT.ASS∪POS' in §14 will not be discussed. In all cases the SP's do not play a central role. They are annotations of other SP's, involving special constructions in KGT. For example the SP 'NON' is just a NEG-frame around a knowledge graph, the SP 'THE' just a reference to another something, for which in KGT an EQU-link to its representing token is used. In principle situations, described in E by Σ (§9), do not differ much from situations where the stratification '/NON' is used. In KGT we would use a BE-frame.

Let us get back to §7. The title is CONVERGENCE, but in a version of 1978 the name was PARALLELISM. The latter name is not very clear. Convergence is referring to one entity being the carrier of properties projected in an SP. In KGT verbs and nouns are the natural centra in a knowledge graph, to which the other types of words "converge". Not without reason the first rule in a generative grammar is S ⟶ NP VP, i.e. Sentence ⟶ Noun Phrase followed by a Verb Phrase. In the example

‘ . . . child = [X; X reads Y]
[Y ; X reads Y] ; book . . . ’

there are two layers. The first line denotes "the child", a noun phrase, whereas the second line denotes "reads a book", a verb phrase. The notation suggests convergence to two entities, however there are three centra: child, read and book. This becomes even clearer when we consider the sentence: "The young clever childs reads with great haste a very large book". "The young clever child" refers to one entity, as does "a very large book". "reads with great haste" clearly is a part of the sentence where the carrier of the projected properties is one and the same act. In my opinion here E is exhibiting a major deficiency.

We must still discuss the signs " - ". " = " and " ; ". The " - "sign will be discussed in the next subsection. The signs "=" and ";" are meant to denote the same: x=child and y=book, but in §9 we read

"If in such a case something converges with a dominated element, it is connected with ";".".

This brings in the concept of *domination*, which is not defined in E. The book is seen as "dominated", probably because it is the object in the sentence. This

means that the sign ";" has a double meaning, and refers to both a typing of a token (y) and the sysntactic ordering of these words. The notation used is not very luckily chosen. In §9 things become even worse. In example (i); *the child was meager* we read

"We indicate the relationship between 'child' and 'meager' by "=", so '...child=meager...'. Writing 'p=q'indicates that 'p' and 'q' are convergent, but 'p' and 'p=q', as well as 'q' and 'p=q', are divergent. The relationship is called NEXUS.".

This is not admissible. The sign " = " for an instantiation as in child=[x;x reads y] cannot also be used in an expression child=meager, where a noun and an adjective would in some sense be said to be "equal". As Σ expresses a situation, in E one might write

$$\text{`} \quad \cdots \quad \Sigma$$

$$\text{child} - \text{meager} \quad \ldots \quad \text{'},$$

expressing something like "child meager be".

As things are described in E convergence cannot be said to be well-described. Neither are "Σ" and the stratification sign " / " to the point with respect to convergence.

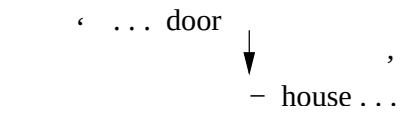
4.3 The other relation symbols

We did not yet discuss the sign " - " as in '...child-meager...' or '...house-high...'. In §11.KINDS OF CONVERGENCE we read

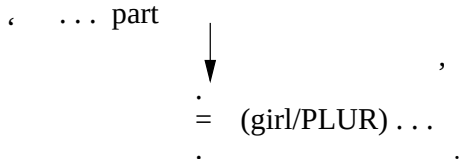
" The DIRECTED LIMITATION in 'p-q' indicates in principle that the referent carries the property |q|, measured according to the norm holding for carrier |p|".

Here the main divergence of E and KGT becomes evident. The link p-q is determined by the fact that the property |q| is also a property of an APR of 'p'. The "direction" is indicated by the ordering , from p, the central concept, to q, the concept that in KGT is called an "adword". Nothing is further said in E about the nature of the "-"-sign. In KGT the direction of the links and, more importantly, the type of the links is chosen from the ontology. For *high house* the syntactic word graphs for the adjective and the noun determine the orientation , of a PAR-link, towards the noun. The semantic word graphs can be used to

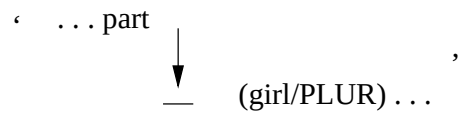
construct the meaning of *high house* in the form of a knowledge graph, see e.g. Zhang [12]. In §12.KINDS OF DIVERGENCE we encounter the notation for *the door of the house*:



The arrow is used to indicate the divergence. As we put in doubt the whole concept of convergence, suggesting that basically verbs and nouns form the central constituents in a sentence, we put in doubt the usefulness of the vertical arrow in the notation of E. *door* and *house* are two separate concepts, two nouns being related by the preposition *of*. In E no reference to *of* is made, although merological relationships play an extremely important role in language. The only explicit remark about merology is found in Example (ll) where “in” is used to indicate PART OF WHOLE as in *part of the girls*:



The expression is ambiguous as, next to “a subset of the set of girls”, it can also mean “the part belonging to the girls” for which E gives:



In the first meaning KGT uses a SUB-link and in the second meaning a PAR-link, from ‘part’ to ‘girls’. In the first case it is unclear why the ‘part’, the subset of girls , is not seen in E as convergent with the SP ‘girls’. The girls of the subset are referents both for ‘part’ and for ‘girls’. We already discussed the unclear status of convergence in the last subsection. In fact, in §8.GRADATION we read some doubt in E as well where the word “strange” (“eigenaardig”) is used in the discussion of the notation ‘$p > q$’. The example given is *remarkably high trees*:

The group ‘high>remarkably’ converges with ‘tree’, but ‘remarkably’ converges with ‘$< high >$’, not with ‘tree’. We recall that ‘$< high >$’ denotes ‘height’. A comparison is made with *the remarkable height of the trees*:

‘. . . [<high>] − remarkable

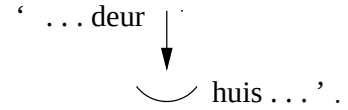
[high] ; tree . . . ’ .

The notation in E is really somewhat confusing. The “ ; ”-sign “equates” high with tree, we saw this kind of flaw before. In the first example one “layer” is used because of a kind of overlapping convergence, i.e. according to E. This could be defended by noting that *trees, being remarkably high* is paraphrasing *remarkably high trees*. However, ‘remarkable’ does not say something about ‘height’, as is said in E. There is a sign (height,”height”) . This can have a DIRECTED LIMITATION ‘height-remarkable’ linguistically, as in the second example, but there is nothing remarkable about a “height”. The height of a tree can be measured in meters. It is the value of the height that may be called remarkable, e.g. when the trees are Banzai trees, usually having extremely low (value of their) height. Also in the first example, what is remarkable is that the trees are so high. This, of course, in comparison with the height of other trees. So here too it is the value of the height that is remarkable.
In KGT these subtle nuances come forward if the word graphs are carefully constructed for the lexicon of word graphs. In a theory where the differences are not taken into account, as in E, the coding will always be difficult if done automatically (per computer), and will strongly depend on the sharpwittedness of the human encoder. This same sharpwittedness is needed in KGT, when the word graphs are constructed. Let me give some examples. During a course on knowledge graph theory 32 students were forming 16 groups of two coders. They were asked to make a word graph expressing the meaning of the word “marriage”. The 16 answers were all different, and even after synonyms had been located still 15 different graphs resulted. Then the graphs were compared and it turned out that they could be combined into one graph, that, according to the whole group, could indeed be seen as THE word graph for the word, or at least as a very good one. The quality of the word graphs considerably differed from student pair to student pair, both with respect to internal consistency and to consequent use of the ontological possibilities.

The picture that comes forward from making a comparison of E and KGT is that Ebeling’s theory was clearly ahead of its time. The goal was very similar to that of KGT. But, as is usual for such theories, the pioneering takes its toll. The structuralistic nature of the theory is expressed in a list of relationship symbols. Not having a direct grip on the explicit structures corresponding to the meanings of the words, as KGT has, a correct coding of (partial) sentences asks for an

encoder, who knows what was intended in the developed system. A computer approach for E seems therefore doomed to fail. The sharp focus on the concept of convergence distorts the encoding of other features. Merological relationships do not get enough attention, singularity and plurality are dealt with in a way that suggests that no other way of dealing with them was seen. The same holds for tenses or negations, also occurring in stratifications. Articles and prepositions are dealt with stepmotherly. All these things get a much more thorough treatment in KGT. That theory too will have to show how to deal with words and their combination. An open problem is e.g. the representation of nouns that consist of two words. E mentions (s) *een frisdrank*:

and(gg) *de huisdeur*:



In the literature “doublenouns” are classified in many types. In E one example, (s), is said to show convergent FIXED ATTACHMENT, and the other, (gg), is said to be divergent. Just two types are distinguished, on the basis of the consideration of APR’s. Like for *part of the girls* one may have doubt about the divergence of *huisdeur*. Anyhow, the distinction is made on semantic grounds, but then the formalism in E is completely inadequate if, say some 30 types of fixed attachment can be distinguished semantically. KGT shows more promise here. Finally, let us consider verbs like *cover* and *discover*. There is a large group of verbs starting with “dis-”. That part of “discover” has a relationship with “cover”. There is no mentioning of this in E, but it is also still not dealt with in KGT.

Concluding I would say that Ebeling’s theory is a pioneering theory, but inadequate for its goals, as they come forward from the text. I do think that KGT is a “better” theory, able to cover the same linguistic features as dealt with in E. But KGT too still has to show its viability.

Acknowledgement . I thank my colleague Kortlandt for suggesting and enabling me to carry out this comparison.

References

[1] Abney, S.P., *Parsing by chunks*, in: Principle-based parsing (R.Berwick, S.P.Abney and C.Yenny, eds.), Kluwer Academic Publishers (1991).

[2] Berg, H. van den, *Knowledge graphs and logic: One of two kinds*, Ph.D.-Thesis, University of Twente, ISBN 90 900 6360 0 (1997).

[3] Ebeling, C.L., *Een inleiding tot de syntaxis*, Summary (J.Wiedenhof en R.Smeets, eds.), The Hakuchi Press, Leiden, ISBN 90 71176 08 , (1994).

[4] Hoede, C., *On the ontology of knowledge graphs*, in: Conceptual Structures: Applications, Implementation and Theory, Lecture Notes in Artificial Intelligence 954, 308-322 (1995).

[5] Hoede, C. and X. Li, *Word graphs: The first set*, in Conceptual Structures: Knowledge Representation as Interlingua, Auxiliary Proceedings of the fourth International Conference on Conceptual Structures (P.W. Eklund, G. Ellis and G. Mann, eds.), ICCS '96, 81-93 (1996).

[6] Hoede, C. and L. Zhang, *Word graphs: The third set*, in: Conceptual Structures: Broadening the Base (H.S. Delugach and G. Stumme, eds.), Lecture Notes in Artificial Intelligence 2120, ISBN 3 540 42344 3, 15-28 (2001).

[7] Hoede, C. and L. Zhang, *Utterance paths*, in: Using Conceptual Structures (B. Ganter and A. de Moor, eds.) Contributions to ICCS 2003, Shaker Verlag, ISBN 3 8344 1705 3, 15-28 (2003).

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

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

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

[11] Willems, M., *Chemistry of language*, Ph.D.-Thesis, University of Twente, ISBN 90 9005672 9, (1993).

[12] Zhang, L., *Knowledge graph theory and structural parsing*, Ph.D.-Thesis, University of Twente, ISBN 9 0365 1835 0, (2002).