D. José Manuel Muñoz Muñoz

THE STRUCTURE OF MEANING DEFINITIONS IN THE LEXICON: A FG PERSPECTIVE

UNIVERSIDAD DE CÓRDOBA

In recent years there has been a growing interest on the part of linguists to look upon language not only from a systemic perspective but also as a vehicle for the codification and representation of knowledge. This interest is shared by computer scientists and knowledge engineers involved in such fields as Artificial Intelligence and Knowledge representation. Lexicology is one of the disciplines that plays a major role in bridging the gap between these two often diverging views. Most linguistic models nowadays are trying to go beyond formal manipulation of linguistic data to explain how language acts as a substrate of the mental processes involved in knowledge and the most suitable tool for the representation of that knowledge. This idea has been central to functional grammar from the outset with the incorporation of features such as functional logic. The present contribution looks upon the lexicon as a database that plays a major role in a computational model of the natural language user (C.M.N.L.U.) (cf. Connolly and Dik, 1989) and reflects part of the work carried out on a Functional Grammar lexicon of English verbs that is being developed in Cordoba and that was inspired by Martín Mingorance´s Functional Lexematic approach. We will look at lexical knowledge representation from a functional grammar point of view, suggesting how a field-like structure can represent that knowledge and optimise information management processes.

The lexicon in FG.

The lexicon is a central component of FG. Dik (1978a) conceives it as a multi-purpose database in which the basic predicate frames and basic terms of the language are contained. The Fund contáins non-basic predicates and terms. Each lexical entry in the lexicon consists of a predicate structure which allows predicate formation and expression rules to operate in the production and interpretation of expressions. The FG model as envisaged by Dik (Dik 1978a, 1989) places a great burden on predicate formation rules and in fact predicate formation has been given much attention in the FG model. This importance of the rule component tends to keep the information contained in the lexicon to a minimum. In contrast with this view, developments in the framework of FG, especially concerning computational implementations of different aspects of the model tend to minimise rule-driven operations for the generation of derived predicates in favour of storing them directly in the lexicon. Weigand (1986), Junger (1987) take this view and include a varying number of derived predicates ready-made in the lexicon for simplicity. Kahrel (1989) proposes a flexible lexicon of the language user which is more comprehensive than the theoretical lexicon of the linguistic model: He includes derived predicates and converts the formation rules into a relational component whose main task is to account for semantic relations between entries. The inclusion of some derived predicates in the lexicon is not only advised by computational simplicity but also by psychological evidence (cf. Meijs, 1985 and Bybee, 1985).

According to Dik only content words (nouns, adjectives and verbs) are included in the lexicon, adverbs would be derived from their corresponding adjectival forms. Grammar words are accounted for in other components of the FG model. Some computational implementations, however, propose the inclusion in the lexicon of grammar words. Fromkin (1987) proposes a lexicon which is sub-divided into different sub-lexicons containing phonological, grammatical and lexical information.

The structure of a lexical entry.

Following Dik´s knowledge typology (1987) lexical knowledge is a special kind of long-term knowledge which together with grammatical and pragmatic knowledge constitutes our linguistic knowledge.

According to Dik´s proposal (cf. Dik, 1980: 6) the information a lexical entry should contain is the following:

a) The form of the predicate.
b) Its syntactic category (verb, noun, adjective)
c) Its number of arguments.
e) Selection restrictions imposed on each argument.
f) A definition of its meaning.

A typical example of a lexical entry would be:

$$\textbf{kiss}_V\ (x_1\text{: <human>}(x_1))_{Ag}\ (x_2\text{: <concrete>}(x_2))_{Go}$$
$$=_{df}$$
$$\textbf{touch}\ (x_1\text{: <human>}(x_1))_{Ag}\ (x_2\text{: <concrete>}(x_2))_{Go}$$
$$(d2x_3\text{: lip}_N(x_3))_{Instr}$$
$$(dl_4\text{: [express}_V\ (x_1)_{Ag}$$
$$(x_5\text{: affection}_N\ (x_5)_P(x_2)_{Ref})_{Go}]\ (x_4)_{Purp}$$

The first line of the entry contains a predicate frame which provides information on the form of the predicate, its syntactic category and its argument structure. The definition gives semantic information about the meaning of the entry in a dictionary fashion:

> $Kiss_{V,\ def}$=To touch someone with the lips in order to express affection.

From a computational point of view the information contained in the lexicon can be used in different natural language processing tasks. So, predicate frames can be used for the analysis and generation of sentences following a procedure consisting in identifying the verbal element in the string and reconstructing the terms that meet the selection restrictions that the verbal entry imposes (see Kwee Tjoe Liong, 1989, and Dignum, 1989, for a discussion of computational alternatives in parsing within the FG framework).

The inclusion in the lexicon of meaning definitions, together with the fact that only lexical entries of the object language can be used in these definitions following the principle of stepwise lexical decomposition (cf. Dik, 1978) make the FG lexicon especially suitable from the point of view of its psychological adequacy and equate the concepts of lexicon and dictionary. Also, from a knowledge representation point of view the lexicon is the main source of input for inference mechanisms and the crucial element in the implementation of functional logic. Dik identifies logical form with FG grammatical form: in other words logical properties and logical mechanisms can be specified using FG underlying clause structures. Thus FG would act as a means of integrating grammar, logic and cognition in a model of the natural language user.

How does FG represent the meaning of lexical entries? To answer this question it is necessary to enquire into the principle of SLD and see some of its computational implications. The principle of SLD establishes that:

> The definiens of a meaning definition defining some predicate *Pi* may not contain a proper subconfiguration of predicates such that this subconfiguration in itself constitutes the definiens of a meaning definition

defining some predicate ϕj.

(Dik, 1978: 24)

This assumption has several advantages from a theoretical point of view: In the first place, the use of abstract predicates for the specification of meaning is avoided, thus meeting one of FG requirements; secondly, any lexical item appearing at surface level can be represented by its underlying definiens.

It seems that a full application of the principle of SLD would inescapably lead to circularity, since all words are defined by means of other words of the object language. This is in fact what happens in normal dictionary practice where circularity is unavoidable. In order to overcome this difficulty, Dik asserts:

> In every language there is a set of semantically simple lexical items the meaning of which cannot be defined by means of meaning definitions.
>
> (Dik, 1978: 31)

This conception implies a pyramidal structure of the lexicon: with a number of basic predicates acting as the seeds of the meaning definitions of all other lexical elements in the language. From a theoretical point of view the use of defining predicates leads to two confronting alternatives:

a) To establish a reduced set of undefined lexical elements to act as basic.

b) To establish some abstract defining vocabulary. The elements of this vocabulary would constitute a theoretical defining metalanguage.

Dik states that the second of these alternatives does not in effect solve the problem of the definition of lexical items:

> In this respect, a theory which assumes that defining predicates are lexical items of the object language is in precisely the same position as a theory which assumes that they are drawn from some theoretical metalanguage. In the latter case, too, there must be a set of undefined defining predicates.
>
> (Dik, 1978, 31)

It is outside the scope of our work to explore the theoretical consequences of these alternatives. Instead we would like to concentrate on the computational implications of Dik's approach and examine possible solutions to the practical problems the model presents.

This conception has computational implications from two points of view:

a) the mechanics of SLD implementation and

b) the structure of basic (semantically simple) lexical items.

Stepwise Lexical Decomposition.

The principle of SLD implies that for defining any (non basic) lexical item of the language we use other lexical items of the language. The model proposed here is operationally different from other models in which abstract predicates are used in meaning definitions. CDN´s for instance assigns each lexical entry to a frame-like structure in which its meaning and the meanings of all semantically related lexical items are specified (cf. Shanck, 1980). In these systems lexical searches are unique, in the sense that once the appropriate entry has been located no additional searches are necessary for the specification of information. In a SLD lexicon searches are multiplied if we need a full semantic mapping of the lexical entry. The number of searches will depend on the number of predicates appearing on the meaning definition of the required entry plus the number of other predicates eventually found in subsequent meaning definitions. This procedure may result in a quite complicated process for even the simplest searches.

$actor_n$.- $_{def}$= a man who acts a part in a play.
$person_n$.- $_{def}$ = a human being.
act_v.- $_{def}$= to play a part.
$play_v$.- $_{def}$= a piece of writing to be performed in a theatre.
$writing_n$.- $_{def}$= a written work
theatre.-
work....
....

The result of such a search would have the advantage of rendering plenty of information about the lexical entry concerned; from a practical point of view the disadvantages are:

a) It is computationally very costly. Even the simplest sentences would require a very high number of searches and the temporal memory necessary to store the results of those searches would have to be (practically) unlimited.
b) Once the search is launched, it would render information which is not required in the particular application making it difficult to draw the line between relevant and irrelevant information.
c) It would place a tremendous burden on the compilation of the lexicon in order to avoid circularity and redundancy. In applications that require a fairly large lexical database it would make it impossible to

keep track of all possible circularities, which would hinder data integrity and coherence.

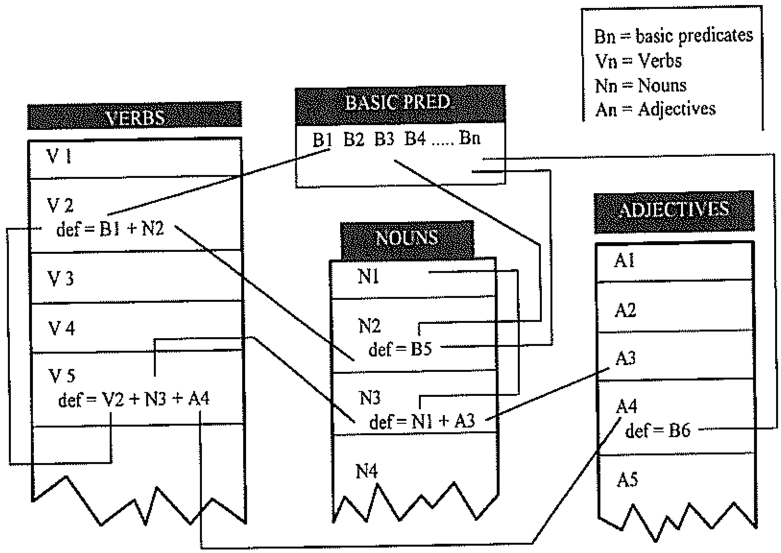

Figure 1: Scheme of a lexical search according to the SLD principle.

To avoid the inconvenience and complications of this retrieval mechanism it would be necessary to incorporate some sort of knowledge representation mechanism such as semantic mappings, or graphs and establish clear search strategies to impose the appropriate constrains on searches so that irrelevant or redundant information may be avoided.

Basic predicates.

Dik makes no claims as to the criteria to establish basic predicates in the lexicon. In fact, none of the implementations of FG up to date makes use of this feature of the model. In ProFGlot, a Prolog implementation of FG, Dik includes meaning definitions in the lexicon but no inference is carried out on those meaning definitions and no distinction is made between basic and non basic predicates (cf Dik, 1991). My own view is that without some kind of knowledge representation mechanism at the top of the lexical pyramid any attempt at representing lexical knowledge would be an empty one. This specification of what we might call the *basic vocabulary of definitions* is a previous step to the inclusion of non-basic vocabulary in the lexicon. Weigand (1990: 101) proposes that this basic module could take the form of a relatively reduced set of primitives which would be the final reference of dictionary definitions. This set of primitives would be made up of a taxonomy derived from a FG perspective consisting in the classification of entities into first order entities (concrete things), second order entities (states of affairs), third order entities (propositions and intensions), and fourth order propositions (utterances).

Lexical fields and FG.

Our proposal is that the incorporation of a field-like structure to the lexicon would help avoid redundant or unnecessary information in most cases. Several authors within the framework of FG have made reference to the possibility of incorporating lexical field criteria to the structure of the lexicon but these references are only tentative and not systematic (cf. Weigand, 1990 and Vossen, 1990). The incorporation of lexical-semantic field criteria to the structure of a FG lexicon is the main contribution of Martin Mingorance´s functional lexematic model:

> «This [Dik´s] grammar, based on a functional communicative view of language, belongs to the synthetic type of grammars, i.e. it has been devised from the encoder's point of view.
> The onomasiological approach seems to be the most appropriate criterion for the organisation of the lexicon in this type of grammar, as the procedure of 'stepwise lexical decomposition' clearly shows (cf. Dik, 1978b: 3.6.).
> Conversely, the application of an analytic model such as Coseriu's Lexematics seems to be, from a practical point of view, the most appropriate procedure for the structuring of the lexicon in semantic fields; in this way it would be able to provide a 'stepwise' description in accordance with the principles of FG."
>
> (Martín Mingorance, 1990: 229-30)

According to Dik´s model "the structure of the definiens of meaning definitions is of the same formal type as the structure of underlying sentences." (Dik, 1978b, 21). Following this principle we could say that a defining predicate consists of a nuclear predication containing an article of the object language whose argument structure provides slots for each one of the arguments and slots for selection restrictions plus a (unspecified) number of satellites whose structure would have to be sufficiently flexible as to be appropriate for the inclusion of the different formal types of satellites. This constitutes quite a complicated structure in which most of the slots would be empty in the great majority of the cases. Our assumption is that a simplified structure consisting of an archilexeme and a specification slot for establishing the difference between the lexical entry and the archilexeme would be more appropriate for the representation of lexical knowledge in the lexicon.

In our model the content of a lexical field is summarised by its archilexeme. Archilexemes contain information about the functional and cognitive structure of the field. Hyponyms inherit the frame structure of the archilexeme and offer certain specifications for the slots provided in the field structure. This not only reduces the amount of unnecessary information but offers the possibility of adding relevant elements about the cognitive structure of the field. Archilexemes would in effect act as the basic words that are not defined by lexical means. The structure of the meaning definitions of lexical entries would be the following:

$$\textbf{lexeme} =_{def} = \textbf{(arch)} + \textbf{(spec)}$$

where arch would be a (basic or non basic) archilexeme and spec is the specification of one of the functional slots of the predicate frame of arch. This element establishes the difference between the lexical entry and its archilexeme. The information in **spec** can refer to one of the argument slots in the nuclear predication of the archilexeme or to a satellite (see examples below). Thus assassinate would have the following definiens:

assassinate$_v$
arch= murder
spec= (important(x_2):politician(x_2))$_{Go}$

According to this the definition for *assassinate* would be the same as for *murder* except for the fact that the slot provided for the second argument in the nuclear structure of murder would in this case be occupied by the specification "*important politician*". This specification acts in much the same way as selection restrictions do.

This structure permits the application of the principles of contrast and opposition characteristic of lexical field conceptions and applies them from a FG perspective.

The use of this structure is consequent with Dik's proposal that the definiens of lexical items belong to the same formal type as the underlying predications of the language: Notice that although the structure proposed above does not correspond to an underlying structure, the result of a full search would produce an underlying structure, since **spec** only acts as a specification of one of the elements in the structure of the basic lexical entry.

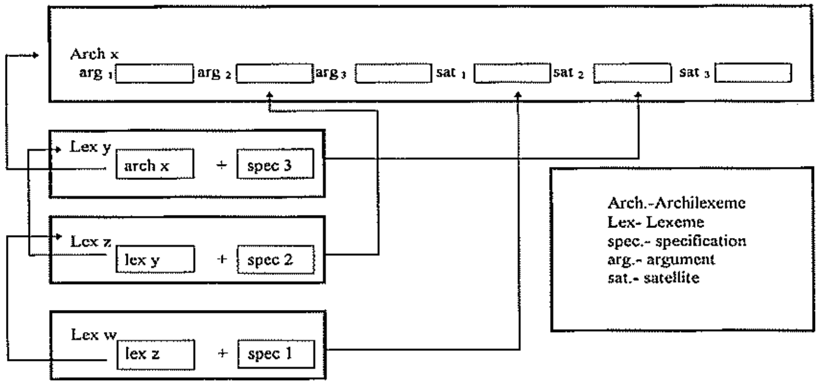

Figure 2: Scheme of a lexical search.

Thus, in the lexical field of physical perception the basic predicates for *perceive* would be:

perceive.
Type: *process.*
(animate(x_1))*Exp*
(material(x_2))*Proc* / (SoA(x_2))*Proc*
(sense(x_3) (bodily (x_3)))Instr
Cond:(proximity (x_1) (x_2))
Cons: (know(x_1) (attribute (x_4)))

This would define the verb *perceive* as a process by which an animate entity (x_1) is in the presence of a certain material object or state of affairs (x_2) and gets to know something (x_3) about (x_2) this process is produced by the instrumental use of one of the bodily senses. Apart from that information we have added the condition that there must be a proximity (visual or otherwise) and the cognitive consequence of perception which is knowledge by the experiencer of a certain attribute of the processed.

Similarly, an action verb such as look at could be defined as:

look at
Type: action
(animate:x_1)Ag

$(concrete{:}x_2)Go$
$((sense{:}x_3)\ (sight{:}\ x_3))Instr$
Purp: $(perceive(x_1)\ (attribute{:}x_4))$

Using this structure for basic entries, hyponims can be defined by reference to that particular frame, for example:

see$_v(x1)\ (x2) = perceive\ (sight(x_3))_{Instr}$
feel$_v\ (x1)\ (x2) = perceive\ (tact(x3))\ _{Instr}$
hear$_v\ (x1)\ (x2)= perceive\ (hearing(x3))\ _{Instr}$
glance$_v\ (x1)\ (x2) = look\ at\ (x1)\ (x2)\ (quick)\ _{Man}$
scan$_v\ (x_1)\ (x_2) = look\ at\ (x_1)\ (x_2)\ (specific(x_4){:}\ information(x_4))_{Proc}$
stare at$_v\ (x_1)\ (x_2) = look\ at\ (x_1)\ (x_2)\ (open\ eyed\ (x_1))\ _{Man}$
glare atv $(x_1)\ (x_2) = stare\ (x_1)\ (x_2)\ (angry\ (x_1))\ _{Man}$
glower at$_v\ (x_1)\ (x_2) = glare\ at\ (x_1)\ (x_2)\ (long)_{Dur}$
contemplate$_v\ (x_1)\ (x_2) = look\ at\ (x_1)\ (x_2)\ (thoughtful\ (x_1))_{Man}$

According to this model basic predicates fulfil three main functions:

a) They act as basic structures providing slots for the insertion of specifications found in the subsequent searches of normal predicates.
b) They also provide a default set for the information not specified in the frames of the normal predicates.
c) They contain additional information which will provide the functional and (some) cognitive framework against which non-basic predicates must be analysed.

It must be noted that in these non-basic lexical entries we make full use of inheritance mechanisms which allows for only the information which is specific to that particular entry to be included, thus avoiding unnecessary duplication and contributing to database integrity. From a computational point of view, this structure would simplify lexical searches while retaining the necessary information for lexical logic to be fully implemented.

We would like to suggest that this structure is psychologically adequate since when we try to characterise the meaning of lexical items we tend to concentrate on assigning it to an appropriate archilexeme and then establishing its differential peculiarities rather than mentally trying to analyse the whole of a general defining structure.

The principles underlined in this article are being applied to a prototype of a lexical database of English verbs from a FG perspective (cf. Muñoz Muñoz, 1994) which was inspired and guided by Prof. Martín Mingorance to whose enthusiasm and insight we are gratefully indebted.

BIBLIOGRAPHY

BYBEE, J.L. 1985 *Morphology; a study of the relation between meaning and form*. Amsterdam: John Benjamins.

CONNOLLY, J.H. and DIK, S. (eds.) 1989. *Functional Grammar and the computer*. Dordrecht: Foris Publications.

DIK, S. 1991. *Functional Grammar in Prolog. An integrated Implementation for English, French and Dutch*. (Prefinal version) Amsterdam: Mouton.

DIK, S. 1978a. *Functional Grammar*. Amsterdam: North-Holland Linguistic Series.

DIK, S. 1978b. *Stepwise Lexical Decomposition*. Lisse: Peter de Ridder Press.

DIK, S.C. 1987. «Linguistically Motivated Knowledge Representation» en Nagoa, M. 1987. *Language and Artificial Intelligence*. Amsterdam: North Holland.

DIK, S.C. 1989. *The Theory of Functional Grammar*. Dordrecht: Foris Publications.

FROMKIN, V.A. 1987. "The lexicon: evidence from acquired dyslexia". Language 63 (1987), 1-22.

GROOT, C. de. 1987. "Predicate formation in F.G". Working Papers in Functional Grammar. n.20.

JUNGER, J. 1987. "Predicate formation in the verbal system of Modern Hebrew". Dordrecht: Foris.

KAHREL, P. 1989. "On the representation of the Lexicon in FG". in Connoly and Dik, 1989: 135-150

MARTIN MINGORANCE, L. 1990. «Functional grammar and lexematics in lexicography». En Tomasczyk, J. y Lewandowska-Tomasczyk (eds.) 1990 *Meaning and Lexicography*. Amsterdam: John Benjamins 227-254.

MEIJS, W. 1985. "Lexical Organization from three different angles". JALLC 6 (1985), 1-10.

MUÑOZ MUÑOZ, J.M. 1994. Criterios para el diseño de un programa de gestión de una base de datos léxica. Tesis Doctoral. Universidad de Córdoba.

WEIGAND, H. 1986 "The Structure of the conceptual language KOTO". Report nr. IR-112. Dpt. of Mathematics and Information Theory, Free University. Amsterdam.