LEXICAL REPRESENTATION: LEVELS AND STRUCTURES

Some important recent studies have concentrated on the connection between lexical semantics and generative syntax. Interest in this field has been motivated by the recognition of the fact that many aspects of syntax of a clause are determined by the semantics of the predicator heading this clause. Research in the properties of the lexicon resulted in a shift from the use of phrase structure rules and subcategorization frames to formulation of appropriate lexical representations.

The aim of this paper is to propose a double-level structure for lexical representations of verbs (that is prototypical predicators). Our proposal incorporates ideas from a number of recent publications: Hale and Keyser [1986, 1987]; Higginbotham [1985]; Kegl and Fellbaum [1988]; Levin and Rappaport [1986]; Rappaport and Levin [1988]; Rappaport, Levin and Laughren [1988]; and Zubizarreta [1985, 1987]. Additionally, the discussion of the semantic properties of lexical entries will serve as an introduction to Conceptual Semantics, as developed in Jackendoff [1983, 1987].

As a first approximation we may assume that a lexical entry of a verb must include:

1. the morphophonological form,
2. specification of the categorial type,
3. information about the organization of the argument(s) selected by the verb,
4. semantics of the verb.

Here, we shall be interested in parts (1.iii) and (1.iv), i.e. syntactic and semantic aspect of the lexical representation. The
semantic aspect will become central to our concern in further parts of this paper, however, in the following discussion we shall assume that the semantic content of each argument is identified by the so-called theta-roles (θ-roles), as introduced in early generative literature [Gruber 1976, Jackendoff 1972] and also used in more recent studies [Chomsky 1981, Williams 1981]. An alternative approach to this issue will be presented later on.

As far as (1.iii) is concerned the lexical representation has to specify the following information about arguments:

(2)  
   i. the number of arguments that a verb takes,
   ii. the distinction between external and internal arguments,
   iii. the distinction between direct and indirect arguments,
   iv. optionality of arguments,
   v. the referential properties of arguments.

The lexical representation specifies the number of arguments the verb takes by mentioning the θ-roles in a form of a list, called the θ-grid:

(3)  
   put: (Agent, Theme, Location)

(4)  
   John put the books on the shelf.

The θ-grid (3) for the verb put specifies that it is a triadic verb, i.e. it takes three arguments bearing the roles of Agent (John), Theme (the books), and Location (on the shelf).

As observed in Williams [1981, 1984] predicates make a distinction between the external argument and the internal one(s). Internal arguments are syntactically realized inside the VP, as sisters to the verb. The external argument, on the other hand, is syntactically identified as realized in the [NP, S] (subject) position in a clause. These arguments are external to the maximal projection of the verb and therefore their θ-roles are assigned by the VP via predication [see Williams 1980 and 1987]. And so in the sentence (4) with the simplified structure (5) the internal θ-roles (Theme and Location) are assigned to NPs internal to the maximal projection of V (this assignment is represented by indexing) whereas the external argument (Agent) is bound by the index of the whole VP ("vertical binding" in Williams 1987), and is assigned its θ-role through the process of predica-
tion [for different accounts of the problem of licensing external arguments see Chomsky 1981, 1986 and Kegel and Fellbaum 1988]:

(5) 

Now we can reformulate the lexical entry (3) as (6), with the external argument outside the brackets:

(6) put: Agent Theme Location

As it is clear from (5) the subject (external argument) is outside the VP and thus the verb does not subcategorize for this position. Also the Projection Principle as stated in Chomsky [1981: 29] requires objects (internal arguments) to be syntactically realized, but says nothing about subjects:

(7) Projection Principle:
Representations at each syntactic level (i.e. LF, and D- and S-structure) are projected from the lexicon, in that they observe the subcategorization properties of lexical items.

The requirement that clauses have subjects has been added to the basic version of the Principle [in Chomsky 1982: 10] giving rise to the Extended Projection Principle (EPP):

(8) EPP:
   i. = (7),
   ii. Every sentence must have a subject.

From the fact that the external argument is lexically un-associated with a syntactic frame it follows that it may be syntactically unrealized, as in for example passives and derived nominals. Compare in this respect (9), where both the external
and internal arguments are present, with (10) and (11) (passive and nomi
nal, respectively) where the external argument is missing from the syntactic realization:

(9) Tom writes letters everyday.

(10) Letters are written everyday.

(11) Letter-writing can be fun.

Apart from the distinction between external and internal arguments its seems necessary to distinguish direct arguments from indirect ones. An NP argument which is assigned its θ-role directly by the verb is the direct argument, whereas any arguments assigned their θ-role by a preposition are called indirect arguments [see Marantz 1984 and Zubizarreta 1985]. This distinction is motivated by the fact that external and direct arguments are semantically unrestricted, and so for example the subject may be assigned other θ-roles than Agents (we underline the direct argument in the θ-grid):

(12) Fred received a cheque from Bob.

receive: Goal  \langle Theme, Source \rangle

The indirect arguments on the other hand, are restricted by appropriate prepositions: the object of to must be the Goal, the object of from must be the Source, the object of of must be the Theme, the object of in must be the Location, etc.

The way arguments of a verb are organized depends on the syntactic structure in which the verb occurs; compare the behaviour of the verb break (expressed by the θ-grid) in transitive (13), passive (14) and unaccusative (15) use:

(13) Sam broke the bottle.

break: Agent  ⟨Patient⟩

(14) The bottle was broken (by Sam).

break: (Agent)  ⟨Patient⟩

(15) The bottle broke.

break:  ⟨Patient⟩
Above sentences suggest that there are two types of θ-role association with syntax. Following the terminology of Safir [1987] we shall say that if a θ-role is actually present in the thematic array of the verb at D-structure it is "projected" and if it is mapped onto a structural position then it is "linked". Obviously, a relevant level of representation will have to capture the common features and aspects of meaning shared by the verb break in sentences (13)-(15); we return to this issue of formulating an appropriate semantic representation later on. From the above description of linking and projecting it follows that a θ-role may be:

(16) a) projected and linked (arguments of transitive verbs);
    b) projected but unlinked (missing agent of passives);
    c) unprojected and unlinked (missing agent of unaccusatives).

Now the Projection Principle for the thematic structure can be interpreted as in (17), see also Safir [1987]:

(17) Thematic structure projected at D-structure must be expressed at every syntactic level.

Furthermore, some θ-roles may be connected with optional arguments of polyadic verbs. However, as noted by Levin and Rappaport [1986], optionality of arguments depends to large extent on the pattern of θ-role assignment, as illustrated by examples (18) and (19) from Levin and Rappaport [1986: 640]:

(18) i. I read the book to Jane.
    read: Agent ⟨Theme, Goal⟩

   ii. I read the book.
     read: Agent ⟨Theme⟩

   iii. I read to Jane.
      read: Agent ⟨Goal⟩

(19) i. I read Jane the book.
    read: Agent ⟨Theme, Goal⟩

   ii. *I read Jane.*
      * read: Agent ⟨Goal⟩

When the Goal role is assigned directly [as in (19)] both arguments are obligatory, on the other pattern of θ-role assignment
both internal arguments are optional. The two θ-role assignment options for \textit{read} are shown in (20), with optional elements in parentheses:

\begin{align*}
\text{(20)} & \quad \begin{array}{l}
\text{A. Agent} \quad \langle \text{Theme}, \ (\text{Goal}) \rangle \\
\text{read:} \\
\text{B. Agent} \quad \langle \text{Theme, Goal} \rangle
\end{array} \\
\end{align*}

As postulated in (2.v) the lexical representation must show coreferentiality between the internal and external arguments of inherent reflexives [Zubizarreta 1985: 252], as demonstrated by (21):

\begin{align*}
\text{(21)} & \quad \begin{array}{l}
i. \text{John behaved.} \\
ii. \text{John behaved himself.} \\
iii. * \text{John behaved Mary.}
\end{array} \\
\end{align*}

\text{behave: Agent}_1 \quad \langle (\text{Patient}_1) \rangle

So far we have been using annotated θ-role labels in our θ-grids, however, as observed by Jackendoff [1983, 1987] these roles are not disembodied entities but derive from the position of an argument within an explicit semantic structure. It is also worth mentioning that since the specific content of such notions as Agent, Theme, etc., though relevant in some way for the ultimate semantic representation, is not relevant for purposes of sentence grammar it is advantageous to postulate that syntactically relevant lexical structures do not contain θ-role labels, they do however, preserve all information previously supplied by annotations.

Following Rappaport, Levin, Laughren [1988], Rappaport and Levin [1988] and Hale and Keyser [1986, 1987], we introduce the level of Predicate Argument Structure (PAS) as the syntactically relevant level of lexical structure. PAS embodies the categorial projection of the category \textit{V} to a phrasal node \textit{VP} and the organization of the arguments of the verb. The arguments of the verb are represented by variables; following an established convention we use "x" to indicate the external argument, "y" to indicate the direct argument and "z" to indicate the indirect argument. We also use "w" to indicate obligatory adjuncts (secondary arguments) and subscripts to express referentiality. Below we give a list of various configurations
for Predicate Argument Structures, together with names of constructions and example sentences [based on Kege1 and Fellowbaum 1988]:

(22) a) \( x \langle y \rangle \) transitive
The boy hit the dog.

b) \( x \) unergative
The woman cried.

c) \( \langle y \rangle \) unaccusative
The bottle broke.
middle
Ford Fiesta sells very well.
passive
The ball was hit.

d) \( x \langle y Pz \rangle \) ditransitive
The man gave the book to the woman.

e) \( x \langle y z \rangle \) double object
The man gave the woman the book.

f) \( \langle y z \rangle \) double object passive
The woman was given the book.
psych-verb
The play amused the children.

\( g) \emptyset \) weather verb
It rained.

h) \( \langle y w \rangle \) measure verb
The chicken weighed 3 lbs.
copular
Mary is a teacher

(The symbol \( \emptyset \) is used to indicate an expletive it.)

Only in case of idiomatic expressions there occurs a constant in the PAS of a verb [Zubizarreta 1985]:

(23) i. kick - non-idiomatic: \( x \langle y \rangle \)
idiomatic: \( x \langle \text{the bucket} \rangle \) (to kick the bucket)

ii. lose: \( x_i \langle y_1 \text{s way} \rangle \) (to lose one's way)

One of the central constraints on the relation between syntactic structures and thematic/semantic structures is the Theta-Criterion [Chomsky 1981: 36]:
Theta-Criterion:
Each argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument.

By this criterion each variable in the PAS of a verb must be filled or "saturated" [Higginbotham 1985] by corresponding to an appropriate NP in D-structure (and by the EPP, in syntax in general). This process of saturation establishes the association between NPs occupying the argument positions of a verb in the syntax and the variables in the PAS, giving rise to a structure schematically represented in (25):

(25) verb: $x_i^j y_j^k z_k^l$, PAS

Apart from the EPP and the θ-criterion the syntactic representation must also meet another well-formedness condition - the Case Filter [see Chomsky 1981]:

Case Filter:
Lexical (phonetically realized) NPs must be assigned Case.

Case assigning features are not specified in lexical entries and Case is assigned according to the following rules (rules in (27) are for English):

(27) i. NP governed by INFL [+tense] is assigned Nominative Case
    ii. NP governed by VP is assigned Accusative Case.
    iii. NP governed by PP is assigned Oblique Case.

See however Márcz [1986] for arguments that in nonconfigurational languages - such as ex. Hungarian, Basque, Warlpiri, etc. - the lexical entry of a verb contains the Case frame.

As it has already been observed the PAS gives the syntactically relevant representations and contains variables. These variables are related to a more fully specified structure - the Lexical Conceptual Structure (LCS). LCS is a lexical semantic representation which encodes certain aspects of a verb's meaning through predicate decomposition. The background assumption here is that at some level of representation the meanings of verbs are not unanalyzable entities. The introduction of the level of LCS is motivated by a number
of phenomena, among them the behaviour of verbs in transitive alternations [Guerssel et. al 1985, Hale and Keyser 1986], the locative alternation [Rappaport and Levin 1988], and also by the inquiry into general properties and principles of linguistic theory [Jackendoff 1983, 1987].

LCS is a representation of the concept named by the verb and the participants in the action (represented by variables) and it is composed from a universal set of primitive functions.

Before we proceed it is necessary to remark that our knowledge of universal semantic primitives is still very limited and therefore the following discussion serves only as an introduction to this very complex issue, also our descriptions of LCSs are often not complete and rather simplified.

As noted above, the level of LCS is motivated by a number of phenomena connected with the behaviour of verbs under different alternations. The verb in (28) displays the causative/inchoative alternation, whereas the verb in (29) has the causative variant only [examples based on Guerssel et. al 1985]:

(28) a) Jane broke the cup.
    b) The cup broke.

(29) a) Jane cut the bread.
    b) *The bread cut.

The appropriate Predicate Argument Structures are given in (30) and (31), respectively:

(30) break: x \<y\> (causative)
     \<y\> (inchoative)

(31) cut: x \<y\>
      *\<y\>

The above PASs do not explain the difference in behaviour of break and cut. The explanation, however, can be found in semantics, i.e. in fundamentally different LCSs: (32) versus (33):

(32) a) BREAK: [x cause [y become BROKEN]] (causative)
    b) BREAK: [y become BROKEN] (inchoative)

(33) CUT: [x produce CUT in y]
In (32) cause and become are two potential primitives, BROKEN is a constant representing that element of meaning that sets the state being broken apart from other states (or rather STATES). In (33) produce is another candidate for a primitive, CUT a constant. In both cases (i.e. both predicates and constants) the elements need further refinement, however, even these crude representations enable a coherent analysis of the inchoative/causative alternation. Thus, in case of break the basic LCS is the monadic (32b) and to render the causative reading this LCS is embedded into a general dyadic LCS of the form (34):

(34) CAUSE: [x cause [y become in STATE]]

That is, the basic LCS is embedded as the complement of a dyadic causative predicator - CAUSE - and gives ultimately the complex LCS (32a). The option of embedding does not apply for cut, as it does not have an appropriate LCS. For elaboration of this view, and a discussion of middle and unaccusative constructions, see Hale and Keysters [1986, 1987].

Another motivation for LCS comes from the analysis of the locative alternation. Verbs showing this alternation (35) have traditionally been associated with a θ-grid like (36), and caused numerous problems for an appropriate interpretation within such grids [example and discussion based on Rappaport and Levin 1988]:

(35) a) Tim loaded hay onto the truck.
    b) Tim loaded the truck with hay.

(36) Load: Agent (Theme, Goal)

The θ-grid (36) is a mere list of θ-roles and it does not account for the presence of the two variants associated with two distinct PASs (the locative variant and the with variant, respectively):

(37) a) load: x (y, Ploc^z)
    b) load: x (y, Pwith^z)

Now, in order to relate the representations in (36) and (37) complicated linking rules ought to be devised. However, such rules, though complicated and often idiosyncratic, do not account for the so-called affected interpretation of the with variant (35b). To make explicit all these relevant aspects of meaning, Rappaport
and Levin employ the possibility of using definitions with variables and predicate decomposition. The following LCS are given for the two variants of *load*:

(38) LOAD: \([x \text{ cause } [y \text{ become at } z]]\)

(39) LOAD: \([x \text{ cause } [z \text{ become in STATE}]\)
    \quad \text{BY MEANS OF } [x \text{ cause } [y \text{ become at } z]]\]

The representation in (38) indicates that *load* names an event which involves a change of location. The representation in (39) indicates that *load* names an event in which a change of state is brought about by means of a change of location. The use of the same set of variables \((x, z)\) in both clauses of the definition in (39) identifies the participants as the same. Obviously, a more elaborate representation would require a full specification of the manner component in order to set *load* apart from the other locative alternation verbs.

The introduction of two levels of lexical representation (PAS and LCS) has proved to give correct results in describing a number of constructions across different languages: causative, inchoative, and middle in English, Berber, Warlpiri, Winnebago [Guerssel et al 1985]; ergative and passive in Norwegian [Áfarli 1987]; transitivity in Hungarian, Basque, and Warlpiri [Marczenko 1986]. LCS is, however, motivated also by the inquiry into general properties and principles of linguistic theory.

According to Jackendoff [1983, 1987] thematic - i.e. semantic or conceptual - structures are autonomous structures with their own primitives and principles of combination and organization into subcomponents. Innate formation rules include "a vocabulary of primitive conceptual categories or 'semantic parts of speech'" [Jackendoff 1987: 375]. The LCS of a verb is composed from a universal set of primitive functions such as CAUSE, GO, STAY, and BE. CAUSE, GO, STAY define EVENTS; BE defines STATES. Each of these functions, together with its arguments, forms a conceptual clause. Arguments PLACE and PATH can be further decomposed into a function (AT, IN, ON) and another argument. Basic categories are expanded by formation rules of following type into more complex expressions [Jackendoff 1987: 375]:

(40) a) PLACE → \([\text{Place PLACE-FUNCTION (THING)}]\)
Rule (40a) expands the basic category PLACE into a PLACE-FUNCTION and argument of the function which is itself of the category THING. In the Jackendoffian model of Conceptual Semantics the argument of the function PLACE serves as a spatial reference point, and so for ex. [all examples from Jackendoff 1987] in the phrase under the table, the table designates a reference object, whereas under (the PLACE-FUNCTION) defines a region. Rule (40b) expands the basic category PATH into one of the five functions that map a reference object (THING or PLACE) into a related trajectory. An example of a PATH with a reference THING is the phrase to the house; a PATH with a reference PLACE is from under the table with the PLACE category (under the table) embedded as an argument of PATH.

From (40c) it follows that there are two EVENT functions: GO and STAY, each of which expands the category EVENT and each takes two arguments. The arguments of GO are the THING in motion and the PATH it traverses, as in Bill went to New York. In contrast with GO, which denotes motion, STAY denotes stasis over a period of time. The arguments of this function are the THING (standing still) and its location, i.e. PLACE, as in Bill stayed in the kitchen (40.d) expands the basic category STATE into two functions: BE and ORIENT. BE specifies the location of objects as in the dog is in the park; ORIENT specifies the orientation of objects - the sign points toward New York.
A number of basic correspondence rules and subsidiary principles (some of them language specific) govern the relationship between conceptual constituents and syntactic ones. And so a sentence (in unmarked case) can express an EVENT or a STATE, noun phrases can express almost any conceptual category, a prepositional phrase can express a PLACE or a PATH, etc. This relation between the syntactic (41a) and conceptual (41b) constituent structure can be best illustrated with a single example [Jackendoff 1987: 376]:

(41) a. \[S [NP^John] [VP ran [PP into [NP^the room]]]]

b. Event GO ([Thing John, [Path TO (Place IN ([Thing room]))]])

In (41) we have the following correspondences:

(42) Syntactic Structure  Conceptual Structure
S  Event
VP  Event-function GO
NP^i  first argument of GO (Thing)
PP  second argument of GO (Path)
NP^j  argument of Path (Thing)

The second argument of the Event-function GO, i.e. Path is composite: the Path-function TO takes a Place as its argument, Place in turn decomposes into the Place-function IN and a Thing argument.

Lexical entries include phonological information, categorial information, a subcategorization frame and a conceptual structure, with appropriate coindeixation between the last two, as in entries for run and into, respectively [Jackendoff 1987: 376]:

(43) run: \[[-N, +V]

[—— (PP^j)]

[Event GO ([Thing ]^i, [Path ]^j)]

(44) into: \[[-N, -V]

[—— NP^j]

[Path TO ([Place IN ([Thing ]^j)])]
The verb *run* has a simple PAS (45) in which only the external argument (the subject) is present:

(45) run: x

However, in the Jackendoffian framework of Conceptual Semantics the verb *run* requires two arguments: the Thing in motion and the Path specifying the trajectory of this motion. The Thing is indexed "i" in (43), which in Jackendoff's convention indicates the subject position. The second argument, on the other hand, is filled in by the postverbal PP, with which it is coindexed in the subcategorization frame ("j"). If no PP is syntactically present, the Path is unspecified. The important point is that well-formedness conditions on conceptual structure require this argument to be present conceptually even if it is not expressed syntactically. The case of the argument-taking preposition *into* is simpler: it requires an NP object, and this object is coindexed with the argument position in conceptual structure.

In the derivation entries (43) and (44) - and the lexical entries for the relevant nouns - are concatenated in a particular expression and the conceptual structures of these items are subject to a process of Argument Fusion [Jackendoff 1987: 386]:

(46) Argument Fusion:
Into each indexed constituent in the reading of the verb or proposition, fuse the reading of the syntactic constituent in the sentence that satisfies the coindexed position in the verb's subcategorization feature. Into the position indexed i in the reading of the verb, fuse the reading of the subject.

Thus the conceptual structures in the lexical entries of *run* and *into* are combined with each other and with noun phrases to give the fused conceptual structure (41b).

The conceptual structures allow an explicit and natural representation of the selectional restrictions which particular lexical items impose on their complements. The same structures provide a basis for the rules of inference in natural language. Moreover, the structures allow an explicit account of the similarities and differences in the meanings of various words and groups and classes of words and they relate these meanings to the representations generated by other (non-linguistic) cognitive systems.

Jackendoff introduces in his lexical entries (such as (43) or
subcategorization frames, however, it seems plausible to substitute argument structures (i.e. PASs) for frames - this move, though, would require an elaboration of linking rules. For some recent proposals see Rappaport and Levin [1988] and Rappaport, Levin and Laughren [1988]. Though this issue is crucial for an adequate theory of lexical representation we will at present refrain from discussing it in any detail.

Apart from a brief discussion of the Jackendoffian model of Conceptual Semantics, we have in this paper presented two levels of lexical structure: Predicate Argument Structure and Lexical Conceptual Structure, eliminating thus θ-roles understood as primitives of semantic theory. The PAS, which is an abstract syntactic projection of the category verb, has variables identified in the LCS. LCS defines the meaning of the verb through predicate decomposition.

In this model θ-relations become reduced to structural configurations in conceptual structure and θ-roles, as derived notions, are defined over LCSs. The content of a θ-role (and hence of the θ-grid) is a compositional function, and may therefore vary according to the particular pattern of lexicalization, explicit in the LCS.

REFERENCES


Piotr Stalmaszczuk

STRUKTURA I POZIOMY HASEŁ SŁOWNIKOWYCH

Celem artykułu jest zaproponowanie dwupoziomowych hasel słownikowych w obrębie semantyki leksykalnej uprawianej w ramach gramatyki generatywnej Chomskiego. Przedmiotem zainteresowania są hasła słownikowe dla czasowników - proponuje się dwa poziomy struktury leksykalnej: strukturę predykatowo-argumentową (PAS), oraz strukturę leksykalno-pojęciową (Lexical Conceptual Structure - LCS). PAS stanowi modelową projekcję składniowych właściwości czasownika, natomiast LCS określa semantykę czasownika poprzez definicje tzw. rozkład predykatów.

Dodatkowym celem artykułu jest zarysowanie modelu semantyki zaproponowanego przez R. Jackendoffa w pracach z lat siedemdziesiątych i osiemdziesiątych.