DEPENDENCY SYNTAX IN A FORMAL DESCRIPTION

Eva Hajicová, Charles University, Prague
e-mail: hajicova@ufal.mff.cuni.cz

I. INTRODUCTORY REMARKS:

Necessary condition: to combine requirements of formal methodology with empirical findings of classical research.

Prototypical expression of grammatical values - morphemes (endings, affixes, function words); word order (with intonation) expresses first of all the topic-focus articulation, also in English.
"Free" word order corroborates the use of dependency syntax, present in other frameworks - Lexical Functional Grammar, Head Driven Phrase Structure, X-bar theory, theta roles: dependency is necessary, constituency perhaps is not. Interactive nature of language - topic-focus articulation (TFA) as an aspect of sentence structure.

Classical postulates of synchronic linguistics:
(A) to distinguish language as a system
  (i) from its use in communication (*langue* vs. *parole*, linguistic competence vs. performance);
    communicative (pragmatic) competence;
  (ii) from cognitive patternings of content;
(B) to acknowledge the word as a basic unit (to store much of grammatical information in the lexicon) and
  sentence as the largest complex unit of the system of language;
(C) to establish (underlying) syntax as the core of linguistic description ("Cartesian Linguistics").

Prague Linguistic Circle - V. Mathesius, R. Jakobson, B. Havránek, B. Trnka, V. Skali čka

Functional Generative Description (FGD): interface level of the Saussurean 'form of content' ('linguistic meaning'), 'tectogrammatics' - parenthesized strings, complex symbols:

(a) the dependency structure of the sentence (basicaly a projective tree); kinds of the dependency relation – edge labels; linearized: subscripts at the parentheses;
(b) coordination - more dimensions, but (thanks to projectivity) linearization still is possible;
(c) topic-focus articulation (Hajicová): contextual boundness, communicative dynamism; represented by the left-to-right order of nodes and by indices;
(d) values of morphological categories (conveyed by function words, affixes, alternations, word order) - rendered by indices within node labels, so that the representations of sentences do not display too many nodes.

Non-terminals - only during the derivation, not in the (terminal) representations. The transition to morphemics (strings without parentheses) includes movement rules, covering differences between deep and surface word order

II. DEPENDENCY SYNTAX

L. Tesnière: the verb with its valency (*actants* and *circonstants*, i.e. inner participants and circumstantials, or arguments and adjuncts,...) as the core of the sentence;

subject as one of the participants: *premier actant*

Formalisation of dependency syntax:

dependency tree with complex symbols:

arrive.Pret.Decl

Actor Direct.2 Temporal.when

brother.Sing.Def *there* *yesterday*

Appurt Gener

I *young*.Compar

complex symbols (lexical units, morphological grammatemes, syntactic reations);

linearization of the tree (subscripts point to the head):

$((my)_{Apurt}\ (younger)_{Gener}\ brother)_{Act}\ arrived\ (_{Dir.2}\ there)\ (_{Temp}yesterday)$

relational notions - primitive, not derived from 'structural' ones (from constituency); subscripts as terminal symbols.

VALENCY AS A BASIS FOR FORMAL DESCRIPTION

A. Classifying the valency slots:

criteria for classifying the data on valency:

(i) inner participants vs. circumstantials:

bound to certain verbs, occurring just once

objects: direct (Objective, Patient) - *to build* (*destroy*) a house; *to see a house; to address someone; to elect the chairman; to choose a spokesman;*

indirect (Addressee) - *to give Mary a book*;

second - *to do sth. as chairman; to elect so. the chairman; to choose him as chairman*;

subject (Actor/Bearer) - ?

(ii) obligatory vs. optional with a given head:

participants - prototypically obligatory: *Jim met Eve.*

- possibly optional, e.g. *to read a book* (*to so.*)

circumstantials - prototypically optional, e.g.*to be*

sitting (*somewhere*) (*for a reason*) (*for some time*);

- possibly obligatory: *to behave somehow; to last for some time; to arrive at a place;*

obligatory, although deletable (i.e. present in the underlying structure): 'dialogue test':

A: *Jerry arrives tomorrow.*

B: *Where to?* (A's assumption that the place is known to the hearer is not valid)

A: **I don't know.*

This answer would disqualify the speaker, but it is acceptable after *In which way?, Why?*

(iii) Different kinds of complementations (valency slots, dependency relations):

if meanings differ, speaker is able to distinguish:

John rolled down the hill and broke the window.

speaker may not know if this was an intentional action; thus Agentive and Experiencer (Theme) differ only as cognitive roles; in the system of language only Actor/Bearer is present: vagueness, underspecification;

if impossible to combine in coordination, the slots differ:

**Jim said this three times and yesterday. Temp.2, Temp.1*

**We came home and yesterday. Direction, Temp.1*

**He bought a car from Jim and twice. Origin, Temp.2*

B. Repertoire of complementations:

(a) dependent mainly on verbs:

inner participants (arguments): Actor, Addressee, Objective, Origin, Effect

She changed her hair-do from a braid into a fringe.

Act Obj Or Eff

free modifications (circumstantials, adverbials, adjuncts):

Temporal - when, how many times, since when, till when, how long, for how long

Manner

Regard

Extent (*He spent his money to the last penny.*)

Norm (*in accordance with*)

Criterion (*according to*)

Substitution (*instead of*)

Accompaniment (*with someone*)
Means (Instrument)
Difference (*two inches taller*)
Benefit (*for someone*)
Comparison (*as bright as something; brighter than sth.*)
Locative
Directionals - 1.from where, 2.which way, 3.where to
Condition
Cause
Aim (*in order to, for the sake of*))
Concession (*although*)
Result (*so that*)

(b) dependent mainly <u>on nouns</u>:

arguments: Material (Partitive, *two baskets of sth.*)
Identity (*the river Don, the notion of operator*)
adjuncts (free modifications):
Appurtenance (my table, Jim's brother, Mary's car)
Restrictive (rich man)
Descriptive (the Swedes, who are a Scandinavian nation)
combinations with semantic features:
Loc: *in, on, under, between,...*; Accomp or Regard: +/-;

necessity of a more subtle classification (e.g. Manner:
quickly - like a lightning; Loc.where: *she got hurt in the kitchen - on her shoulder).*

C. Examples of simplified valency frames:
optional adjuncts (see above) not included

read V Act1 Addr Obj
change V Act1 Obj1 Or Eff1
give V Act1 Addr1 Obj1
rain V
brother N Appurt1 superscript 1 - obligatory;
man N
glass N Material
full A Material1
green A

asymmetry: shifting (inferencing necessary, context)

Addr
Act <-- Obj <-- Orig
Eff

(1) The book appeared. Obj --> Act
(2) She prepared soup from potatoes. Eff --> Obj
(3) He abandoned his family. Orig --> Obj
(4) Jim addressed the whole family. Addr --> Obj
(5) John built a house. Act - Agentive
(6) John got a letter. Act - Experiencer
(7) John broke a window. Act - ?

III. SYNTAX AND SEMANTICS:

each level - its own syntax; syntax, semantics and pragmatics are not three levels of the system of language;

'underlying' syntax - Saussurean meaning - a level within the system of language, differs from the non-linguistic cognitive content; cf. Chomskyan Logical Form;
'meaning' - a presystemic notion requiring a series of explicata:

(i) underlying structure, tectogrammatics - level of linguistic patterning or 'form' of cognitive content, with F. de Saussure, L. Hjelmslev, E. Coseriu; ambiguity of an expression: *light* vs. *dark* and vs. *heavy*, vagueness or indistinctness: now (time point or interval), *desk/table*, *book* (text or physical object);

(ii) semantic(-pragmatic) interpretation:
intension (truth conditions), extension (truth values):

Formal semantics: R. Montague, B. Partee, D. Lewis; contextual knowledge in semantic interpretation:
I. Heim: file change semantics, context change potential;
H. Kamp: discourse representation theory;
sentence as an operation on the hearer's memory state;

My neighbour is a widow.
different extensions in different contexts (possible worlds or situations, reference assignment); sentence vs. utterance;

meaning (i) - just another language; however: disambiguated; still a linguistic structure (language specific; no logical devices such as overt variables, operators, parenteheses and prenex quantifiers denoting scopes); relevant for a division of labor; interface between the study of language systems and the interdisciplinary domains of formal semantics, cognitive linguistics, communication (discourse):

reference assignment (based on inferencing using
contextual and other knowledge; salience (Hajicová);
figurative meanings,
discourse patterns (typical frames, scenarios), and so on.

IV. SYNTAX AND MORPHEMICS:

not every word is a syntactic unit; **function words** are syntactically fixed to words - prepositions to nouns, conjunctions and auxiliaries to verbs; it is not adequate (economical) to represent them by nodes;

as far as they are immediately semantically relevant, they are indicated by indices of node labels (grammatemes, values of tense, aspect, modalities, number, definiteness); cf. analytic verb forms, prepositions vs. case endings;

as far as they are relevant just for grammatical agreement, they concern only the relationship between syntax (tree) and morphemics (string): gender, number, case with adjectives, person, number with verbs.

V. DEPENDENCY AND COORDINATION

coordination (and apposition) - another dimension, cannot be rendered by a planary graph (tree);
coordination of sentences and of their parts, i.e. roots or other nodes in the trees are coordinated:

(1) Jim went to Boston and Mary to Chicago. (deletion)
(2) Jim and Mary went to Boston.
(a) Jim (went to Boston) and Mary went to Boston.
(b) Jim and/with Mary went to Boston.
(3) Jim and Mary are a nice pair. (just (b): group reading)
(4) Jim and Mary or Mike and Jane played against Bill and Martha.
(5) Jim and Mary came there together.
(6) Jim and Mary met (each other).

Kinds of coordination: conjunction, disjunction (strings of unlimited length), adversative (pairs).

Apposition: another kind of a similar relation (index at rightside bracket), or an underlying predication (deletion):
Mary and John, (who are) our neighbours,...

VI. THREE LAYERS OF WORD ORDER

Tectogrammatics - topic-focus articulation:

systemic ordering of valency slots (kinds of dependency relation): in grammar (specified in valency frames);

communicative dynamism: in the TRs (underlying word order);

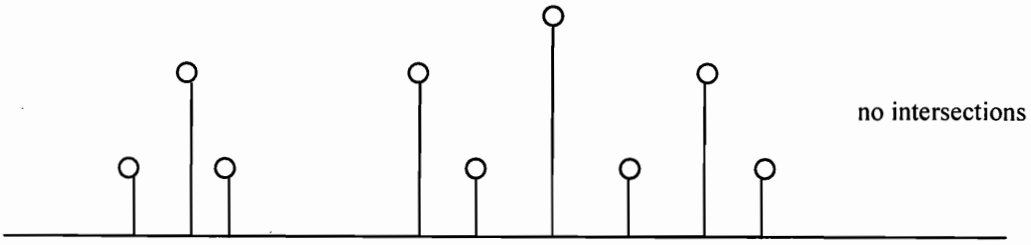
surface word order: on the level of morphemics:

(7) Jim visited me, since he wanted to ask me for advice.

subordinating conjunction in front of the embedded clause,
preposition in front of the noun group;

Projectivity: a dependency tree (accounting both for the dependency relations between heads and their modifiers, and for the left-to-right order) is projective if for every set of nodes a, b, c present in the tree, it holds that if a depends on c, and b is placed between a and c in the left-to-right order, then b immediately or through mediation of other nodes depends on c.

condition of projectivity:



Similar restrictions hold between coordination and the basic two dimensions; thus, one-to-one **linearization** is possible:

(8) Mary and John, our neighbors, who are a nice pair, arrive.

(8') (((Mary John) $_{Conj}$ (we) $_{Appurt}$ neighbor.Plur.Def) $_{Appos}$
($_{Gener}$ (Rel) $_{Actor}$ be.Indic.Pres ($_{Obj}$ (nice) $_{Gener}$

pair.Sing.Indef)))$_{Actor}$ (here)$_{Dir.2}$ arrive.Indic.Pres

non-projectivity (exceptions, strongly limited) can be described by movement rules concerning morphemics:

(9) I met a man yesterday, who asked me for your address.
underlying: I - yesterday - met - a man, who...

(10) a larger town than Boston: larger - than B. - town

VII. LEXICAL AND GRAMMATICAL INFORMATION

much of grammatical information - in the lexicon (especially in the valency frames or grids);
word - a central unit of the language system;

the **lexical entry** contains the following parts:

(a) the lexical meaning; with ambiguity there are several representations, i.e. several lexical units;

(b) the slots for values of relevant grammatical categories, i.e. of grammatemes; e.g. gender with nouns is specified here, but number and definiteness with nouns, or tense, aspect, modalities with verbs, degrees of comparison with adjectives, get their values only for individual occurrences of the word forms in discourses);

(c) the valency frame: possible complementations, ordered in accordance with systemic ordering (see **Haji čová**; inner participants (occurring at most once with a head node) and obligatory modifications are indicated (arrive at a place, behave somehow, last how long,...); optional free modifications - lists for word classes;

(d) obligatoriness vs. optionality; operational criteria - see **Panevová**;

(e) surface deletability: e.g. Directional.2 with *to arrive* is deletable, Objective with *to meet* is not;

(f) optional or obligatory controller (e.g. Actor - obligatory with *to try*, optional with *to decide*; Addressee - optional with to advise, to forbid);

(g) ability to occupy certain syntactic positions (e.g. of Subject with Passivization, of a *wh*-element) or of barriers for movement;

(h) subcategorization conditions (e.g. Obj possibly having the shape of a verb clause.

Lexical entries can share their lexical part proper, differing in their frames: *to swarm* has an obligatory Means in

The garden swarms with bees,

but just an optional one in *Bees swarm in the garden*,

to load has an obligatory, though deletable Means in

They loaded the truck with hay.

and an obligatory Dir.2 in *They loaded hay on the truck*

VIII. SPECIFICATION OF SYNTACTIC (UNDERLYING) STRUCTURES

The class of syntactic representations can be specified by means of a small number of general principles describing the core of grammar and by specific rules for secondary (marked, marginal) patterns.

(A) Generative procedure:

(i) generate a node:

(a) create the node either as the root of a representation, or as a node dependent on another one and placed to the right of all its sister nodes;

(b) choose its lexical value and the values of its grammatemes, meeting (by unification) the subcategorization conditions of its mother node and the restrictions on the combinations of grammatemes as specified by the lexical entry of the head or by its word class;

(c) if a root is being generated, its grammatemes determine it as a finite verb form, either CB (contextually bound, in topic) or NB (non-bound, in focus); the generated node is considered as node n in the next step;

(ii) if the frame of node n contains a complementation, then it is possible to generate

(a) either a left daughter of n, assign it 'CB' and a complementation value chosen from the frame of n,

(b) or a right daughter of n, assign it 'NB' and a complementation value from the left end of n's frame.

NOTE: If the chosen complementation value is an inner participant, it is deleted in the frame of n (as saturated); "from the left end" means that optional complementations can be skipped and deleted in the frame of n.

(iii) if the frame of n contains no complementation, then the mother node of n is considered as node n; if no mother node is present, the procedure (generating the tree from the top down and from the left to the right) is finished;

(iv) only representations containing a focus are accepted, i.e. only those whose branch going from the root to the rightmost daughter of the rightmost daughter of ... of the root includes a NB node.

(B) Declarative specification:

if the concept of unification is so complemented as to

(a) distinguish saturated items (cf. the Note above) and

(b) check the correspondence of the order of NB sister nodes with the systemic ordering, then it can be checked if a tree unifies with conditions stated in the lexicon:

all obligatory complementations of n are among the daughters of n; only complementations contained in the frame of n (or in the relevant list of free modifications) occur as daughters; subcategorization restrictions, etc., are met.

(C) This specification covers the core of sentence syntax; it has been completed to cover coordinated structures (a third dimension of the network), negation and other focalizers (*only*, *also*, *willingly*, etc.), in specific positions (prototypically at the beginning of the focus).

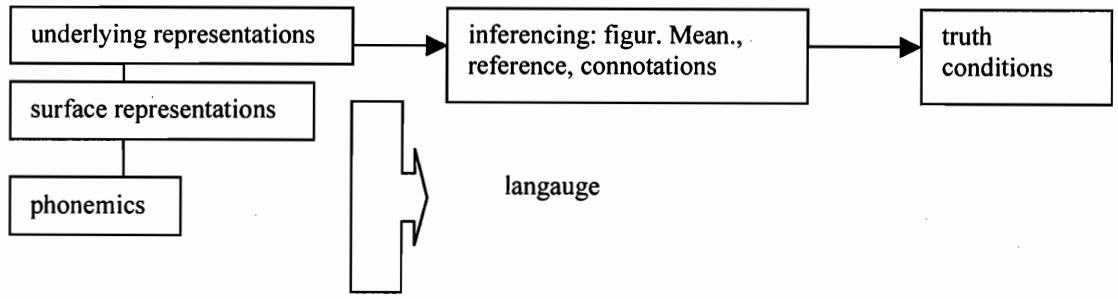
IX. UNIVERSAL GRAMMAR

common core of language systems; how much is innate, how much gained with interactivity?

An alternative to Chomsky's Universal Grammar:

dependency syntax, partially "free" word order in the underlying representations and much of grammatical information included in lexical entries; the core of syntax can be described in a minimalized, economical way.

This account of innate properties helps explain the acquisition of language, moreover, as embedded in communicative interaction, in context; it corresponds to the requirements of what Chomsky called Cartesian Linguistics:



LITERATURE:

Functional Generative Description:

E. Hajicová: Issues of sentence structure and discourse patterns. Charles University 1993.

P. Sgall, E. Hajicová and J. Panevová: The meaning of the sentence in its semantic and pragmatic aspects. Ed. By J.Mey. Dordrecht:Reidel and Prague:Academia, 1986.

E. Hajicová: "Free" word order described without unnecessary complexity. Theoretical Linguistics 17, 1991, 99-106.

J. Panevová: On verbal frames in Functional Generative Description. Prague Bulletin of Mathematical Linguistics 22, 1974, 3-40; 23, 1975, 17-52; a revised version in Prague Studies in Mathematical Linguistics 6, 1978, 227-254.

P. Sgall: Kommunikativnyj dinamizm i ponjatie sistemnogo uporjadochenija. In the journal Nauchno-texnicheskaja informacija, Serija 2, 1980, No. 6, pp. 24-29.

P. Sgall: Topic, presupposition and systemic ordering of participants. In: Tekst w kontekscie. Ed. by Tereza Dobrzynska. Wroclaw/Warszawa:Ossolineum, 1990, 33-39.

P. Sgall: Underlying structure of sentences and its relations to semantics. In: Wiener Slawistischer Almanach, Sonderband 33. Festschrift f r V. Ju. Rozencvejg, ed. By Tilmann Reuther, 1992, 273-282.