Logics in Constraint-Based Lexicalized Grammar (CBLG). Linguistically Significant Generalizations

Roussanka Loukanova

5 November, 2014
(Some) objectives of adequate linguistic theory

- Linguistic theory as theory of language phenomena
- The target is adequate linguistic theory
  - Correctness: distinguishing syntactically well-formed from ill-formed expressions in a given HL
  - Completeness: potential representation of all well-formed expressions that can be expressed in a given HL
- Semantic coverage
- Linguistically significant generalizations: Why? E.g.:
  1. to understand the nature of HL and other natural languages
  2. for adequateness of linguistic information
  3. for effective language learning and education
  4. for efficient human language processing (HLP)

- The above objectives require criteria of adequateness of a linguistic theory:
  Chomsky, on one side, and Barwise on the other, cover major phenomena partly by dividing syntax and semantics.
CBLG approach makes a significant step to a comprehensive grammar theory that covers

- lexicon
- lexical syntax
- phrasal syntax
- semantic representations on a par with syntax
- in a syntax-semantics interface fashion
- phonological (speech) and/or orthographical (written) representations

Traditionally, the above components are treated as separate subjects. As a consequence, typically, HLP (NLP) treats them in a relative isolation.

CBLG takes a unified approach by putting these components together as comprehensive grammar theory.
Is it possible to realize all of the following tasks

We target to put together the above components (modules) into a unified grammar system that is

- a computational, grammatical framework
  
  *Computational grammar:* as framework that uses methods of logic and other mathematical theories: computability, algebraic structure, recursion theory, probability, geometry (3D space & time grids, virtual worlds), etc.

- correct rendering of particular languages: syntactically well-formed expressions associated with adequate meanings

- to implement computerised grammar for HL processing

- for correct multilingual translation

- to represent adequately linguistic information

Yes, to some extend these tasks are covered, but work is largely open.
Major Characteristics of CBLG Approaches

- **Constraint-Based Approach**
  Well-formed syntactic structures are defined by satisfaction of constraints, instead of strict derivation process, e.g., as
  - feature-structure descriptions
  - dependent types
  - combined representation

- **Lexicalized Approach**
  Lexicon, Lexical Formation, and Phrase Syntax are (relatively) separated modules. Lexicon maintains the lexical items
  - as structured descriptions
  - by rules for lexical formation of lexemes (not necessarily words)
  - by rules for word formation — e.g., inflection rules

- **Interface between Lexicon, Lexical Formation, and Phrase Syntax**
  Word structures (formed in the lexicon module) are used in syntactic analyses of phrases.
Syntax-Semantics Interface

- Semantic Compositionality is formulated in the terms of syntactic rules.
- Dependence of syntactic well-formedness on the semantics of components: Binding phenomenon, in particular, reflexives.
Formal Syntax of HL in the CBLG Approaches

- **Initiation**: Pollard and Sag, 1987 [6]
- **Classics** in HPSG literature: Pollard and Sag, 1994 [7]

The ideas of Situation Theory and Situation Semantics taken for representing linguistic knowledge

- Information flaw: a piece of information carries other, new information.
- Partiality of information: in general, and in particular, of linguistic information and knowledge.
- Information accessible by human cognition is partial.
- How to represent partial information and knowledge, in a formal, mathematically precise way?
- New developments for distinctions between **partiality**, **underspecification**, **ambiguity**
Existing formalizations of CBLG

- The most complete works on formalization of HPSG by mathematical logic:
  - Mark Johnson, 1988 [2]
  - Paul J. King, 1989 [3]
  - Bob Carpenter, 1992 [1]
  - Gerald B. Penn, 2000 [5]
  - Frank Richter, 2004 [10]

- The formal syntax in CBLG, by Sag et al., 2003 [11], is based on models of typed functions, see:
  - Ch.9 [11]
  - A manuscript by Roussanka Loukanova [4]

- The most complete works on formalization of GF:
  - Aarne Ranta, 1994 [8, 9]

Note: I would classify GF as a new direction in (a new kind of) CBLG, under active development.
Here we give an introduction to major linguistic notions by using AVM notations popular in linguistics. The formal approach is specific to Sag et al. 2003 [11]:

**A formal CBLG language for expressing linguistics concepts**
- Feature structure descriptions are well-typed, well-formed expressions of the CBLG language.
- The grammar constraints are expressions of the language.

**Models of CBLG**
- The model domains consist of typed, feature-value functions, called feature structures.
- A feature-value function represents
  - linguistic information
  - the grammatical structure of a well-formed HL expression (e.g., of a sentence)
An Example for feature structure descriptions

S

phrase

[HEAD 0]

[COMPS ⟨ ⟩]

[SPR ⟨ ⟩]

SYN

[VAL]

NP

word

[HEAD 4]

[AGR 1]

[num 3]

[COMPS ⟨ ⟩]

[SPR ⟨ ⟩]

SYN

[VAL]

Maja

noun

[HEAD 1]

[AGR 3]

[NUM sg]

[PER 3rd]

[COMPS ⟨ ⟩]

[SPR ⟨ ⟩]

SYN

[VAL]

the toy

put

V

phrase

[HEAD 0]

[COMPS ⟨ ⟩]

[SPR ⟨ ⟩]

SYN

[VAL]

VP

[HEAD 4]

[COMPS ⟨ ⟩]

[SPR ⟨ ⟩]

[agr 3]

[agr 3]

NP

[agr 3]

[agr 3]

PP

on the table

[agr 3]

the table
An example of a (small) realistic type hierarchy in CBLG
Major constraints of $G$: their formulation may vary depending on the HL and the version of CBLG

- a set of grammar rules, e.g.:
  - The Head-Specifier Rule
  - The Head-Complement Rule
  - The Head-Modifier Rule(s)
  - The Coordination Rule, etc.

- a set of principles:
  - The Head Feature Principle
  - The Valence Principle
  - Semantic principles, etc.

- a finite set of lexical entries

- a set of lexical rules (word formation, inflection, etc.)

The constraints of $G$ are expressed in a language of feature structure descriptions.
Head Specifier Rule (HSR)

A phrase can consist of a (lexical or phrasal) head preceded by its specifier \[1\] (which can be a Det or NP, etc.):

\[
\begin{align*}
\text{phrase} & \rightarrow \text{[SYN } \text{VAL } \text{[SPR } \langle \rangle \text{]} \text{]} \\
\text{word} & \rightarrow \text{[SYN } \text{VAL } \text{[COMPS } \langle 1, \ldots, n \rangle \text{]} \text{]} \\
\end{align*}
\]

Head Complement Rule (HCR)

A phrase can consist of a lexical head followed by its complements:

\[
\begin{align*}
\text{phrase} & \rightarrow \text{[SYN } \text{VAL } \text{[COMPS } \langle 1 \rangle \text{]} \text{]} \\
\text{word} & \rightarrow \text{[SYN } \text{VAL } \text{[COMPS } \langle 1, \ldots, n \rangle \text{]} \text{]} \\
\end{align*}
\]

where \(n \geq 0\)
Head-Modifier Rule (HMR) (post-modifier)

A phrase can consist of a (lexical or phrasal) head followed by a compatible modifier:

\[
[\text{phrase}] \quad \rightarrow \quad \text{H}[1\text{syn}\text{[comps} \langle \rangle]\text{]} \quad \text{[syn}\text{val}\text{[comps} \langle 1\rangle]\text{]} \quad \text{[syn}\text{val}\text{mod}\text{[comps} \langle 1\rangle]\text{]}]
\]

MOD is a VAL feature:

- **SPR, COMPS and MOD** are similar w.r. to specifying co-occurrence restrictions on the lexical head.
- The way of imposing those restrictions for MOD is different from that of SPR and COMPS (compare with the HSR and the HSR w.r. to what is the head daughter).
Valence Principle (VPr)

Unless the rules dictate otherwise, the mother’s values of the VAL features (spr, comps, mod) are identical to those of the head daughter.

- The kind of expression a modifier can modify is determined by the MOD value of its head. This value is passed up to the modifier phrase by the Valence Principle.
- Packing together MOD, SPR, and COMPS as VAL features facilitates the statement of the Coordination Rule, which implies that:
  - the conjuncts have identical VAL values

For an example with coordination see p. 21.
HMR does not require anything about MOD in 7.

∴ By the VPr, all VAL values of the modifier structure 7 are identical to the VAL values of its head NOM structure 1.
Head Feature Principle (HFP)

- In any headed feature-structure, the HEAD value of the mother is identical to the HEAD value of the head daughter.

- A local tree structure \( \Phi \) satisfies the HFP with respect to a headed rule \( \rho \) iff \( \Phi \) satisfies:

\[
\Phi
\begin{array}{c}
\text{HEAD} \\
\hline
1
\end{array}
\]

\[
\begin{array}{c}
\cdots \\
\phi_h \\
\cdots
\end{array}
\]

\[
\begin{array}{c}
\text{HEAD} \\
\hline
1
\end{array}
\]

where \( \phi_h \) is the head daughter of \( \Phi \).
Agreement and the notion of a head

- The notion of a head: intuitively, the head of a given expression is the sub-expression, which contributes major linguistically significant information.

- E.g., the head daughter determines the pos of the mother (or the other way around):
  The pos of a NP is noun because its head daughter is a N, or a NOM, the pos of which is noun.

The head feature-structure is determined by:
- the feature HEAD in the type hierarchy
- the Specifier-Head Agreement Constraint (SHAC)
- Head Specifier Rule (HSR)
- Head Complement Rule (HCR)
- Head Feature Principle (HFP)
Characteristics of CBLG Approaches
Components of CBLG in Feature-Structure Descriptions
The notion of a head
The Type Hierarchy of the Lexicon

The Specifier-Head Agreement Constraint (SHAC)
Propagation of the Head

feat-struc
- l-rule
  - i-rule
  - d-rule
  - pi-rule
- synsem
  - SYN syn-cat
  - SEM sem-cat

lexeme
- ARG-ST list(exp)

expression
- word ...

infl-lxm
- SHAC: SYN HEAD
- VAL SPR [AGR U]
- const-lxm
- pn-lxm
- pron-lxm

cn-lxm
- cn-lxm
- massn-lxm
- iv-lxm
- verb-lxm
- tv-lxm
  - siv-lxm
  - piv-lxm
  - stv-lxm
  - btv-lxm
  - ptv-lxm

Roussanka Loukanova Logics in CBLG
Propagating the Head: an example for agreement by the HFP and SHAC
S

phrase
HEAD 3
VAL [COMPS ( )]

phrase
HEAD 3
VAL [COMPS ( )]

word
HEAD 1
AGR 2
VAL [num sg]

word
HEAD 3
AGR 2
VAL [num sg]

word
HEAD 6
AGR 2
VAL [num sg]

word
HEAD 4
AGR 2
VAL [num sg]

word
HEAD 6
AGR 2
VAL [num sg]

word
HEAD 4
AGR 2
VAL [num sg]
Bob Carpenter.  

Mark Johnson.  

Paul John King.  
Roussanka Loukanova.
An Approach to Functional Formal Models of Constraint-Based Lexicalized Grammar (CBLG).
draft, 2010.

Gerald B. Penn.
*The Algebraic Structure of Attributed Type Signatures.*

Carl Pollard and Ivan A. Sag.
*Information-Based Syntax and Semantics, Part I.*

Carl Pollard and Ivan A. Sag.
*Head-driven Phrase Structure Grammar.*
Aarne Ranta. 
*Type-theoretical Grammar.*  

Aarne Ranta. 
Grammatical framework: A Type-Theoretical Grammar Formalism.  
URL:  

Frank Richter.  
*A mathematical formalism for linguistic theories with an application in Head-driven Phrase Structure Grammar.*  