Stratificational Approach to Language Description

Markéta Lopatková
Institute of Formal and Applied Linguistics, MFF UK
lopatkova@ufal.mff.cuni.cz
Stratificational Approach – Basic Principles

- every language comprises a restricted number of structural layers or strata,
- strata hierarchically related in such a way that
  - units or combinations of units on one stratum realize units or combinations of units of the next higher stratum
  - strata are linearly ordered
Stratificational Approach – Basic Principles

• every language comprises a restricted number of structural layers or strata,
• strata hierarchically related in such a way that
  • units or combinations of units on one stratum realize units or combinations of units of the next higher stratum
  • strata are linearly ordered
• the number of strata vary from (linguistic) theory to theory
  • semantics, constituted by …
    – sememic stratal system (semantics, deep structure)
  • grammar, constituted by
    – lexemic stratal system and
    – morphemic stratal system (surface structure)
  • phonology, constituted by
    – phonemic system
every language comprises a restricted number of structural layers or strata, 

strata hierarchically related in such a way that

- units or combinations of units on one stratum realize units or combinations of units of the next higher stratum
- strata are linearly ordered

the number of strata vary from (linguistic) theory to theory

- semantics, constituted by …
  - sememic stratal system (deep structure)
- grammar, constituted by
  - lexemic stratal system and
  - morphemic stratal system (surface structure)
- phonology, constituted by
  - phonemic system

grammar relates to semantics and phonology in a same way as the lexemic and the morphemic stratal systems within the grammar

language as a network of relationships (rather than system of rules)
S. Lamb and Stratificational Approach

- Sydney M. Lamb … *Outline of Stratificational Grammar* (1966)
- Berkeley, follower of glossematic school
- four necessary levels of sentence analysis:
  - the sememic stratum … structure of clauses and sentences
  - the lexemic stratum … structure of phrases
  - the morphemic stratum … structure of word forms
  - the phonemic stratum … syllable structure
- each stratum has its elementary units
- each stratum has its own combinatorial pattern
- strata are hierarchically related
  - each “realized” by the elements in the level structurally beneath it
  - without making use of rules that convert one entity into another
S. Lamb and Stratificational Approach

- Sydney M. Lamb … *Outline of Stratificational Grammar* (1966)
  - Berkeley, follower of glossematic school
  - four necessary levels of sentence analysis:
    - the sememic stratum … structure of clauses and sentences
    - the lexemic stratum … structure of phrases
    - the morphemic stratum … structure of word forms
    - the phonemic stratum … syllable structure
  - each stratum has its elementary units
  - each stratum has its own combinatorial pattern
  - strata are hierarchically related
    - each “realized” by the elements in the level structurally beneath it
    - without making use of rules that convert one entity into another

- Functional Generative Description (FGD)
- Meaning ⇔ Text Theory (MTT)
Stratificational grammar as an alternative to transformational grammar (Chomsky)

vs. Noam Chomsky's
Aspects of the Theory of Syntax (1965)
- system of three components (syntax, semantics, phonology)
- three types of rules (phrase-structure, transformational, morphophonemic)
Functional Generative Description
**Functional Generative Description**

- motivation: machine translation

---

![Diagram](image)

- ‘interlingua’
- language independent representation
- sentence ~ string of graphemes/phonemes

---

Sgall, P. (1967) *Generativní popis jazyka a česká deklinace*. Academia, Praha
Functional Generative Description

• motivation: machine translation

Sgall, P. (1967) *Generativní popis jazyka a česká deklinace*. Academia, Praha
Basic characteristics of FGD

'classical' version of FGD:

- dependency framework
  - formal description
  - suitable mathematical formalism
Basic characteristics of FGD

'classical' version of FGD:
• dependency framework
• stratificational approach

synonymy  ambiguity

language meaning \sim function

string of graphemes/phonemes \sim form
Basic characteristics of FGD

'classical' version of FGD:

• dependency framework
• stratificational approach
• relation between a form and its function / a function and its form
Basic characteristics of FGD

'classical' version of FGD:
• dependency framework
• stratificational approach
• relation between a form and its function / a function and its form

structural linguistics:
• language meaning (not cognitive content)
• language as a system ~ langue vs. individual utterances ~ parol
• stress on testable criteria for distinguishing lang. phenomena
Two components of FGD

- **generative component**
  ~ to define all formally correct meaning representations
  (of possible sentences of a given language)
- formalism: 1) phrase rules, phrase structure trees + functors
  2) dependency trees
- push-down automaton
Two components of FGD

• **generative component**
  ~ to define all formally correct meaning representations
  (of possible sentences of a given language)
  • formalism: 1) phrase rules, phrase structure trees + functors
  2) dependency trees
  • push-down automaton

• **translation component**
  ~ translating meaning representations to lower layers
  • sequence of push-down transducers plus finite-state automaton
## System of layers in FGD

<table>
<thead>
<tr>
<th>meaning</th>
<th>deep / underlying syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tectogrammar</td>
</tr>
<tr>
<td></td>
<td>surface syntax</td>
</tr>
<tr>
<td></td>
<td>morphematics</td>
</tr>
<tr>
<td></td>
<td>morphonology</td>
</tr>
<tr>
<td>expression</td>
<td>phonology/phonetics</td>
</tr>
</tbody>
</table>
System of layers in FGD (cont.)

sentence … full representation on each layer of description

each layer ~ set of descriptions for all possible sentences
  • finite set of elementary units
  • finite set of operations and relations → set of complex units
  • finite set of relations between sentence representations
    on a particular layer and its representations on adjacent layers

Dependency Grammars and Treebanks – Stratificational Approach
System of layers in FGD (cont.)

sentence … full representation on each layer of description

each layer ~ set of descriptions for all possible sentences
  • finite set of elementary units
  • finite set of operations and relations → set of complex units
  • finite set of relations between sentence representations on a particular layer and its representations on adjacent layers

\[
\begin{array}{c}
\text{n+1} \\
\downarrow \\
\text{n} \\
\end{array}
\begin{array}{c}
\text{R} \\
\text{C} \\
\text{C} \\
\end{array}
\begin{array}{c}
\text{function} \\
\text{form} \\
\end{array}
\]

type C relations (composition): elementary units constitute complex units i.e., relations between units of the same layer

type R relations (representation): form-function relation i.e., relation between adjacent layers
**System of layers in FGD (cont.)**

**Layer of phonetics**
- distinctive features ... elementary units
- phones (~ a speech sound) ... complex units
- suprasegmental units ... prosody, intonation
System of layers in FGD (cont.)

**layer of phonetics**

- distinctive features ... elementary units
- phones (~ a speech sound) ... complex units
- suprasegmental units ... prosody, intonation

**layer of phonology**

- distinctive features ... elementary units
- phonemes (~ ‘smallest’ units that distinguish meaning) ... complex units
- asymmetry ... allophones ~ variants of a single phoneme
- language dependent (sing vs. sin)
System of layers in FGD (cont.)

layer of morphonology

morphoneme ~ set of phoneme variants  e.g. \(k|c|č|.k\) in "matka"
morph ~ string of morphonemes
lexical variants \((matk, matc, matč, mat.k)\) ... 4 allomorphs \(mat(k|c|č|.k)\) 1 morph
lexical variants \((foot, feet)\) ... 2 allomorphs \(f(oo|ee)t\) 1 morph

Dependency Grammars and Treebanks – Stratificational Approach
Layer of morphematics

- morpheme ~ the smallest component that has semantic meaning
- lexical morpheme
  - roots
    e.g. lex. morpheme for *matka* consists of 4 allomorphs (*matk, matc, matč, mat.k*);
    for *to write* (*writ, wrot*); for *leaf* (*leaf, leav*)
  - derivational morphemes (affixes: prefixes, infixes, suffixes, …)
    *il-* (as in *illegal*), *non-* (as in *nonproblematic*)
    *-ly* (as in *legally*), *-ess* (as in *actress*)
layer of morphematics

- morpheme ~ the smallest component that has semantic meaning
- lexical morpheme
- grammatical morpheme
  - inflectional affixes  e.g. Cz: suffixes
    - nouns: case, gender, number, …
    - verbs: gender, number, tense, voice,
  - Eng: suffixes
    - nouns: plural -s
    - verbs: past tense -ed , continuous –ing
System of layers in FGD (cont.)

layer of morphematics

- morpheme ~ the smallest component that has semantic meaning
- lexical morpheme
- grammatical morpheme
  - inflectional affixes  e.g. Cz: suffixes
    - nouns: case, gender, number, …
    - verbs: gender, number, tense, voice,
  - Eng: suffixes
    - nouns: plural -s
    - verbs: past tense -ed, continuous –ing
  - sema … a combination of grammatical morphemes that characterize a lexical morpheme (or strings of lexical morphemes)
System of layers in FGD (cont.)

layer of morphematics

- morpheme ~ the smallest component that has semantic meaning
- lexical morpheme
- grammatical morpheme
- formeme:
  - sequence of morphs realizing a single tagmeme / sentence member
  - lexical f., case f. (i.e., prep+case), conjunction formemes (i.e., conj+verb mood)

Cz: vysok+á škol+a; lamp+a; na+lavíc+i; chod+í; bud+e+chod+it
Eng: white-collar; lamp; on+ table; walk+s; will+be+walk+ing
System of layers in FGD (cont.)

morpheme ~ the smallest component that has semantic meaning

Czech … (inflection language):

nejneobhospodařovatelnější
nej-ne-ob-hospod-ar’-ova-teln-ější - í
most-non- cultivate - [iter]- [adj] - [super]-[sg+nom+fem|sg+acc+neutr|… pl+voc+masc]

23 combinations ("meanings")

grammatical morphemes
System of layers in FGD (cont.)

morpheme ~ the smallest component that has semantic meaning

Hungarian (agglutinative language):

- **fi-ú**  
  boy

- **fi-a**  
  his/her son

- **fi-á-é**  
  his/her son's (singular object)

- **fi-á-é-i**  
  his/her son's (plural object)

- **fi-a-i**  
  his/her sons

- **fi-a-i-é**  
  his/her sons' (singular object)

- **fi-a-i-é-i**  
  his/her sons' (plural object)
System of layers in FGD

two layers of syntax
- tree-based dependency structure
  - nodes for tagmemes / sememes (complex symbols)
  - edges labeled with a type of a respective syntactic relation
The layer of **surface syntax**

My brother often sleeps in his study.

```
```

Po babiččině příjezdu půjdou rodiče do divadla.
[After grandma's arrival the parents will go to the theatre.]

```
půjdou.fut.Pred
```

Dependency Grammars and Treebanks – Stratificational Approach
The layer of **surface syntax**

**surface syntactic tree**
- ~ nodes for formemes → **tagmemes** / sentence members
  (cz school syntax: větné členy)
- ~ edges for syntactic relations

+ **surface word order** … linear ordering of tree nodes
The layer of **surface syntax**

**surface syntactic tree**

~ nodes for formemes \( \rightarrow \textbf{tagmemes} / \text{sentence members} \)
(cz school syntax: větné členy)

~ edges for syntactic relations

+ **surface word order** ... linear ordering of tree nodes

3 types of elementary units:

• lexical: units from a dictionary

• morphological: set of morphological features \( \sim \textbf{tags} \)
  (a pair of) trousers ... sema - plural

• syntactic: subject, object, attribute, adverbial, complement,...
The layer of **deep syntax**

*Po babiččině příjezdu půjdou rodiče do divadla.*

[After grandma's arrival the parents will go to the theatre.]

---

Dependency Grammars and Treebanks – Stratificational Approach
The layer of **deep syntax**

~ meaning of a sentence:

I. **semantemes**: only content (lexical) words as nodes, their lexical and morphological features and mutual relations terminology: deep / underlying / tectogrammatical representation (TR)

   • modal verbs
     
     *Peter wants* *to* *attend the concert.* [to attend + volitive]
     
     *Charles has* *to* *pass the exam.* [to pass + debitive]

   • nominalization
     
     *After grandma's* *arrival* … → [to arrive]

   • active / passive verbs → [active form]
     
     *Tato krásná kniha byla vydána nakladatelstvím Albatros.*
     
     [This beautiful book was published by the Albatros publishing house.]
The layer of **deep syntax**

~ meaning of a sentence:

II. 3 basic types of elementary units:

- lexical: units from a (tectogrammatical) dictionary
- morphological: **grammatemes**
  - *meaning* of individual morphological categories
  - (a pair of) trousers ... singular
    - denominating *(pojmenovávací)*
  - vs. correlating *(usouvztažňující)* categories
- syntactic: types of relation, **functors** and **subfunctors**
  - Actor, Patient, Addressee, ... local, temporal modifications ...
The layer of deep syntax

~ meaning of a sentence:

III. completeness of the representation

- (surface) ellipses are restored
- omitted surface subject, object, comparison … valency


Spanish: ¿Ves este tronco? [(Do) you see this log? ]
The layer of **deep syntax**

~ meaning of a sentence:

IV. **deep word order**

- information structure / topic focus articulation
- increasing communicative dynamism: word order reflects "relative degree of importance in comparison with other expressions in the sentence […]"
- condition of **projectivity** !!!
The layers of **surface** vs. **deep syntax**

1. different sets of elementary units
   - 'morphological' lemma vs. tectogrammatical lemma
   - morphological categories vs. grammatemes
   - surface sentence members vs. functors

2. different sets of complex units
   - tagmeme vs. semanteme
The layers of **surface vs. deep syntax**

I. different sets of elementary units
   - 'morphological' lemma vs. tectogrammatical lemma
   - morphological categories vs. grammatemes
   - surface sentence members vs. functors

II. elipses and completeness
   - nodes represent formemes as appear in the surface sentence vs. **completeness of the representation** (valency)

Dependency Grammars and Treebanks – Stratificational Approach
The layers of **surface vs. deep syntax**

I. different sets of elementary units
   - 'morphological' lemma vs. tectogrammatical lemma
   - morphological categories vs. grammatemes
   - surface sentence members vs. functors

II. elipses and completeness
   - nodes represent formemes as appear in the surface sentence vs. **completeness of the representation** (valency)

III. surface vs. deep word order
   - **order of words** in a surface sentence vs. information structure
   - nonprojective trees allowed vs. just projective trees

Dependency Grammars and Treebanks – Stratificational Approach
Meaning ⇔ Text Theory (MTT)
Meaning⇔Text Theory (MTT)

- Aleksandr Žolkovskij and Igor Mel’čuk, Moscow, 1965-…
  - formal representation of natural language (Russian)
  - aim: applications in NLP (machine translation, phraseology, lexicography)
- basic principle:
  - language consists in a mapping from the content or meaning (semantics) of an utterance to its form or text (phonetics)
  - sequence of mappings
    - the unordered network of the semantic representation (SemR)
    - dependency tree-structures of the syntactic representation (SyntR)
    - linearized chain of morphemes of the morphological representation (MorphR)
    - (the temporally-ordered string of phones of the phonetic representation, PhonR)
Meaning⇔Text Theory (MTT) and FGD

- both originated almost at the same time (Žolkovskij, Mel’čuk, 1965), (Sgall, 1967)
- the same roots in European structural linguistics
- both stratificational
- both dependency oriented (syntax)
- distinguishing deep and surface syntactic representations
- independence between the dependency structure and word order in a sentence.
- orientation to languages typologically different than English
- both proved to be useful and successful in large-scale implementations
  - machine translation system ETAP for MTT (Apresian et al., 2003);
  - Prague Dependency Treebank (PDT) for FGD, (Hajič et al., 2001),
    machine translation systems, esp. CZ-Eng pair

Based on (Žabokrtský, 2006)
Meaning ⇔ Text Theory (MTT) vs. FGD

- **Dependence Grammars and Treebanks – Stratificational Approach**

  - **Meaning**
    - semantic representation
    - deep-syntactic representation
    - surface-syntactic representation
    - deep-morphological representation
    - surface-morphological representation
    - deep-phonological representation
    - surface-phonological representation
    - tectogrammatical representation
    - surface-syntactic representation
    - morphematic representation
    - morphophonological representation
    - phonological representation

  - **Text/sound**

Based on (Žabokrtský, 2006)

Dependency Grammars and Treebanks – Stratificational Approach
References

• Sgall, P. (1967) Generativní popis jazyka a česká deklinace. Academia, Praha

https://www.britannica.com/science/linguistics/Stratificational-grammar