The Complexity Sources and their Compensation in Language Processing

Philippe Blache

Laboratoire Parole et Langage
CNRS & Aix-Marseille Université
Situation: different kinds of complexity

- System vs. Structural Complexity (Dahl, 2004)

<table>
<thead>
<tr>
<th>System complexity</th>
<th>Structural complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of categories in each domain</td>
<td>Number of categories</td>
</tr>
<tr>
<td>Number of features for each categories</td>
<td>Depth</td>
</tr>
<tr>
<td>Grammar size for each domain</td>
<td>Number of rules used to build the structure</td>
</tr>
<tr>
<td>Lexicon size</td>
<td>Number of words</td>
</tr>
<tr>
<td>Average local complexity</td>
<td></td>
</tr>
</tbody>
</table>

- Absolute vs. Relative Complexity (Miestamo, 2008)
  - *Absolute complexity*: theory-oriented
  - *Relative complexity*: user-dependent

- Our perspective: structural relative complexity

Difficulty Models in Psycholinguistics

- **Existing models**
  - *Incomplete Dependency Hypothesis*
  - *Dependency Locality Theory*
  - *Early Immediate Constituents Principle*
  - *Activation*

- However, they fail at:
  - Describing language in its natural environment
  - Explaining interaction between sources of information

---

Language processing in the real world

- New challenges for Linguistics, Psycholinguistics and NLP
  - Dealing with natural data
  - Language in its context: spoken language, natural interaction

- Issues
  - Units are not always possible to determine, difficult to categorize (gradience)
  - Information can be parcimonious
  - Language processing relies on domains interaction
Overview

- Hypothesis: *Difficulty depends on the search space size*
  The larger the search space, the more difficulty

- Question: *How search space size can be controlled?*

  The more properties, the smaller the search space

- Questions:
  - How to represent properties?
  - How properties implement structural complexity?
  - How structural complexity predict processing difficulty?
Outline

- Basis: a constraint-based representation for syntax
- Cohesion: a model for syntactic complexity
  - How to measure cohesion
  - Experimenting cohesion in human language processing: an interplay between difficulty and facilitation
Part I

Representing syntactic information
Classical Generative Syntax

Language and grammar

- Derivation
- Language = set of derived strings
- Recursively enumerable

Parsing

- Finding a derivation
- Building a tree
- Consequences:
  - There exists a complete grammar of the language
  - The initial system is a complete grammar (acquisition)
Our Proposal: *Property Grammars*

- Describing the characteristics of an input (not building a structure)
- Linguistic statements as constraints
- Declarative approach: no specific mechanism but constraint evaluation

**Basics:**
- Constraints are independent
  - No ranking (contra OT)
  - Separate evaluation (contra (GEN))
- No hierarchical structure (contra PSG)
- Constraints are at the same level (contra DEP)
What kind of syntactic information?

- Linear precedence
- Mandatory cooccurrence between two categories
- Impossible cooccurrence between two categories
- No repetition of the same category within a construction
- Dependency between two categories
Linearity

\[ Prec(A, B) : (\forall x, y)[(A(x) \land B(y) \rightarrow y \not\prec x)] \]

- Example: nominal construction

<table>
<thead>
<tr>
<th>Det \prec Adj</th>
<th>Det \prec N</th>
<th>Det \prec ProR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj \prec N</td>
<td>N \prec ProR</td>
<td>N \prec Prep</td>
</tr>
</tbody>
</table>

\[
\text{the} \rightarrow \text{very} \rightarrow \text{famous} \rightarrow \text{reporter} \rightarrow \text{who} \rightarrow \text{the senator attacked}
\]
Requirement (cooccurrence)

\[ \text{Req}(A, B) : (\forall x, y)[A(x) \rightarrow B(y)] \]

- Example:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ProR ⇒ N</td>
<td>V[ditrans] ⇒ Prep</td>
<td>Prep ⇒ N</td>
</tr>
</tbody>
</table>

- Relations without government:

1. a. The most interesting book of the library
   b. *A most interesting book of the library

Sup ⇒ Det[def]
Exclusion (cooccurrence restriction)

\[ Excl(A, B) : (\forall x)(\exists y)[A(x) \land B(y)] \]

Examples

- Nominal construction:
  \[
  \begin{align*}
  \text{Pro} \otimes \text{N} & \quad \text{N[prop]} \otimes \text{N[com]} & \quad \text{N[prop]} \otimes \text{Prep[inf]}
  \end{align*}
  \]

- Relative construction:
  \[
  \begin{align*}
  \text{ProR[subj]} \otimes \text{N[subj]}
  \end{align*}
  \]
Uniqueness

Uniqueness

\[ \text{Uniq}(A) : (\forall x, y)[A(x) \land A(y) \rightarrow x \approx y] \]

Example (nominal construction):

\[ \text{Uniq} = \{\text{Det, ProR, Prep[inf], Adv}\} \]

\[ \text{The book that I read} \]
Dependency: the type hierarchy

- **dep**: generic dependency relation
- **mod**: modification (typically adjunction)
- **spec**: specification (typically Det-N)
- **comp**: head-complement
- **subj**: subject
- **obj**: direct object
- **iobj**: indirect object
- **xcomp**: other complements (e.g. $N - Prep$)
- **aux**: auxiliary-verb
- **conj**: conjunction
Example:

The most interesting book of the library

Syntactic role: $N[subj] \sim_{subj} V$

Agreement: $Det[agr_i] \sim_{spec} N[agr_i]; Adj[agr_i] \sim_{mod} N[agr_i]$
Example: the nominal construction

<table>
<thead>
<tr>
<th>Det ⊲ {Det, Adj, ProR, Prep, N}</th>
<th>Det ⊳_{\text{spec}} N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N ⊲ {Prep, ProR}</td>
<td>Adj ⊳_{\text{mod}} N</td>
</tr>
<tr>
<td>Pro ⊗ {Det, Adj, ProR, Prep, N}</td>
<td>ProR ⊳_{\text{mod}} N</td>
</tr>
<tr>
<td>N[\text{prop}] ⊗ Det</td>
<td>Prep ⊳_{\text{mod}} N</td>
</tr>
<tr>
<td>Uniq = {Pro, Det, N, ProR, Prep}</td>
<td>Det ⇒ N[\text{com}]</td>
</tr>
<tr>
<td></td>
<td>{Adj, ProR, Prep} ⇒ N</td>
</tr>
</tbody>
</table>
Syntactic Representation: Constraint Graph

The ➙ most ➙ interesting ➙ book ➙ of ➙ the ➙ library

spec

r

l

mod

mod

mod

mod

comp

spec

u

u

u

u

r

l

l

l

l
Constraint violation

<table>
<thead>
<tr>
<th>Constraint graph</th>
<th>Characterization</th>
</tr>
</thead>
</table>
| ![Constraint graph 1](image1) | \[ P^+ = \{Det \prec Adj, Det \prec N, Adv \prec Adj, Adj \prec N, Det \prec N, Adj \Rightarrow N, Adv \Rightarrow N, Adj, Adj \Rightarrow N\} \]
| ![Constraint graph 2](image2) | \[ P^- = \emptyset \]

| ![Constraint graph 3](image3) | \[ P^+ = \{Det \prec N, Adv \prec Adj, Adj \prec N, Det \Rightarrow N, Adj \Rightarrow N, Adv \Rightarrow N, Adj, Adj \Rightarrow N\} \]
| ![Constraint graph 4](image4) | \[ P^- = \{Det \prec Adj\} \]
Part II

Measuring cohesion
Graph-based measures

- Definition: $\text{degree} = \text{number of edges incident to the vertex}$
- Degree of a category in the grammar:

$$deg_{\text{gram}}(N) = 9$$
$$deg_{\text{gram}}(ProR) = 2$$
$$deg_{\text{gram}}(Adj) = 1$$

- Degree of a category in the sentence:

$$deg_{\text{sent}}(N) = 5$$
$$deg_{\text{sent}}(Adj) = 1$$
$$deg_{\text{sent}}(Det) = 1$$
Category Completeness

- The completeness level of a category depends on the number of relations in its description.
- This measure also depends on the number of relations for the category in the grammar.
- Completeness ratio: the higher the number of relations in the grammar is verified, the higher the completeness value.

\[
\text{completeness}(\text{cat}) = \frac{\text{deg}_{\text{sent}}(\text{cat})}{\text{deg}_{\text{gram}}(\text{cat})}
\]
Sentence *Density*

- Measure based on 4 types or properties: *uniqueness, requirement, dependency, linearity*

- **Density of a construction**: the ratio btw the evaluated properties and the possible properties:

\[
\text{density}(\text{sent}) = \frac{|\text{properties}(\text{sent})|}{|\text{words}(\text{sent})|}
\]
Satisfaction ratio

- All constraints can be violated
- A characterization contains both satisfied and violated constraints
- The “quality” of a construction depends on the ratio satisfied/violated
- All constraints can be weighted. We note $W^+$ (resp. $W^-$ the sum of the weights of the satisfied (resp. violated) constraints:

\[
satisfaction(sent) = \frac{W^+ - W^-}{W^+ + W^-}
\]
Cohesion function

Given $S$ a sentence, $w$ the set of its words:

$$
\text{cohesion}(S) = \sum_{i=1}^{|S|} \text{completeness}(w_i) \times \text{density}(S) \times \text{satisfaction}(S)
$$

Hypothesis: cohesion is correlated with difficulty
le côté hystérique un peu de enfin c’est normal tu vois elle souffre et machin

*the hysterical side rather of well it’s normal you see she suffers and that’s it*
Example 2

le chien apparemment connaissait parfaitement –
– le coin
et euh quand on est partis
le chien a décidé de nous suivre

the dog apparently knew perfectly –
– the area
and hmm when we left
the dog decided to follow us
Evaluation

<table>
<thead>
<tr>
<th>Cat</th>
<th>Degree-gram</th>
<th>Deg-sent</th>
<th>Deg-gram</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Det</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>34</td>
<td>0</td>
<td>34</td>
<td>0.15</td>
</tr>
<tr>
<td>Adj</td>
<td>11</td>
<td>3</td>
<td>11</td>
<td>0.27</td>
</tr>
<tr>
<td>Adv</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Prep</td>
<td>31</td>
<td>2</td>
<td>31</td>
<td>0.06</td>
</tr>
<tr>
<td>Pro</td>
<td>4</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Conj</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>Aux</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>0.43</td>
</tr>
<tr>
<td>V</td>
<td>7</td>
<td>2</td>
<td>11</td>
<td>0.18</td>
</tr>
<tr>
<td>Conj</td>
<td>21</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word</th>
<th>Deg-sent</th>
<th>Deg-gram</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>le</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>côté</td>
<td>5</td>
<td>34</td>
<td>0.15</td>
</tr>
<tr>
<td>hystérique</td>
<td>3</td>
<td>11</td>
<td>0.27</td>
</tr>
<tr>
<td>un</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>peu</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>de</td>
<td>2</td>
<td>31</td>
<td>0.06</td>
</tr>
<tr>
<td>enfin</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>c'</td>
<td>1</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>est</td>
<td>3</td>
<td>7</td>
<td>0.43</td>
</tr>
<tr>
<td>normal</td>
<td>2</td>
<td>11</td>
<td>0.18</td>
</tr>
<tr>
<td>tu vois</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>elle</td>
<td>1</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>souffre</td>
<td>2</td>
<td>7</td>
<td>0.29</td>
</tr>
<tr>
<td>et</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>machin</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>
Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Words</th>
<th>Constraints</th>
<th>Completeness</th>
<th>Density</th>
<th>Cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sent. 1</strong></td>
<td>15</td>
<td>19</td>
<td>0.13</td>
<td>1.18</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>Sent. 2</strong></td>
<td>21</td>
<td>38</td>
<td>0.17</td>
<td>1.80</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Part III

Difficulty and Compensation: the Case of Idiom Processing
The situation

**Idiom:** multiword expression with a figurative meaning separate from the literal meaning

**Examples**

- **Decomposable idioms (variables)**
  
  ‘‘let the cat out of the bag’’

- **Non-decomposable idioms (opaque semantics, no variability)**
  
  ‘‘spill the beans’’, ‘‘kick the bucket’’

**Experimental perspective**

- Idioms are read faster
- Idioms are related with specific brain activities
- Two different models according to the way they are processed
Compositional models (nonlexical models)

Main ideas

▶ Idiom comprehension uses normal language processing
▶ Idioms are represented as *configurations* of lexical items, no separate representation in the lexicon

The Configurational Hypothesis


▶ A sufficient portion of an idiomatic expression must be processed literally before the idiom can be identified
▶ After the “Recognition Point”, rest of the string is not processed literally
The *Cohesion Model*: interplay between difficulty and facilitation

- Our hypothesis: *Difficulty can be compensated by Cohesion*

- Experiment:
  - Idioms have high cohesion values
  - Introducing a difficulty into an idiom (a syntactic violation) is compensated by the cohesion
  - We compare idiomatic vs. non-idiomatic sentences, with and without violation
Experimental design

Material

- Idiom (IDNV)
  
  Paul a une idée derrière la tête depuis ce matin

- Idiome with violation (IDV)
  
  Paul a une idée derrière le tête depuis ce matin

- Control (CTRNV)
  
  Paul a une douleur derrière la nuque depuis ce matin

- Control with violation (CTRV)
  
  Paul a une douleur derrière le nuque depuis ce matin
Experimental design

Specific positions to study

- Recognition point (RP)
  
  *Paul a une idée derrière la tête depuis ce matin*

- Modified word, where the violation is introduced (MM)
  
  *Paul a une idée derrière le tête depuis ce matin*

- Detection word, where the violation is detected (MD)
  
  *Paul a une idée derrière le tête depuis ce matin*
Results: recognition point

- Different processing idiom vs. control
- More positive amplitude in the P300 and N400 windows for idioms: facilitation
More negative N400 for violated idiom (IDV) than non violated (IDNV): surprisal at the unexpected (modified) word for idioms
No significative P600 : no repair
Results: detection word

- Small N400, small P600 for the violated control
- Positive deflection for IDV (wrt IDNV) at N400+P600 : repair
Results

- Idioms are processed differently: more positive amplitude after RP
- Modifying a word after the RP in idioms generates N400: surprisal
- At the position of syntactic violation detection:
  - High negativity (difficulty) for control sentences
  - Earlier positivity (P300) for idioms: expectancy confirmation
Paul a une idée derrière le tète depuis ce matin.

Paul a une douleur derrière la nuque depuis ce matin.
Paul a une douleur derrière la nuque depuis ce matin

Paul a une idée derrière la tête depuis ce matin
Violation Compensation

Paul a une douleur derrière la nuque depuis ce matin

Paul a une idée derrière la tête depuis ce matin

Paul a une idée derrière le tête depuis ce matin
Conclusion

What can be done with constraints

▷ Describing whatever the input (including ill-formed)
▷ Measuring structural complexity

Complexity Model : a Cognitive Perspective

▷ An interplay between difficulty and facilitation
▷ An interaction between different sources of information
▷ Complexity depends on the quantity of information to reduce the search space
▷ Necessary to take into account the cognitive matrix as well as the context