Distribution of Valency Complements in Czech Complex Predicates: Between Verb and Noun

Václava Kettnerová and Eduard Bejček
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague, Czech Republic
complex predicates (CPs)

- verbal MWEs composed of light verbs and predicative nouns

  - vést jednání ‘to hold talks’
  - míť potíže ‘to have difficulties’
  - udělat chybu ‘to make a mistake’
  - dostat příkaz ‘to get an order’

- light verbs: syntactic center of CPs
- predicative nouns: semantic center of CPs
• complex predicates (CPs)
  ~ verbal MWEs composed of light verbs and predicative nouns

  vést jednání  ‘to hold talks’
  mít potíže    ‘to have difficulties’
  udělat chybu ‘to make a mistake’
  dostat příkaz ‘to get an order’

• light verbs: syntactic center of CPs
• predicative nouns: semantic center of CPs

  both contribute to valency complements

Soldiers got from the commander guns
‘The soldiers got guns from the commander.’

Soldiers got from the commander the order to shoot
‘The soldiers were ordered to shoot by the commander.’
Introduction (2)

• CP: a redundant number of valency slots for the expression of semantic participants of the action denoted by the CP
  • Jane got something from her father
  • father's order to Jane to watch Jimmy
    ➢ Jane gets an order from her father to watch Jimmy

• some verbal and nominal complements corefer
  ➣ only several verbal and nominal complements are expressed on the surface

which complements of the light verb and which complements of the predicative noun should be present on the surface should be omitted from the surface

argument merger (Grimshaw, Mester, 1988)
argument fusion (Butt, 2010)
argument composition (Hinrichs et al., 1998)
Alonso Ramos (2007) ...
Goals

- formulate principles governing the distribution of valency complements of light verbs and predicative nouns in the surface structure
- verify these principles on well-formed structures

Outline

1. Annotation principles of CPs in the Prague Dependency Treebank
2. Manual analysis → principles of the distribution of valency complements
3. Automatic experiment verifying the proposed principles on the data from the PDT and PCEDT
Complex Predicates in PDT (1)

2,778 CPs in 2,558 sentences
1,695 CPs of a central type* -- only these were studied

* light verb in active form; predicative noun as its direct prepositionless object

A node PRED represents the light verb

A node CPHR represents the predicative noun

Jana dostala od otce příkaz pohlídat mladšího bratra.

‘Jane got from father the order to watch her younger brother.’

Language Resources and Evaluation Conference
Portorož, May 25, 2016
Distinguishing verbal and nominal complements

*dostat* ‘to get’:

\[ \text{ACT}_1 \quad \text{CPHR}_4 \quad ?\text{ORIG}_{od+2,z+2} \]

*příkaz* ‘order’:

\[ \text{ACT}_{2,u} \quad \text{ADDR}_3 \quad \text{PAT}_{k+3,f,aby,ať,že,c} \]

**Jane \text{got} from father order to watch younger brother**

‘Jane was ordered by her father to watch her younger brother.’
Distinguishing verbal and nominal complements

\( \textit{dostat} \) ‘to get’:

\[
\begin{align*}
\text{ACT}_1 & \quad \text{CPHR}_4 & \quad \text{?ORIG}_{\text{od+2},z+2} \\
\text{Jana}_1 & \quad \textit{dostala} & \quad \textit{od otce}_{\text{od+2}} & \quad \textit{příkaz}_4 & \quad \text{pohlídat}_f & \quad \textit{mladšího bratra}.
\end{align*}
\]

‘Jane was ordered by her father to watch her younger brother.’

\( \textit{příkaz} \) ‘order’:

\[
\begin{align*}
\text{ACT}_{2,u} & \quad \text{ADDR}_3 & \quad \text{PAT}_{k+3,f,aby,ať,že,c} \\
\text{Jane} & \quad \textit{got} & \quad \text{from father} & \quad \textit{order} & \quad \text{to watch} & \quad \text{younger brother}
\end{align*}
\]
Distinguishing verbal and nominal complements

**dostat** ‘to get’:

\[
\begin{array}{|c|c|c|}
\hline
\text{ACT}_1 & \text{CPHR}_4 & \text{?ORIG}_{od+2,z+2} \\
\hline
\end{array}
\]

Jane got from father order to watch younger brother

‘Jane was ordered by her father to watch her younger brother.’

**příkaz** ‘order’:

\[
\begin{array}{|c|c|c|}
\hline
\text{ACT}_{2,u} & \text{ADDR}_{3} & \text{PAT}_{k+3,f,aby,ať,že,c} \\
\hline
\end{array}
\]

Ellipsis of Valency Complements

- #QCor: an omitted actant is in grammatical coreference of quasi control with a coreferred element (characteristic of CPs)
- #PersPron
- #Gen
- #Oblfm
- ...

Language Resources and Evaluation Conference

Portorož, May 25, 2016
Complex Predicates in PDT (3)

*dostat* ‘to get’:

\[ \text{ACT}_1 \]

\[ \text{CPHR}_4 \]

\[ ?\text{ORIG}_{od+2,z+2} \]

*příkaz* ‘order’:

\[ \text{ACT}_{2,u} \]

\[ \text{ADDR}_3 \]

\[ \text{PAT}_{k+3,f,aby,at',že,c} \]

\[ Jana_1 \text{ dostala od otce}_{od+2} \text{ příkaz}_4 \text{ pohlídat}_f \text{ mladšího bratra}. \]

\[ \text{Jane got from father order to watch younger brother} \]

‘Jane was ordered by her father to watch her younger brother.’
Jana dostala od otce příkaz pohlídat mladšího bratra.
"Jane was ordered by her father to watch her younger brother."
Complex Predicates in PDT (3)

**dostat** ‘to get’:

**příkaz** ‘order’:

\[
\begin{align*}
\text{Jana}_1 & \quad \text{dostala} & \quad \text{od otce}_{od+2} & \quad \text{příkaz}_4 & \quad \text{pohlídat}_f & \quad \text{mladšího bratra}. \\
\text{Jane} & \quad \text{got} & \quad \text{from father} & \quad \text{order} & \quad \text{to watch} & \quad \text{younger brother} \\
\text{‘Jane was ordered by her father to watch her younger brother.’}
\end{align*}
\]
Manual Analysis (1)

<table>
<thead>
<tr>
<th></th>
<th>40% of PDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPs</td>
<td>659</td>
</tr>
<tr>
<td>All complements</td>
<td>2,034</td>
</tr>
<tr>
<td>Verbal Complements</td>
<td>796</td>
</tr>
<tr>
<td>Nominal Complements</td>
<td>1,238</td>
</tr>
</tbody>
</table>

Principles of the distribution of valency complements

(i) From the valency frame of the light verb:
   • such complements are expressed on the surface that are in the coreference of quasi control with any nominal ones

(ii) From the valency frame of the predicative noun:
   • those complements are expressed on the surface that are not in the coreference of quasi control with any verbal ones
Manual Analysis (2)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>All complements</td>
<td>1,873</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>92.1 %</td>
<td>7.9 %</td>
</tr>
<tr>
<td>Verbal Complements</td>
<td>653</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>32.1 %</td>
<td>7.0 %</td>
</tr>
<tr>
<td>Nominal Complements</td>
<td>1,220</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>60.0 %</td>
<td>0.9 %</td>
</tr>
</tbody>
</table>

• verbal complement CPHR was not counted (it is always present)
• valency complements subject to other types of ellipsis were treated as expressed on the surface
161 valency complements do not comply with the proposed principles due to
• annotation errors
• instigator of the light verb

Televize\textsubscript{V:Instigator:ACT} dává i další příležitosti k podnikání.
‘Television gives even more opportunities for business.’

Instigators violate principle (i)
Experiment (1)

Revised principles of the distribution of valency complements

(i) From the valency frame of the light verb:
   • **ACT** (regardless of its coreference of quasi control) is **always** expressed
   • such complements are expressed on the surface that **are** in the coreference of quasi control with any **nominal** ones

(ii) From the valency frame of the predicative noun:
   • those complements are expressed on the surface that **are not** in the coreference of quasi control with any **verbal** ones

<table>
<thead>
<tr>
<th></th>
<th>60% of PDT</th>
<th>PCEDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPs</td>
<td>1,036</td>
<td>2,116</td>
</tr>
<tr>
<td>All complements</td>
<td>3,264</td>
<td></td>
</tr>
<tr>
<td>Verbal complements</td>
<td>1,293</td>
<td>2,649</td>
</tr>
<tr>
<td>Nominal complements</td>
<td>1,971</td>
<td>N/A</td>
</tr>
</tbody>
</table>
1. identification of a CP

\[ \text{CP} \approx \text{dostat příkaz} \]

‘to get an order’
1. identification of a CP

CP ≈ dostat půříkaz ‘to get an order’

2. extraction of valency complements within a CP

ACT₁ CPHR₄ ?ORIG od+2, z+2
dostat ‘get’

ACT₂,u ADDR₃ PAT k+3,f,aby,ať,že,c příkaz ‘order’
Experiment (2)

1. Identification of a CP

\[ \text{PDT} \rightarrow \text{PDT-Vallex} \]

\[ \text{dostat} \ 'to get' \]

\[ \text{příkaz} \ 'order' \]

\[ \text{CP} \approx \text{dostat příkaz} \]

\[ \text{to get an order} \]

2. Extraction of valency complements within a CP

\[ \text{ACT}_1 \ 	ext{CPHR}_4 \ ?\text{ORIG}_{od+2,z+2} \]

\[ \text{dostat} \ 'get' \]

\[ \text{příkaz} \ 'order' \]

\[ \text{ACT}_2,u \ \text{ADDR}_3 \ \text{PAT}_{k+3,f,aby,at',že,c} \]

3. Coreference of quasi control

\[ \text{ACT}_1 \ 	ext{CPHR}_4 \ ?\text{ORIG}_{od+2,z+2} \]

\[ \text{ACT}_2,u \ \text{ADDR}_3 \ \text{PAT}_{k+3,f,aby,at',že,c} \]
Experiment (2)

1. identification of a CP
2. extraction of valency complements within a CP
3. coreference of quasi control
4. principles
5. verification

CP ≈ dostat příkaz ‘to get an order’

Language Resources and Evaluation Conference
Portorož, May 25, 2016
# Results

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>All complements</td>
<td>3,167</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>97.0%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Verbal complements</td>
<td>1,206</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>36.9%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Nominal complements</td>
<td>1,961</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>60.1%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

1. Table. Statistics on the distribution in PDT.

- the coreference of quasi control was not annotated in PCEDT

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal complements</td>
<td>2,116</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>79.9%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Verbal complements</td>
<td>1,252</td>
<td>68</td>
</tr>
<tr>
<td>(manually corrected)</td>
<td>94.8%</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

2. Table. Statistics on the distribution in PCEDT.
Conclusion

• 97.0% of valency complements follow the proposed principles
  • from a theoretical point of view:
    syntactic formation of CPs is a regular process
  • from a computational point of view:
    generation is possible

• what is necessary?
  • the information on valency complements both
    • of light verbs and
    • of predicative nouns
  • and at the same time the information on coreference

• in the future, extension to other languages