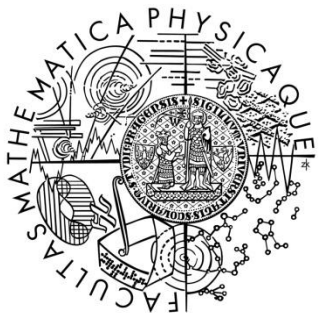


Distribution of Valency Complements in Czech Complex Predicates: Between Verb and Noun

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

Supported by the GA CR grant No. GA15-09979S and partially by the MEYS CR, grant No. LD14117.
Data provided by LINDAT/CLARIN, Infrastructural Project MEYS CR No. LM2015071.

Introduction (1)

- 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

Introduction (1)

- 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

Vojáci	*dostali*	*od velitele*	*zbraně.*	
Soldiers	got	from the commander	guns	

'The soldiers got guns from the commander.'

Vojáci	*dostali*	*od velitele*	*rozkaz*	*střílet.*
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 — 2 complements
  - father's **order** to Jane to watch Jimmy — 3 complements
  - ➢ Jane **gets an order** from her father to watch Jimmy — 3 complements together

- 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

Prague Dependency Treebank 3.0
PDT-Vallex
Prague Czech-English Dependency Treebank 2.0

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

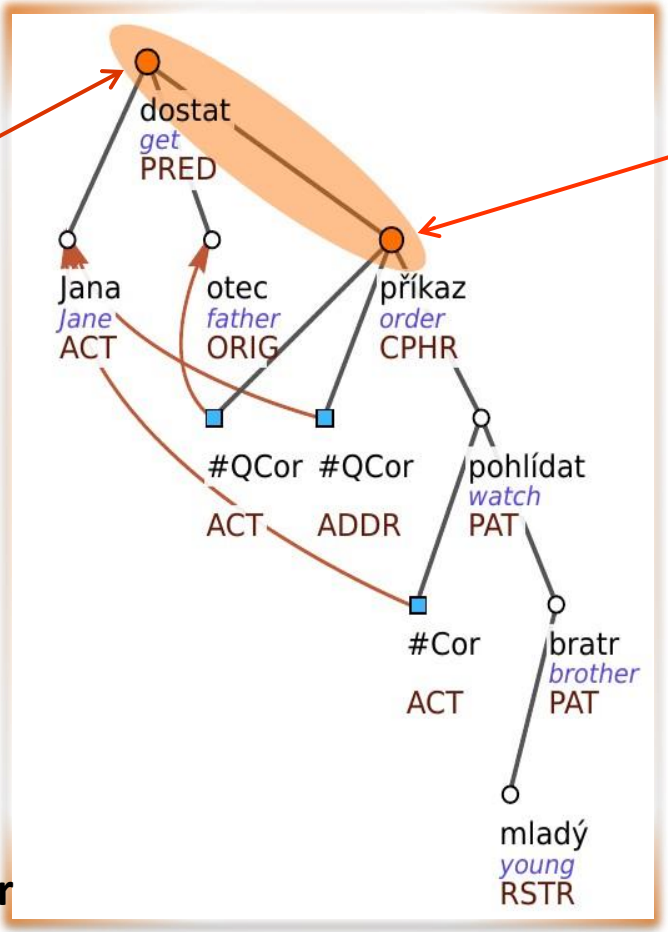
Complex Predicates in PDT (1)

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

$ACT_1\ CPHR_4\ ?ORIG_{od+2,z+2}$



a node CPHR
represents
the predicative noun

$ACT_{2,u}\ ADDR_3\ PAT_{k+3,f,aby,ať,že,c}$

Jana dostala od otce příkaz pohlídat mladšího bratra.

'Jane **got** from father the **order** to watch her younger brother.'

Complex Predicates in PDT (2)

Distinguishing verbal and nominal complements

dostat ‘to get’:

ACT_{1} $CPHR_{4}$ $?ORIG_{od+2,z+2}$

příkaz ‘order’:

$ACT_{2,u}$ $ADDR_{3}$ $PAT_{k+3,f,aby,ať,že,c}$

Jana$_{1}$	***dostala***	*od otce*$_{od+2}$	***příkaz***$_{4}$	*pohlídat*$_{f}$	*mladšího bratra.*
Jane	**got**	from father	**order**	to watch	younger brother

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

Complex Predicates in PDT (2)

Distinguishing verbal and nominal complements

dostat ‘to get’:

ACT$_{1}$ CPHR$_{4}$?ORIG$_{od+2,z+2}$

příkaz ‘order’:

ACT$_{2,u}$ ADDR$_{3}$ PAT$_{k+3,f,aby,ať,že,c}$

Jana$_{1}$	***dostala***	*od otce*$_{od+2}$	***příkaz***$_{4}$	*pohlídat*$_{f}$ *mladšího bratra.*
Jane	**got**	from father	**order**	to watch younger brother

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

Complex Predicates in PDT (2)

Distinguishing verbal and nominal complements

dostat 'to get':

ACT$_{1}$ CPHR$_{4}$?ORIG$_{od+2,z+2}$

příkaz 'order':

ACT$_{2,u}$ ADDR$_{3}$ PAT$_{k+3,f,aby,ať,že,c}$

Jana$_{1}$	***dostala***	*od otce*$_{od+2}$	***příkaz***$_{4}$	*pohlídat*$_{f}$	*mladšího bratra.*
Jane	**got**	from father	**order**	to watch	younger brother

'Jane was ordered by her father to watch her younger brother.'

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
- …

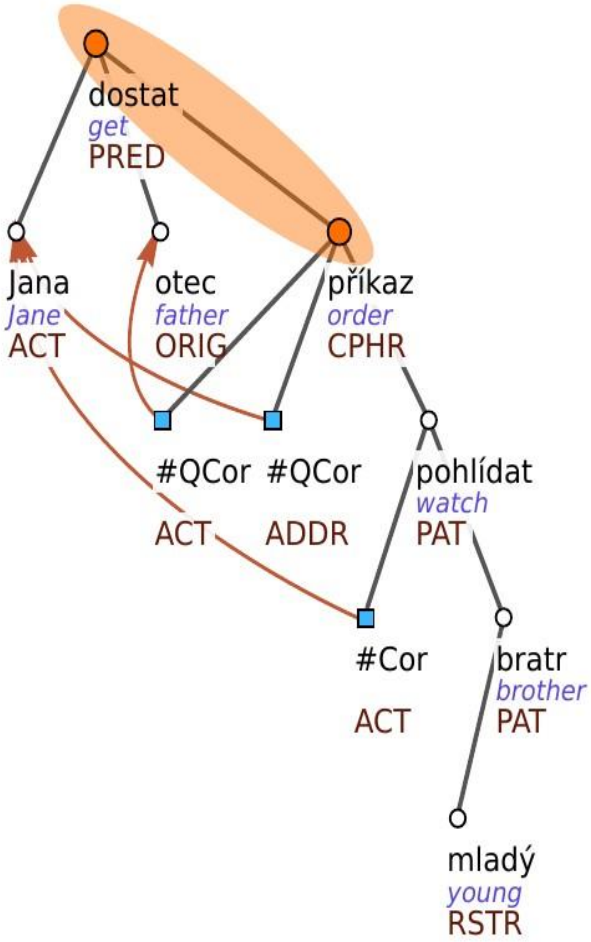
Complex Predicates in PDT (3)

dostat ‘to get’:

ACT_{1}

$CPHR_{4}$

$?ORIG_{od+2,z+2}$



příkaz ‘order’:

$ACT_{2,u}$

$ADDR_{3}$

$PAT_{k+3,f,aby,ať,že,c}$

Jana$_{1}$	*dostala*	*od otce*$_{od+2}$	*příkaz*$_{4}$	*pohlídat*$_{f}$	*mladšího bratra.*
Jane	**got**	from father	**order**	to watch	younger brother

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

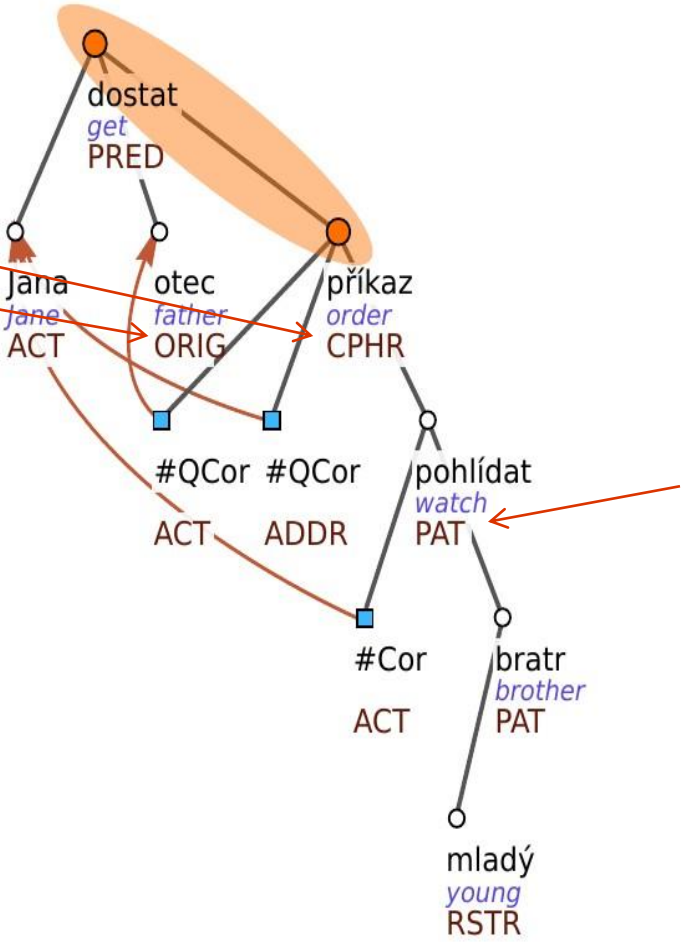
Complex Predicates in PDT (3)

dostat ‘to get’:

ACT$_{1}$

CPHR$_{4}$

?ORIG$_{od+2,z+2}$



příkaz ‘order’:

ACT$_{2,u}$

ADDR$_{3}$

PAT$_{k+3,f,aby,ať,že,c}$

Jana$_{1}$	*dostala*	*od otce$_{od+2}$*	*příkaz$_{4}$*	*pohlídat$_{f}$*	*mladšího bratra.*
Jane	**got**	from father	**order**	to watch	younger brother

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

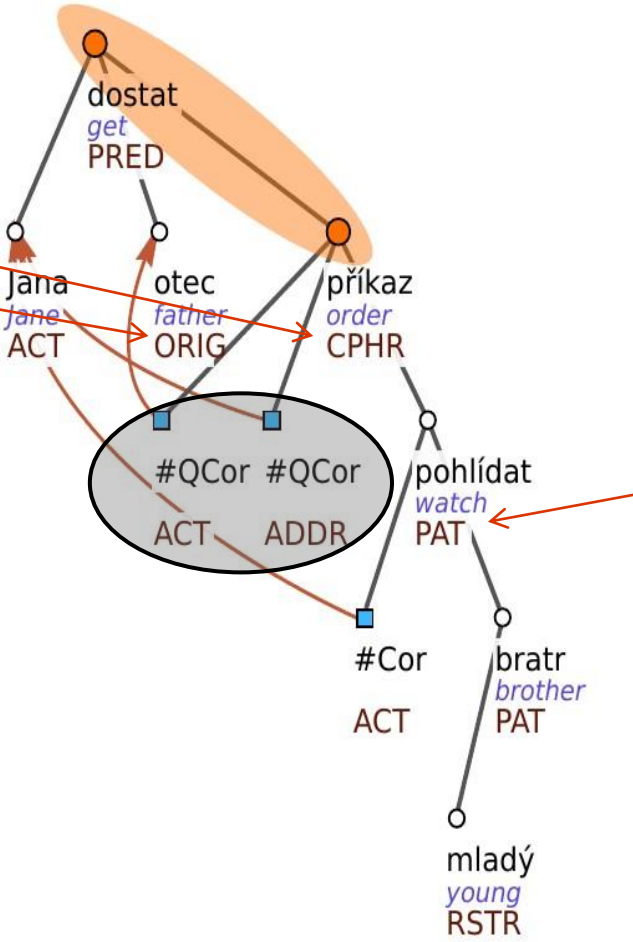
Complex Predicates in PDT (3)

dostat ‘to get’:

ACT$_{1}$

CPHR$_{4}$

?ORIG$_{od+2,z+2}$



příkaz ‘order’:

~~ACT$_{2,u}$~~

~~ADDR$_{3}$~~

PAT$_{k+3,f,aby,ať,že,c}$

Jana$_{1}$	*dostala*	*od otce$_{od+2}$*	*příkaz$_{4}$*	*pohlídat$_{f}$*	*mladšího bratra.*
Jane	**got**	from father	**order**	to watch	younger brother

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

Manual Analysis (1)

	40% of PDT
CPs	659
All complements	2,034
Verbal Complements	796
Nominal Complements	1,238

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)

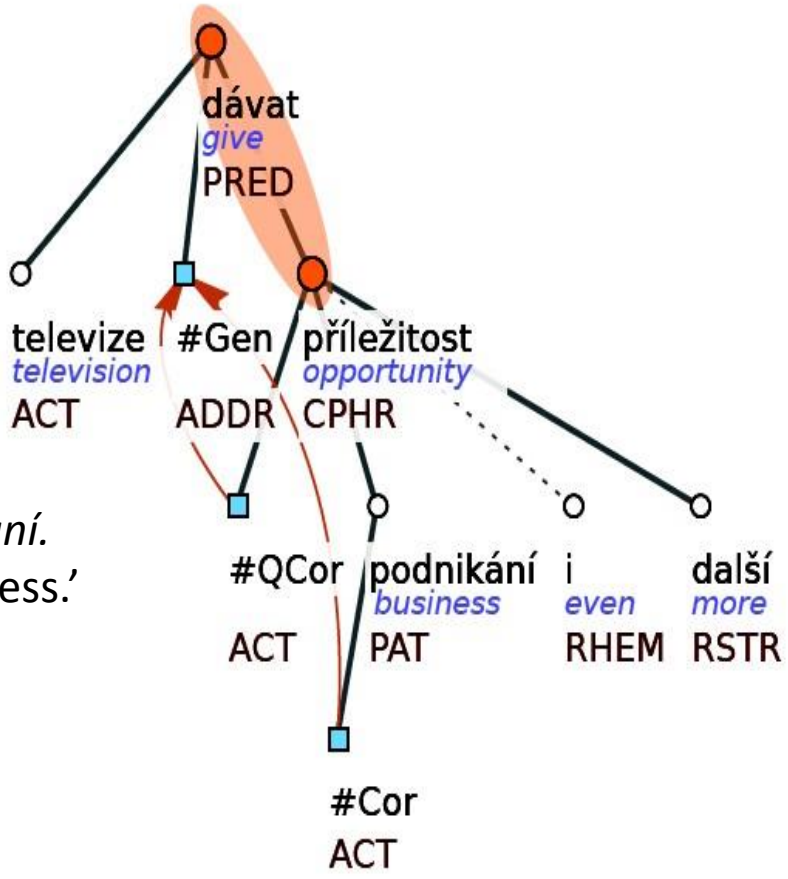
Distribution	Correct	Incorrect
All complements	1,873 92.1 %	161 7.9 %
Verbal Complements	653 32.1 %	143 7.0 %
Nominal Complements	1,220 60.0 %	18 0.9 %

- 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

Manual Analysis (3)

161 valency complements do not comply with the proposed principles due to

- annotation errors
- instigator of the light verb



Televize$_{V:\text{Instigator:ACT}}$ *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:

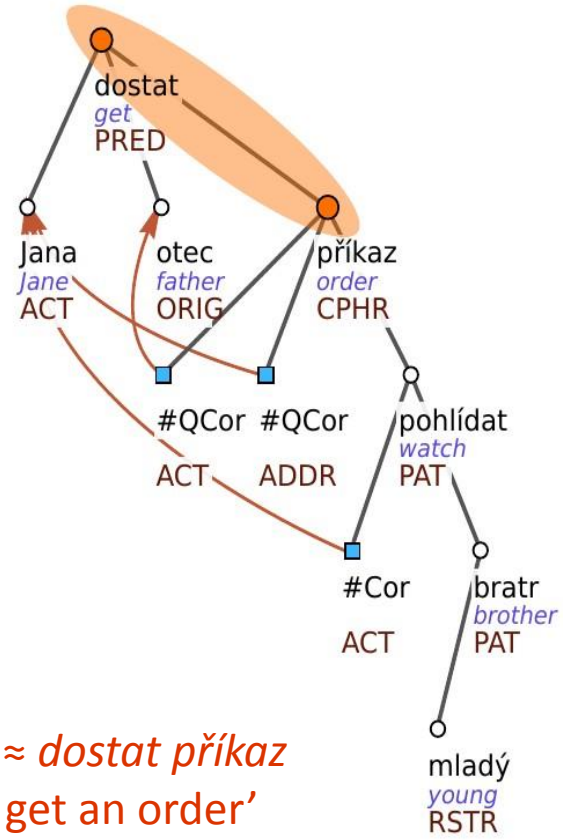
- those complements are expressed on the surface that **are not** in the coreference of quasi control with any **verbal** ones

	60% of PDT	PCEDT
CPs	1,036	2,116
All complements	3,264	
Verbal complements	1,293	2,649
Nominal complements	1,971	N/A

Experiment (2)

1. identification of a CP

PDT



CP ≈ *dostat příkaz*
'to get an order'

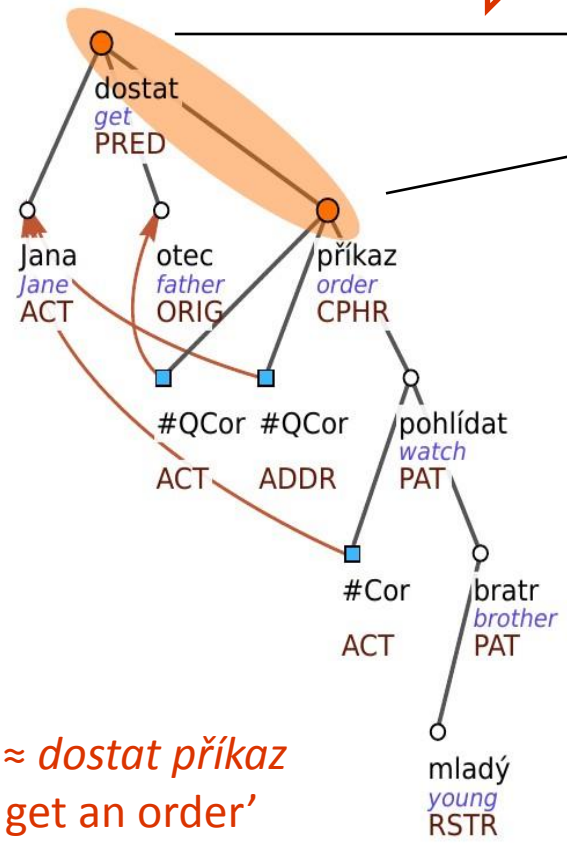
Experiment (2)

1. identification of a CP

PDT

PDT-Vallex

2. extraction of valency complements within a CP



ACT_1 $CPHR_4$ $?ORIG_{od+2,z+2}$ *dostat* ‘get’

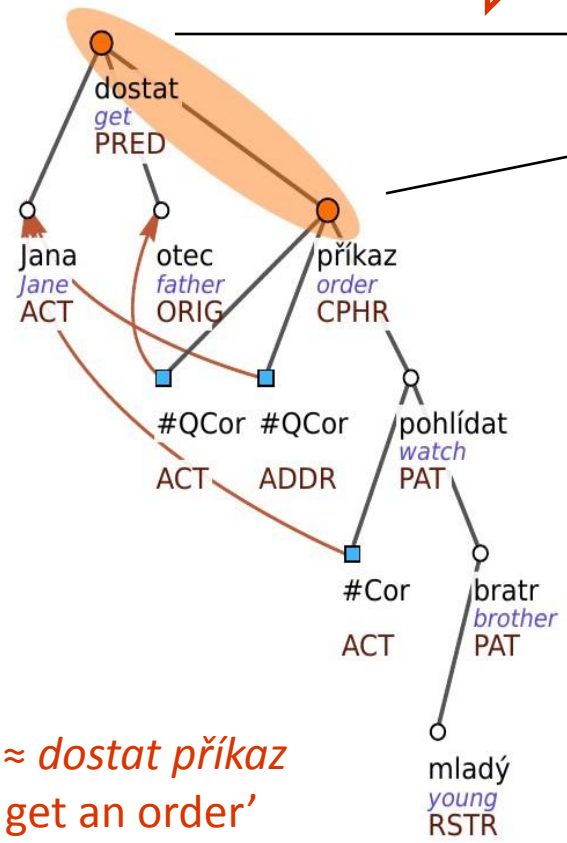
$ACT_{2,u}$ $ADDR_3$ $PAT_{k+3,f,aby,ať,že,c}$ *příkaz* ‘order’

CP ≈ *dostat příkaz*
‘to get an order’

Experiment (2)

1. identification of a CP

PDT → PDT-Vallex



CP ≈ *dostat příkaz*
'to get an order'

2. extraction of valency complements within a CP

$ACT_1\ CPHR_4\ ?ORIG_{od+2,z+2}$ *dostat* 'get'

$ACT_{2,u}\ ADDR_3\ PAT_{k+3,f,aby,ať,že,c}$ *příkaz* 'order'

3. coreference of quasi control

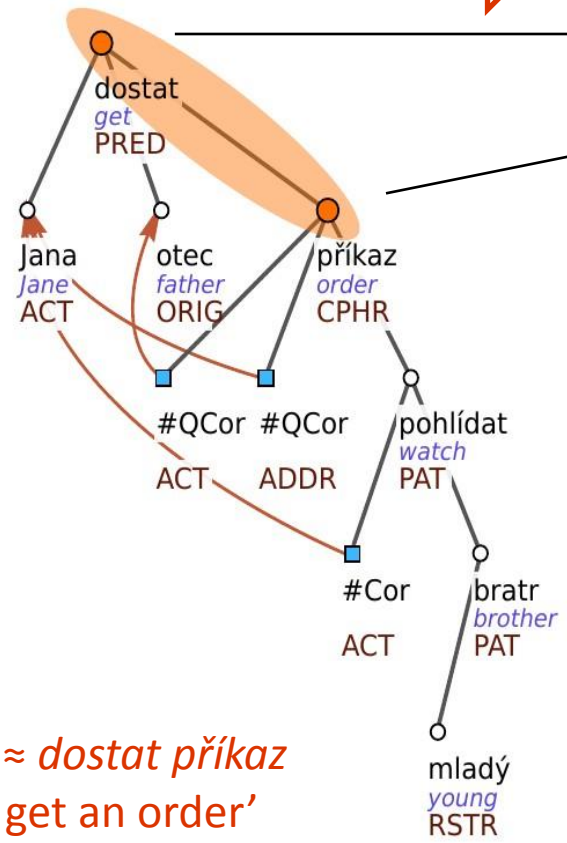
$ACT_1\ CPHR_4\ ?ORIG_{od+2,z+2}$

$ACT_{2,u}\ ADDR_3\ PAT_{k+3,f,aby,ať,že,c}$

Experiment (2)

1. identification of a CP

PDT → PDT-Vallex



CP ≈ *dostat příkaz*
'to get an order'

2. extraction of valency complements within a CP

ACT_1 $CPHR_4$ $?ORIG_{od+2,z+2}$ *dostat* 'get'

$ACT_{2,u}$ $ADDR_3$ $PAT_{k+3,f,aby,ať,že,c}$ *příkaz* 'order'

3. coreference of quasi control

ACT_1 $CPHR_4$ $?ORIG_{od+2,z+2}$

$ACT_{2,u}$ $ADDR_3$ $PAT_{k+3,f,aby,ať,že,c}$

4. principles

$+ACT_1$ $+CPHR_4$ $+?ORIG_{od+2,z+2}$

$-ACT_{2,u}$ $-ADDR_3$ $+PAT_{k+3,f,aby,ať,že,c}$

5. verification

5 out of 5 correct

Results

Distribution	Correct	Incorrect
All complements	3,167 97.0%	97 3.0%
Verbal complements	1,206 36.9%	87 2.7%
Nominal complements	1,961 60.1%	10 0.3%

1. Table. Statistics on the distribution in PDT.

- the coreference of quasi control was not annotated in PCEDT

Distribution	Correct	Incorrect
Verbal complements	2,116 79,9%	533 20,1%
Verbal complements (manually corrected)	1,252 94.8%	68 5.2%

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