Verbal Valency Frame Detection and Selection in Czech and English

Ondřej Dušek, **Jan Hajič**, and **Zdeňka Urešová** {odusek,hajic,uresova}@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague

Valency Frames as Event Mentions

- valency frames: detailed sense description
- valency frame detection (def.) =
  detection & sense disambiguation
  of event mentions
- our system: (auto) valency frame detection
  for Czech and English
- leveraging common annotation style
  of **Prague Dependency Treebanks**
  - PDT 2.5: Czech, PCEDT 2.0: Czech & English
  - based on Functional Generative Description (Tectogrammatics)

Valency in the Functional Generative Description (Tectogrammatics)

- a specific ability of a word to combine with other lexical units (event participants)
- argument-adjunct distinction:

inner participant (argument) × free modifier (adjunct)

specific for particular verbs
can appear only once

ACT, PAT, ADDR...

can modify any verb
can be repeated

MANN, ACMP, MEANS...

obligatory × optional participant

dialogue test:

obligatory = must be known to the speaker

Valency Frame – one per word sense

- all inner participants + obligatory modifiers
- semantic function (functor), obligatoriness, surface form

Automatic Valency Frame Selection

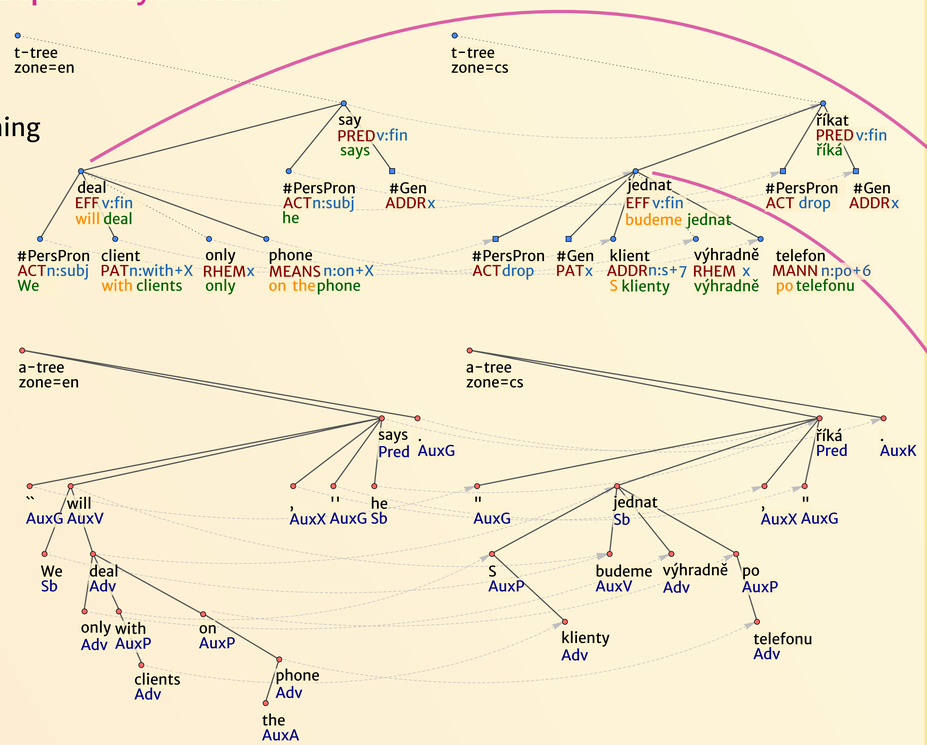
- based on **logistic regression**
- using the **same feature templates** for both languages:
  - word forms
  - morphology
  - formemes
  - a-layer dependency labels
  - neighborhood (topological & syntactic)
- trained on PDT 2.5 (Czech) & PCEDT 2.0 (English)
- **backoff to valency lexicons**
- use: included in the automatic t-layer annotation pipeline

Results

- unlabeled = detecting that a valency frame should be selected
- labeled = with disambiguation, i.e., finding the correct frame ID

	Czech	English
unlabeled precision	99.09	96.03
unlabeled recall	94.81	93.07
unlabeled F1 (detection only)	96.90	94.53
labeled precision	78.38	81.58
labeled recall	74.99	79.06
labeled F1 (detect & disamb.)	**76.65**	**80.30**
frame selection accuracy	79.10	84.95
ambiguous verbs baseline	66.68	68.44
ambiguous verbs classifier	72.41	80.03

Annotation of Prague Dependency Treebanks



t-layer: linguistic meaning

- nodes for lexical words only
- functors (semantic function)
- formemes (syntactic function)
- valency frame reference

a-layer: dependency syntax

m-layer: morphology

``	We	will	only	deal	with	clients	on	the	,	phone	''	he	says	.	"	S	klienty	budeme	jednat	výhradně	po	telefonu	,	"	říká	.
``	we	will	only	deal	with	client	on	the	,	phone	''	he	say	.	"	s	klient	být	jednat	výhradně	po	telefon	,	"	říkat	.
``	PRP	MD	RB	VB	IN	NNS	IN	DT	,	NN	''	PRP	VBZ	.	Z:	RR7	NNMP7	VB-P1FA	Vf	Dg		RR6	NNIS6	Z:	Z:	VB-S3PA	Z:

w-layer: tokenized text

``We will only deal with clients on the phone,'' he says.

"S klienty budeme jednat výhradně po telefonu," říká.

PDT-Vallex & EngVallex

- lexicons interlinked with treebanks

EngVallex: English, based on PropBank
7,699 frames for 4,337 verbs

deal

deal[1] ACT() PAT() ?ADDR()

- *But the computer-guided selling in response to those developments dealt a serious blow to the over-the-counter market, Mr. DaPuzzo said.*

deal[2] ACT() PAT()

(handle, deal with: deal with)

- *By contrast, Value Line said Georgia-Pacific ``is in a comparatively good position *trace* to deal with weakening paper markets,'' ...*

deal[3] ACT() PAT()

(handle, deal with: deal in)

- *The idea was to let small investors, the backbone of the fund business, deal in the money market's high short-term interest rates.*

PDT-Vallex: Czech, created with PDT
9,191 frames for 5,510 verbs

jednat

jednat[1] 161x,85x ACT(1) PAT(0+6;v+6;na-1[téma,4]) ADDR(5+7)

(smlouvat, hovořit) • *jedná s nimi o investicích; j. v této věci; parlament j. o nových zákonech; ministři spolu.MANN j. Rcp. ADDR o novele*

jednat[2] ACT(1) PAT(0+6)

(pojednávat, týkat se) • *román jedná o lásce*

jednat[3] 8x,5x ACT(1) PAT(5+7) MANN()|ACMP()|CRIT()|CPR()

(zacházet) • *jedná s ní špatně.MANN; j. s ním podle pravidel. CRIT; j. s námi bez servítků.ACMP; j. s ním šalamounsky.CPR.*

jednat[4] 22x,42x ACT(1) BEN()|MANN()|ACMP()|CRIT()|CPR()|AIM()

(chovat se, postupovat) • *začal jednat zbrkle.MANN; j. podle regulí.CRIT; j. proti rozhodnutí úřadu.BEN; j. v zájmu zákonného postupu.BEN; j. s razancí.ACMP a bez diskutování.ACMP; j. otrocky.CPR; j. v zájmu zákonného postupu.AIM*

Conclusions

- first results in valency frame selection on Prague Dependency Treebanks using automatic annotation
- encouraging, given the high granularity of the lexicons
- detecting frames – easier for Czech, selecting the correct one – English better
- problems:
  - idiomatic expressions
  - frames with identical set of members

Supported by the Grant No. GPP406/13/03351P of the Grant Agency of the Czech Republic and the project LH12093 of the Ministry of Education, Youth and Sports of the Czech Republic, and by the Charles University SVV project 260 104. Data provided by LINDAT/CLARIN, Infrastructural Project MEYS CR No. LM2010013.