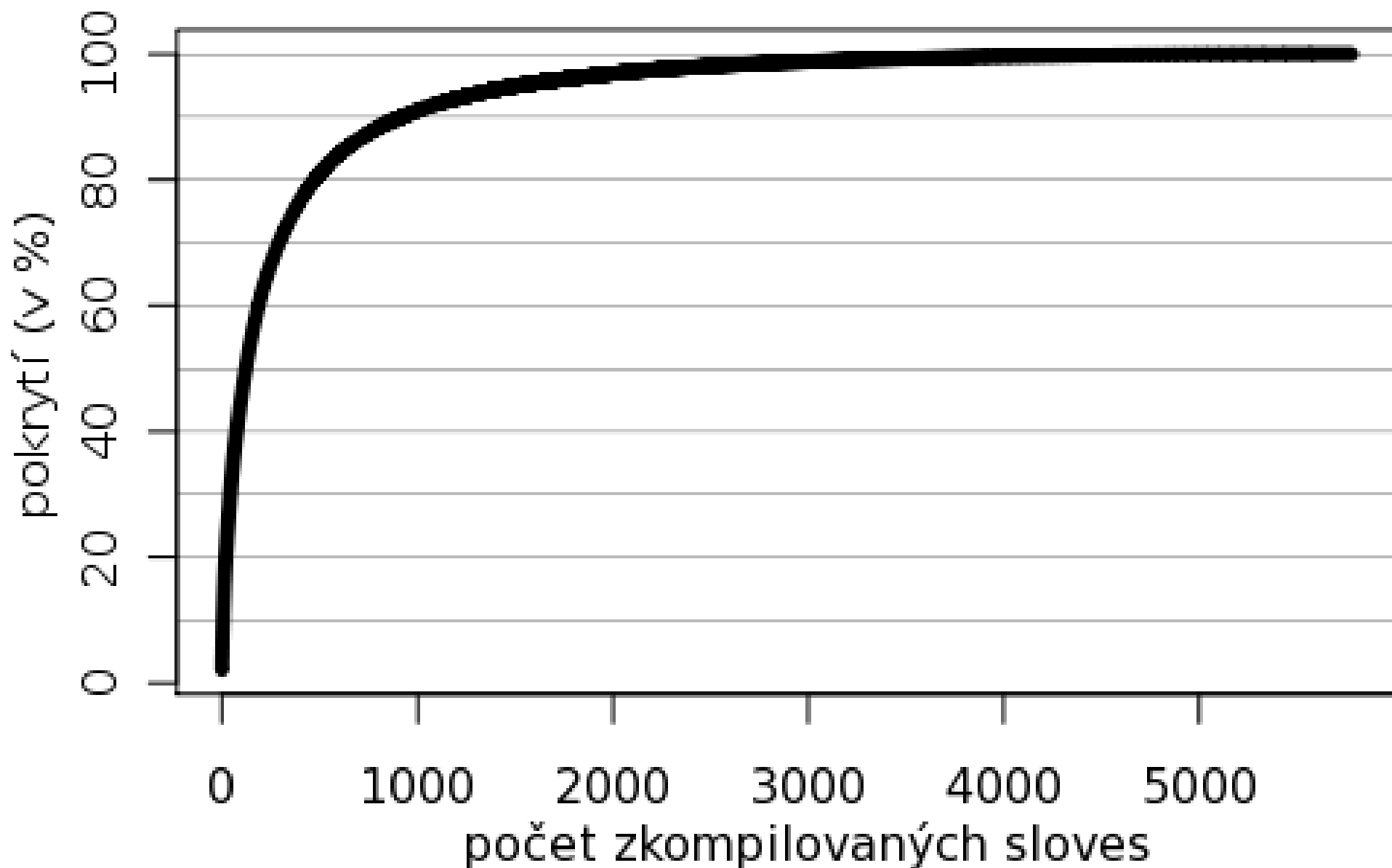


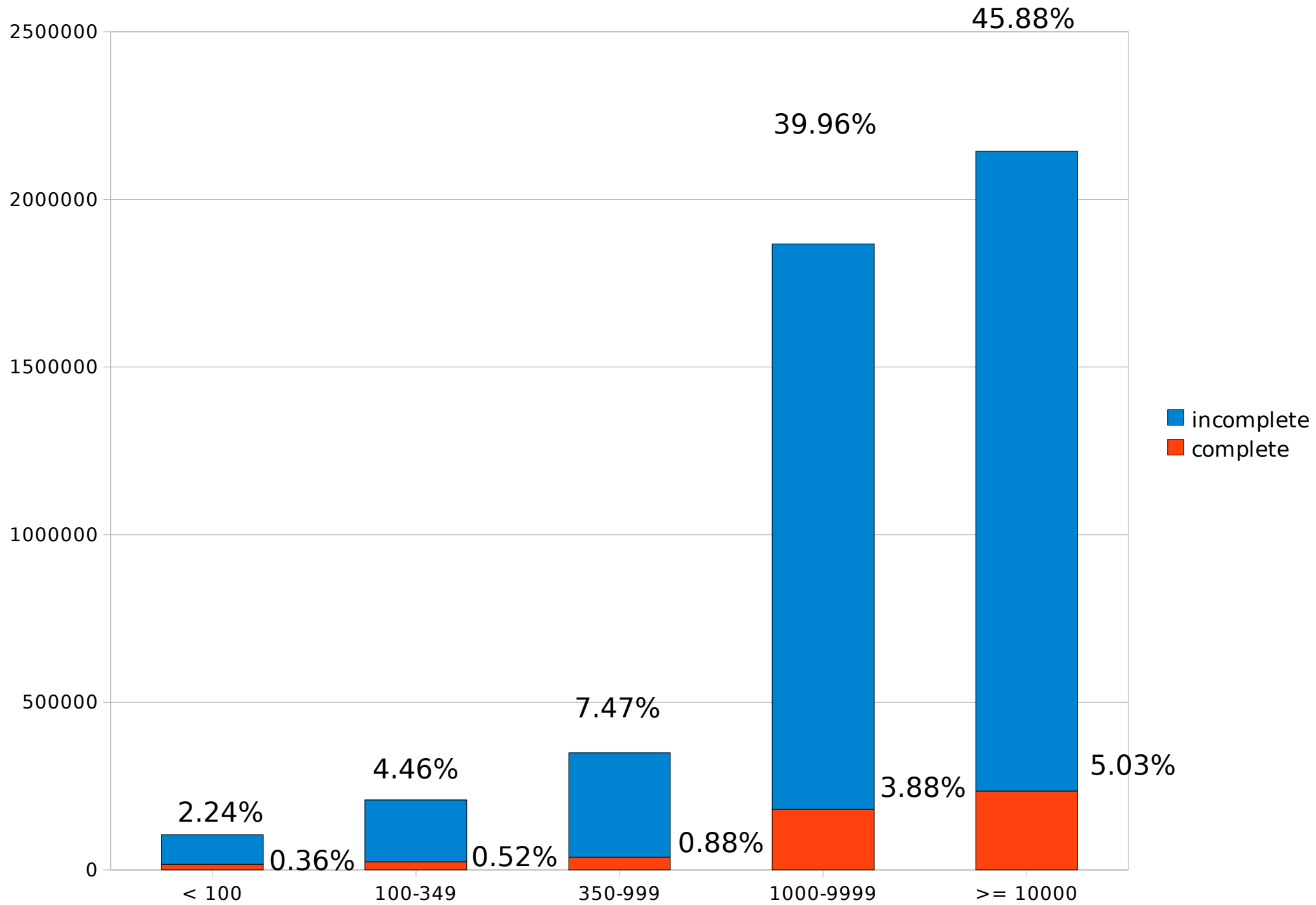
# Verbs in BNC50 - coverage



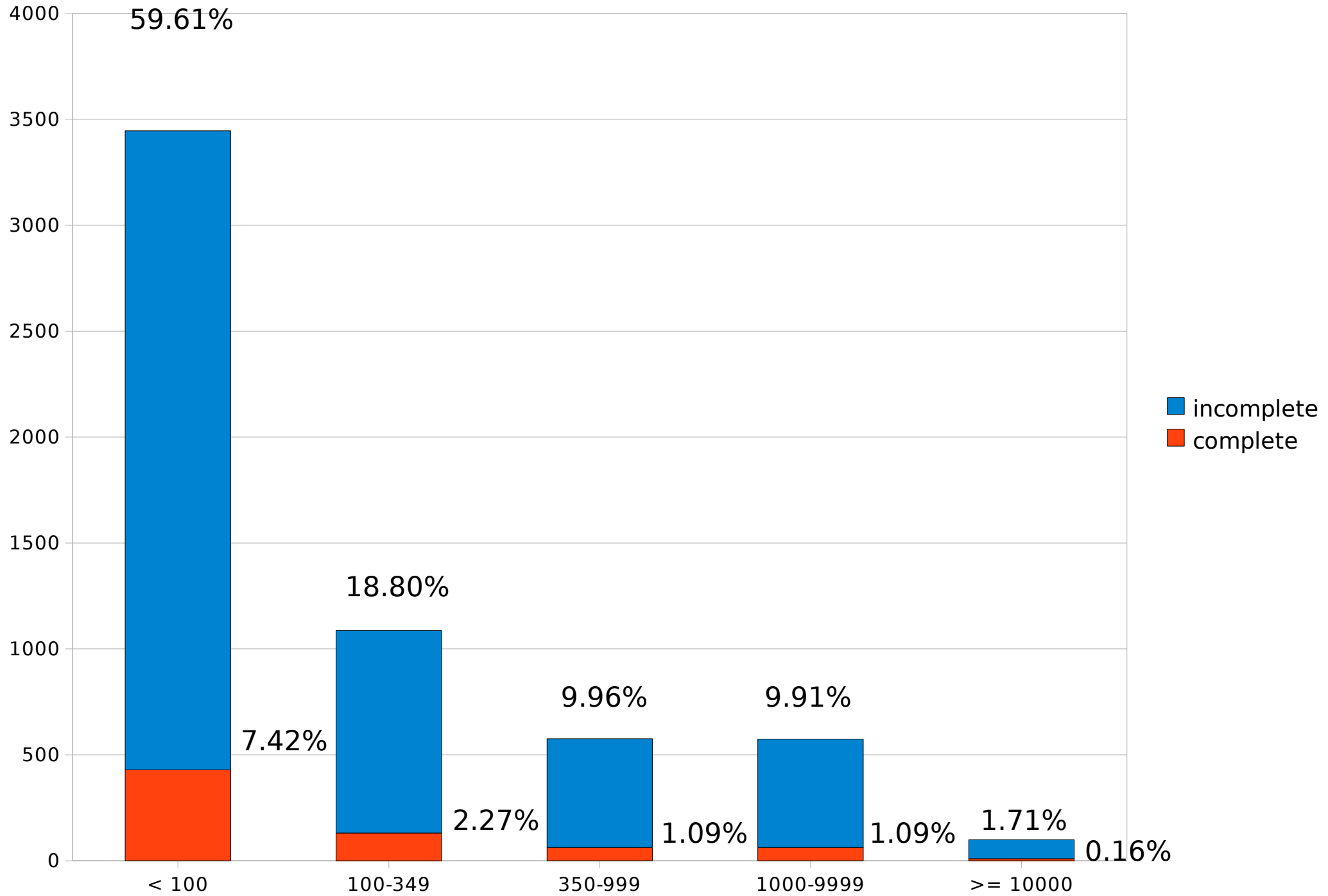
# Verbs in BNC50

	verb tokens (=frequency)			verb types		
	all	%	complete	all	%	complete
<b>&lt; 100</b>	104,564	2.24%	16,698	3,446	59.61%	429
<b>100 - 349</b>	208,466	4.46%	24,358	1,087	18.80%	131
<b>350 - 999</b>	348,945	7.47%	38,094	576	9.96%	63
<b>1,000 - 9,999</b>	1,867,147	39.96%	181,141	573	9.91%	63
<b>&gt;= 10,000</b>	2,143,971	45.88%	235,262	99	1.71%	9
<b>all</b>	4,673,093		495,553	5,781		695

# BNC50 coverage by verb tokens



# BNC50 coverage - types



# Inconsistencies in patterns

- makes patterns unparseable
- types:
  - semantic type and semantic role in one field:
    - abort(1) 'Human = Female'
  - semantic role as a grammatical feature:
    - beat(9) semantic role 'PLURAL'
  - inconsistent separator:
    - beat(1) 'savagely, severely, brutally'
    - hang(13) 'heavy | heavily'

# Inconsistencies in patterns II

- types (continue):
  - non-matching brackets:
    - forge(2) 'Institution2} and {Human 3'
  - complex structure in 'lexset':
    - accept(3) 'responsibility (for [[Eventuality | Entity]])'
    - beat(9) '[[Human 2]]'s door'
  - insufficient form structure:
    - console(3) semantic type '{that-CLAUSE}'
  - whole phrase in 'lexset':
    - bite(14) lexset (object) 'more than [PERSONAL PRON] can chew'

# Interannotator agreement

verb	N	P	BNC50 freq	A	Fleiss' kappa					
					all annotators		Patrick Hanks			
					+exp	-exp	#	+exp	#	-exp
accept	8	4.16	12,190	3	0.54	0.64	13	0.65	7	0.80
address	6	3.85	3,628	3	0.37	0.49	9	0.71	6	0.78
alter	2	1.28	2,489	3	0.59	0.78	3	0.74	3	0.74
argue	7	1.73	11,362	3	0.83	1.00	1	0.74	0	1.00
call	34	6.68	24,439	2	0.69	0.68	5	0.87	4	0.90
claim	6	3.14	12,517	4	0.69	0.72	8	0.72	5	0.82
fire	15	7.96	1,488	2	0.34	0.42	17	0.59	11	0.72
lead	12	3.97	20,180	3	0.72	0.78	10	0.74	7	0.81
need	5	4.12	28,352	3	0.64	0.66	6	0.84	6	0.83
plan	4	3.08	7,294	3	0.57	0.60	8	0.78	5	0.86
spoil	9	3.36	409	3	0.65	0.73	4	0.85	2	0.93
tell	21	3.80	21,550	2	0.27	0.27	9	0.74	5	0.84
visit	3	2.16	5,889	3	0.56	0.60	5	0.80	4	0.83

- size of each sample is 50 concordances in BNC50

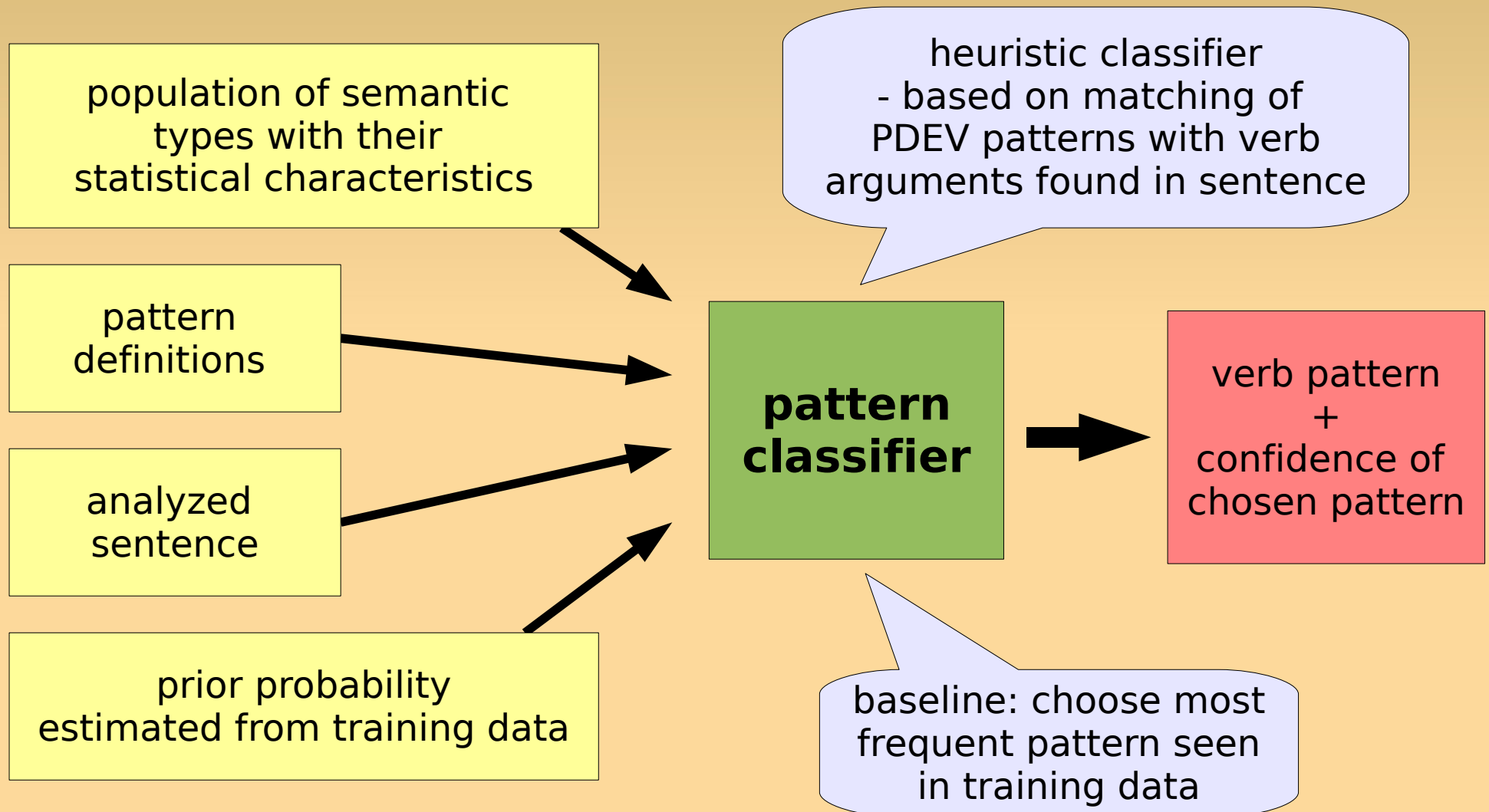
	verb	N	P	BNC50 freq	A	Fleiss' kappa					
						all annotators		Patrick Hanks			
						+exp	-exp	#	+exp	#	-exp
1	abstain	3	2.47	232	4	0.72	0.82	6	0.77	4	0.84
2	accept	8	4.16	12,190	3	0.54	0.64	13	0.65	7	0.80
3	address	6	3.85	3,628	3	0.37	0.49	9	0.71	6	0.78
4	admit	13	7.03	5,410	3	0.67	0.74	15	0.61	10	0.73
5	alter	2	1.28	2,489	3	0.59	0.78	3	0.74	3	0.74
6	announce	4	2.68	8,739	3	0.68	0.76	6	0.80	4	0.86
7	argue	7	1.73	11,362	3	0.83	1.00	1	0.74	0	1.00
8	call	34	6.68	24,439	2	0.69	0.68	5	0.87	4	0.90
9	claim	6	3.14	12,517	4	0.69	0.72	8	0.72	5	0.82
10	engage	9	4.82	2,985	2	0.58	0.66	9	0.76	6	0.84
11	explain	5	3.21	10,064	3	0.66	0.74	9	0.71	8	0.73
12	fire	15	7.96	1,488	2	0.34	0.42	17	0.59	11	0.72
13	lead	12	3.97	20,180	3	0.72	0.78	10	0.74	7	0.81
14	need	5	4.12	28,352	3	0.64	0.66	6	0.84	6	0.83
15	plan	4	3.08	7,294	3	0.57	0.60	8	0.78	5	0.86
16	rush	9	4.99	984	2	0.67	0.65	7	0.80	7	0.78
17	say	9	1.66	94,608	3	0.48	0.48	5	0.46	5	0.46
18	spoil	9	3.36	409	3	0.65	0.73	4	0.85	2	0.93
19	tell	21	3.80	21,550	2	0.27	0.27	9	0.74	5	0.84
20	visit	3	2.16	5,889	3	0.56	0.60	5	0.80	4	0.83



# Current interannotator agreement

	<b>date:</b>	30-11-2010					
	<b>verb</b>	<b>BNC50 freq</b>	<b>patterns</b>	<b>A2</b>	<b>A3</b>	<b>Cohen's kappa</b>	<b>reference sample</b>
1	access	455	4	50	50	0.416	250
2	accrue	347	3	50	50	0.020	208
3	breathe	950	17	50	50	0.617	250
4	crush	435	5	50	50	0.354	250
5	drain	500	15	50	50	0.386	250
6	enlarge	529	3	50	50	0.304	250
7	enlist	347	3	50	50	0.525	205
8	forge	511	8	50	50	0.604	248
9	frighten	305	4	50	50	0.441	246
10	furnish	396	4	50	50	0.351	250
11	hail	449	9	46	0		250
12	part	380	8	50	0		244
13	plug	346	9	50	50	0.551	249
14	pour	914	20	50	0		250
15	smash	530	8	50	0		243
16	smell	416	7	50	0		250
17	steer	432	7	50	0		203
18	swell	398	10	50	0		241
19	trouble	375	6	50	50	0.319	250
20	wake	909	5	50	0		250

# Automatic Pattern Classifier



# Automatic Pattern Classifier - experiments

verb	N	P	f	D	A <sub>B</sub>	A	Y	avg(c)	H	H <sub>Y</sub>	A <sub>H</sub>
abandon	8	4.22	2813	77	54.55	76.62	59	82.61	54	49	90.74
alter	2	1.28	2489	64	90.62	90.62	58	99.01	64	58	90.63
announce	4	2.68	8739	82	48.78	80.49	66	86.18	59	52	88.14
argue	7	1.73	11362	82	92.68	92.68	76	95.59	67	62	92.54
devote	6	4.65	1530	80	47.50	63.75	51	80.95	45	38	84.44
fire	15	7.96	1488	76	28.95	40.79	31	61.95	27	15	55.56
land	12	7.02	1496	66	40.91	59.09	39	67.09	27	18	66.67
need	5	4.12	28352	65	35.38	53.85	35	76.37	31	21	67.74
plan	4	3.08	7294	59	54.24	81.36	48	77.55	35	31	88.57
rush	9	4.99	984	277	53.07	65.34	181	71.14	113	69	61.06
sleep	11	2.44	1575	414	77.78	84.78	351	83.05	90	79	87.78
urge	6	2.84	2603	83	72.29	81.93	68	96.37	75	65	86.67
visit	3	2.16	5889	76	68.42	75.00	57	88.22	55	48	87.27

<b>celkem</b>				<b>2049</b>	<b>61.54</b>	<b>72.86</b>	<b>1493</b>		<b>1016</b>	<b>802</b>	<b>78.94</b>
---------------	--	--	--	-------------	--------------	--------------	-------------	--	-------------	------------	--------------

<b>sloveso</b>	<b>N</b>	<b>P</b>	<b>f</b>	<b> D </b>	<b>A<sub>B</sub></b>	<b>A</b>	<b> Y </b>	<b>avg(c)</b>	<b> H </b>	<b> H<sub>Y</sub> </b>	<b>A<sub>H</sub></b>
abandon	8	4.22	2813	77	54.55	76.62	59	82.61	54	49	90.74
abstain	3	2.47	232	74	58.11	71.62	53	76.88	26	21	80.77
address	6	3.85	3628	80	63.75	72.50	58	83.40	49	39	79.59
alter	2	1.28	2489	64	90.62	90.62	58	99.01	64	58	90.63
announce	4	2.68	8739	82	48.78	80.49	66	86.18	59	52	88.14
argue	7	1.73	11362	82	92.68	92.68	76	95.59	67	62	92.54
devote	6	4.65	1530	80	47.50	63.75	51	80.95	45	38	84.44
engage	9	4.82	2985	79	37.97	44.30	35	62.90	20	9	45.00
fire	15	7.96	1488	76	28.95	40.79	31	61.95	27	15	55.56
handle	6	3.35	2796	80	56.25	66.25	53	87.49	55	41	74.55
land	12	7.02	1496	66	40.91	59.09	39	67.09	27	18	66.67
need	5	4.12	28352	65	35.38	53.85	35	76.37	31	21	67.74
plan	4	3.08	7294	59	54.24	81.36	48	77.55	35	31	88.57
propose	8	3.80	4987	81	65.43	76.54	62	81.43	50	38	76.00
rush	9	4.99	984	277	53.07	65.34	181	71.14	113	69	61.06
signal	6	3.76	892	73	57.53	63.01	46	83.87	39	24	61.54
sleep	11	2.44	1575	414	77.78	84.78	351	83.05	90	79	87.78
treat	4	2.60	6455	81	71.60	81.48	66	82.27	35	25	71.43
urge	6	2.84	2603	83	72.29	81.93	68	96.37	75	65	86.67
visit	3	2.16	5889	76	68.42	75.00	57	88.22	55	48	87.27
<b>celkem</b>				<b>2049</b>	<b>61.54</b>	<b>72.86</b>	<b>1493</b>		<b>1016</b>	<b>802</b>	<b>78.94</b>