System of Semantic Types

- analyse sentences

- find verb arguments – agent, object, …

- match verb argument with semantic type at same position (use definitions of patterns)

- get pairs ({Semantic_types}, word)

- disambiguate the pairs

- create lists of words which realised semantic types

`` We 've devoted the project to social awareness .

devoted

nsubj / aux | dobj | prep_to

We | 've | project | awareness

det | amod

the | social
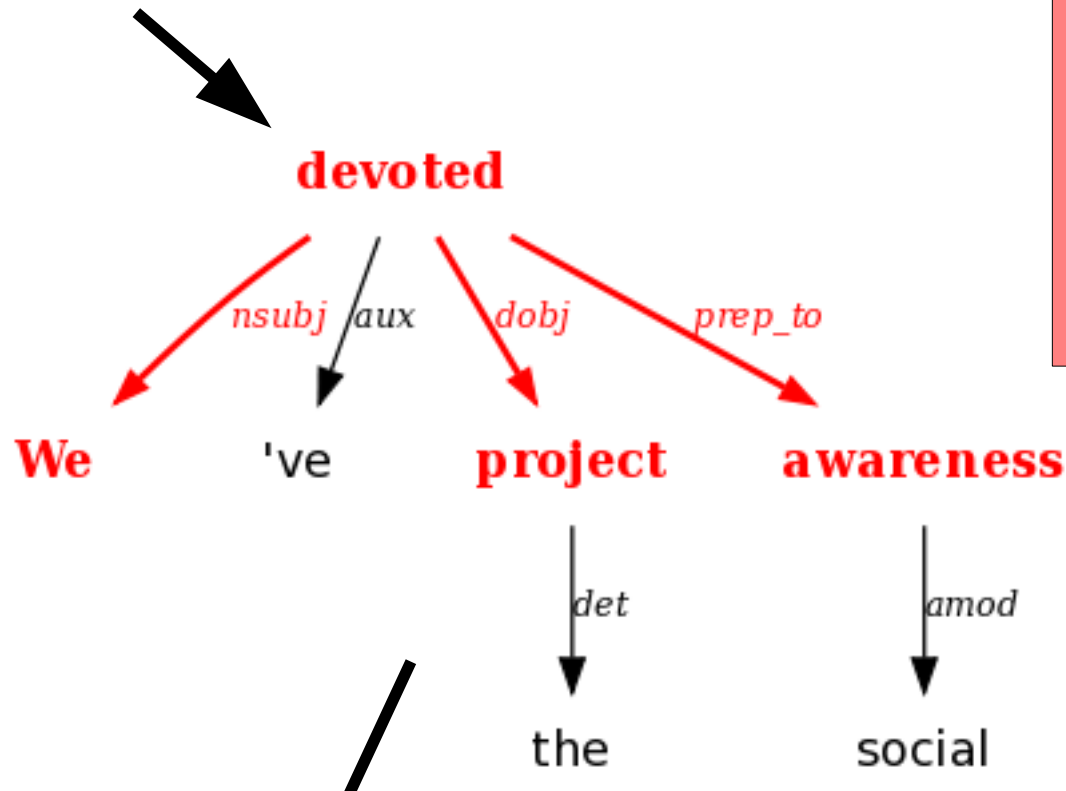
[[Human | Institution]]
  = We
[[Activity]]
  = project
[[Anything]]
  = awareness

agent: We
object: project
adv.: to awareness

+

devote [4]:
agent: [[Human | Institution]]
object: [[Activity]]
adverbial: to [[Anything]]

Our little programs I

- for filtering and sorting after nouns, semantic types or statistical characteristics of doubles (ST, N)

    - example:
      ```
      view_semtypes_lists.sh -file semtypes_lists.txt
      -f3 10 -pmi3 1.5 -sort semtype f3 pmi3
      ```

        - this prints only lines, where *f3* is greater or equal 10 and *pmi3* is greater or equal 1.5

        - the output will be sorted primarily by column *semtype*, than by column *f3* and finally by *pmi3*

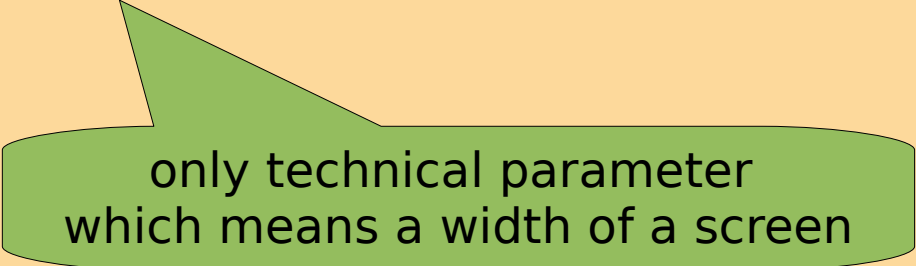# Program for comfortable viewing lists of STs and nouns

- **USAGE:**

  - -h | -help   print this help and exit

  - -file filename   use file with name
    filename as an input [must be set here or in script]

  - (-f1 | -f2 | -f3 | -f4) value   set value as a minimal
    value for frequency

  - (-pmi3 | -pmi4) value   set value as a minimal
    value for pointwise mutual information

  - (-uf1 | -uf2 | -uf3 | -uf4) value   set value as a maximal
    value for frequency

  - (-upmi3 | -upmi4) value   set value as a maximal
    value for pointwise mutual information

  - -sort column1 column2 ... columnN   result will be sorted
    primarily by column1, than by column2, ... and finally by
    columnN

- Columns are: noun, semtype, f1, f2, f3, pmi3, f4, pmi4

- Not depends on ordering of the arguments

Example: Semantic type „Plane"

noun	Semantic Type	F1	F2	F3	PMI3	F4	PMI4
aircraft	Plane	36	20	27.83	8.28	34.43	8.76
plane	Plane	21	12	16.5	8.68	20.31	9.15
helicopter	Plane	21	2	11.25	8.09	14.67	8.64
P-40	Plane	3	3	3	9.59	3	9.76
aeroplane	Plane	3	2	2.5	8.59	3	9.03
caravan	Plane	5	0	2.5	8.33	2.5	8.5
van	Plane	5	0	2.5	6.52	2.5	6.69
vehicle	Plane	5	0	2.5	6.06	2.5	6.23
Hurricanes	Plane	2	2	2	7.42	2	7.59
Sugar	Plane	2	2	2	8.59	2	8.76
glider	Plane	2	2	2	9.59	2	9.76
jet	Plane	3	1	2	7.42	3	8.18

Our little programs II

- for finding original sentences for each double (Semantic Type, Noun)

  - you can also use regular expressions

  - it is useful for detecting parser errors and non-standard constructions

- example:

  - `view_orig_sentences.sh $COLUMNS "Animal" "snake"`

  - `view_orig_sentences.sh $COLUMNS ".*" "snake"`

only technical parameter
which means a width of a screen

Annotation of doubles (ST, N)

- tagged doubles:

	count	%
T	3250	37.37%
M	4316	49.63%
C	1131	13.00%
total	8697	100.00%

Semantic Types	all	T	M	C
total	146	109	120	86
f >= 10	58	26	44	18
f >= 5	83	42	60	26
f >= 3	103	52	78	40
f > 1	118	68	93	51
f = 1	28	41	27	35

- how many different Semantic Types is tagged

The most frequent Semantic Types in tagged sample (size=8697)

2236	Human	85	Animate	18	Privilege
1570	Institution	64	Concept	17	Rule
965	Document	49	Process	17	Building
284	Anything	49	Information	17	Body Part
255	Eventuality	49	Abstract	16	Permission
251	Entity	42	Resource	15	Psych
219	Animal	36	Attitude	15	Drug
187	Activity	32	Device	13	Deity
170	Physical Object	30	Plan	13	Asset
169	Human Group	30	Emotion	12	Self
166	Event	28	Boat	12	Road Vehicle
161	Action	24	Money	12	Opportunity
149	Location	23	Liquid	11	Language
136	Information Source	22	Inanimate	10	Word
111	State	22	Computer	10	Plant
105	Proposition	21	Horse	10	Noise
102	Vehicle	20	Pace	10	Investigation
90	Agreement	19	Speech Act	10	Disease
89	Stuff	19	Property	8	Weapon
87	Artifact	19	Food	8	Route

The most frequent STs by tags

tag T		tag C		tag M	
1489	Human	293	Institution	1031	Institution
283	Anything	148	Document	764	Document
246	Institution	120	Human	627	Human
214	Entity	108	Information Source	201	Animal
156	Eventuality	45	Physical Object	108	Human Group
80	Location	38	Action	101	Activity
67	Physical Object	29	Human Group	93	Event
59	Activity	27	Activity	90	Vehicle
53	Document	23	Eventuality	82	Agreement
50	Event	23	Event	76	Eventuality
48	Action	20	Proposition	75	Action
40	Animate	17	State	71	Proposition
32	Human Group	16	Entity	69	State
27	Abstract	16	Concept	58	Stuff
25	State	14	Location	58	Physical Object
21	Process	14	Artifact	55	Location
20	Inanimate	12	Stuff	53	Artifact
20	Artifact	11	Plan	40	Animate
19	Stuff	9	Abstract	33	Resource
18	Concept	8	Information	30	Concept