WG2: Improving Parsing Medical Discourse using Cascades of Multiword Expression Recognition

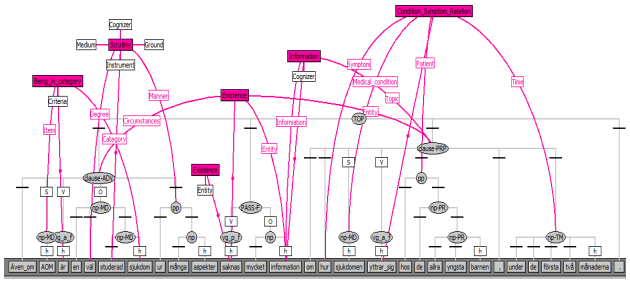
DIMITRIOS KOKKINAKIS

Dep. of Swedish, the Swedish Language Bank &
the Centre for Language Technology (CLT)
University of Gothenburg
Sweden
dimitrios.kokkinakis@svenska.gu.se

OUTLINE

…how a cascaded finite state constituent parser for Swedish behaves, with respect to parsing accuracy, where various types of MWE are automatically and successively recognized and introduced into the parser

```
<s id="id.149">
[MAIN-INF
  [RGOA  <t id="id.149_1"/> Bland_annat]
  [vg_
  h=[V@IPAS-AUX  <t id="id.149_2"/> har]]
SBJ=[np-pronoun
  h=[PF@NSO@S  <t id="id.149_3"/> det]]
V=[vg_a_i
  h=[V@IUAS  <t id="id.149_4"/> visat_sig]]]
[CSS  <t id="id.149_5"/> att]
[...]
<s id="id.230">
[MAIN-FIN
SBJ=[np-entity-work
    [NP00N@0S  <t id="id.230_1"/> The]
    [work-entity
      [NP00N@0S-w  <t id="id.230_2"/> Great]
      [NP00N@0S-w  <t id="id.230_3"/> Smoky]
      [NP00N@0S-w  <t id="id.230_4"/> Mountain]
    h=[NP00N@0S-w  <t id="id.230_5"/> Study]]]
V=[vg_a_f
  h=[V@IPAS  <t id="id.230_6"/> är]]
OBJ=[np
    [DI@US@S  <t id="id.230_7"/> en]
    [AQPUSNIS  <t id="id.230_8"/> amerikansk]
    [AQPUSNIS  <t id="id.230_9"/> longitudinell]
  h=[NCUSN@IS  <t id="id.230_10"/> befolkningsstudie]]]
[...]
```



	Precision	Recall	F-score
i*	77.92%	76.79%	77.35%
iii*	77.96%	76.82%	77.39%
iv*	80.71%	79.93%	80.32%
v*	86.67%	87.10%	86.89%