

## Corpus annotation formats

---

LIN 386.m  
Computational Linguistics 2  
Katrin Erk

## Examples: Part-of-speech tags (from the Penn Treebank)

---

[ An/DT example/NN ]	[ that/DT degree/NN ]
of/IN	of/IN
[ the/DT overall/JJ standards/NNS ]	[ cochannel/NN interference/NN ]
applied/VBN is/VBZ	
[ the/DT 20-to-1/CD ratio/NN ]	[ which/WDT ]
established/VBN for/IN	is/VBZ regarded/VBN as/IN
[ the/DT determination/NN ]	[ objectionable/JJ ]
of/IN	./.

---

## Examples: Part-of-speech tags (from the BNC)

---

```
<s n="2"><w PNP>It <w VBZ>is <w AJ0-VVG>surprising  
<w AVQ>how <w DT0>many <w NN0>people  
<w VVB>take <w NN1>part <w PRP>in  
<w NN1>competition <w PRP>without <w VVG>knowing  
<w AT0>the <w NN2>rules<c PUN>, <w AV0>even  
<w PRP>at <w NN1-AJ0>&eacute;lite <w NN1>level<c  
PUN>.
```

---

## Examples: syntactic annotation (from the Penn Treebank)

---

```
(( (S (" ``")  
  (S-TPC-1  
    (NP-SBJ (PRP He) )  
    (VP (MD must)  
      (VP (VB have)  
        (VP (VBN forgiven)  
          (NP (PRP me) )))))  
    (" ")  
    (.,)  
    (NP-SBJ (NNP Henrietta) )  
    (VP (VBD murmured)  
      (S (-NONE- *T*-1) )  
      (PP (TO to)  
        (NP (DT the) (NN room) )))  
    (.,) )
```

---

## Examples: syntactic annotation (output of the Minipar parser)

---

```
> (  
E0  ((  fin C  *  )  
1  (At  ~ Prep E0  mod  (gov fin))  
2  (the  ~ Det 3  det  (gov time))  
3  (time ~ N 1  pcomp-n (gov at))  
4  (this ~ N 5  s  (gov be))  
5  (was  be VBE E0  i  (gov fin))  
E2  ((  this N 8  subj  (gov attitude) (antecedent 4))  
6  (an  ~ Det 8  det  (gov attitude))  
7  (unfashionable ~ A 8  mod  (gov attitude))  
8  (attitude ~ N 5  pred  (gov be))  
)
```

---

## Examples: syntactic annotation (from the TIGER corpus)

---

```
<s id="s2">  
<graph root="s2_502">  
<terminals>  
<t id="s2_1" word="Konzernchefs" pos="NN" morph="--" />  
<t id="s2_2" word="lehnen" pos="VVFIN" morph="--" />  
<t id="s2_3" word="den" pos="ART" morph="--" />  
<t id="s2_4" word="Milliard&#x00e4;r" pos="NN" morph="--" />  
<t id="s2_5" word="als" pos="APPR" morph="--" />  
<t id="s2_6" word="US-Pr&#x00e4;sidenten" pos="NN" morph="--"  
 />  
<t id="s2_7" word="ab" pos="PTKVZ" morph="--" />  
</terminals>
```

---

## Examples: syntactic annotation (from the TIGER corpus, ctd.)

---

```
<nonterminals>  
<nt id="s2_500" cat="NP">  
<edge label="NK" idref="s2_3" />  
<edge label="NK" idref="s2_4" />  
</nt>  
<nt id="s2_501" cat="PP">  
<edge label="AC" idref="s2_5" />  
<edge label="NK" idref="s2_6" />  
</nt>  
<nt id="s2_502" cat="S">  
<edge label="SB" idref="s2_1" />  
<edge label="HD" idref="s2_2" />  
<edge label="SVP" idref="s2_7" />  
<edge label="OA" idref="s2_500" />  
<edge label="MO" idref="s2_501" />  
</nt>  
</nonterminals>  
</graph>  
</s>
```

---

## Examples: semantic role annotation (Example from FrameNet)

---

```
<lexunit-annotation ID="123" name="push.v" frame="Manipulation"  
pos="V" incorporatedFE="">  
<definition>COD: exert force on (someone or something) in order to  
move them a way from oneself.</definition>  
<lexemes>  
<lexeme ID="31272" pos="V" breakBefore="false"  
headword="false">push</lexeme>  
</lexemes>  
<subcorpus name="V-690-trans-other">  
<annotationSet ID="420230" status="MANUAL">
```

---

## Examples: semantic role annotation (Example from FrameNet)

---

```
<layers>
  <layer ID="3111648" name="FE" rank="1">
    <labels>
      <label name="Time" ID="32042021" start="0" end="57" />
      <label name="Agent" ID="32042024" start="61" end="70" />
      <label name="Entity" ID="32042027" start="84" end="100" />
    </labels>
  </layer>
  <layer ID="3111649" name="GF">
    <labels>
      <label name="Dep" ID="32042023" start="0" end="57" />
      <label name="Ext" ID="32042026" start="61" end="70" />
      <label name="Obj" ID="32042029" start="84" end="100" />
    </labels>
  </layer>
</layers>
```

## Examples: semantic role annotation (Example from FrameNet)

---

```
<layer ID="3111650" name="PT">
  <labels>
    <label name="PPing" ID="32042022" start="0" end="57" />
    <label name="NP" ID="32042025" start="61" end="70" />
    <label name="NP" ID="32042028" start="84" end="100" />
  </labels>
</layer>
<layer ID="3111653" name="Target">
  <labels>
    <label name="Target" ID="9693266" start="77" end="82" />
  </labels>
</layer>
</layers>
```

---

## Example: semantic role annotation (Example from FrameNet)

---

```
<sentence ID="388901" aPos="7529682">
  <text>After reaching the desired revs by using the
  foot throttle , the driver then pushes the centre button
  .</text>
</sentence>
</annotationSet>
</subcorpus>
[... ]
</lexunit-annotation>
```

---

## Examples: semantic role annotation (Example from Salsa)

---

```
<s id="s11753">
  <graph root="s11753_VROOT">
    <terminals>
      <t id="s11753_1" word="``" pos="$(" morph="--"/>
      <t id="s11753_2" word="Ob" pos="KOUS" morph="--"/>
      <t id="s11753_3" word="die" pos="PDS" morph="--"/>
      <t id="s11753_4" word="da" pos="ADV" morph="--"/>
      <t id="s11753_5" word="drin" pos="ADV" morph="--"/>
      <t id="s11753_6" word="uns" pos="PPER" morph="--"/>
      <t id="s11753_7" word="&#x00fc;berhaupt" pos="ADV" morph="--"/>
      <t id="s11753_8" word="bemerken" pos="VVFIN" morph="--"/>
      <t id="s11753_9" word="?" pos="$. " morph="--"/>
      <t id="s11753_10" word="&apos;&apos;" pos="$(" morph="--"/>
    </terminals>
  </graph>
</s>
```

---

## Examples: semantic role annotation (Example from Salsa)

---

```
<nonterminals>
<nt id="s11753_500" cat="NP">
  <edge label="NK" idref="s11753_3"/>
  <edge label="MNR" idref="s11753_4"/>
  <edge label="MNR" idref="s11753_5"/>
</nt>
<nt id="s11753_501" cat="S">
  <edge label="CP" idref="s11753_2"/>
  <edge label="OA" idref="s11753_6"/>
  <edge label="MO" idref="s11753_7"/>
  <edge label="HD" idref="s11753_8"/>
  <edge label="SB"
    idref="s11753_500"/>
</nt>
```

---

```
<nt id="s11753_VROOT"
  cat="VROOT">
  <edge label="--" idref="s11753_1"/>
  <edge label="--" idref="s11753_9"/>
  <edge label="--" idref="s11753_10"/>
  <edge label="--" idref="s11753_501"/>
</nt>
</nonterminals>
</graph>
```

## Examples: semantic role annotation (Example from Salsa)

---

```
<sem>
<frames>
  <frame name="Perception_experience" id="s11753_f0">
    <target>
      <fenode idref="s11753_8"/>
    </target>
    <fe name="Phenomenon" id="s11753_f0_e1">
      <fenode idref="s11753_6"/>
    </fe>
    <fe name="Perceiver_passive" id="s11753_f0_e2">
      <fenode idref="s11753_500"/>
    </fe>
  </frame>
</frames>
</sem>
</s>
```