

Exercise 1
Working with annotated corpora

1. The text below is a paragraph from the electronic version of the Travels of Sir Mandeville.

the Soudanne, of hem that speken with him. And whan that Messangeres of straunge Contrees comen before him, the Meynee of the Soudan,° whan the Straungeres speken to hym, thei ben aboute the Souldan with Swerdes drawn and Gysarmez and Axes, here Armes lift up in highe with the Wepenes, for to smyte upon hem, zif thei seye ony Woord, that is displesance to the Soudan. And

(from J.O. Halliwell, 1883, www.archive.org)

How would you go about if your task was to find all verb forms? What would you have to do?

2. Below you see the tagged version of the same paragraph. How would you go about finding all verb forms?

And_CONJ whan_P +tat_C messangeres_NS of_P strange_ADJ contrees_NS
comen_VBP before_P him_PRO the_D meynee_N of_P the_D Soudan_N ,_
Whan_P the_D strangeres_NS speken_VBP to_P hym_PRO ,_, +tei_PRO ben_BEP
aboute_P the_D souldan_N with_P swerdes_NS drawn_VAN &_CONJ
gysarmez_NS &_CONJ axes_NS ,_, here_PRO\$ armes_NS lift_VAN vp_RP in_P
high_ADJ with_P +to_D wepenes_NS for_FOR to_TO smyte_VB vpon_P hem_PRO
+gif_P +tei_PRO seye_VBP ony_Q woord_N +tat_C is_BEP displesance_N to_P
the_D soudan_N ._. CMMANDEV,24.593_ID

3. Finally, you find a parsed version of the text below. What additional information is given in this format? What could you say about the position of all occurring verbs in this text? Mark all verb forms in red and try to analyse its position in the sentences relative to the subject and the object (if there is any).

((IP-MAT (CONJ And)
 (PP (P whan)
 (CP-ADV (C +tat)
 (IP-SUB (NP-SBJ (NS messangeres)
 (PP (P of)
 (NP (ADJ straunge) (NS contrees))))
 (VBP comen)
 (PP (P before)
 (NP (PRO him))))))
 (NP-LFD (D the) (N meynee)
 (PP (P of)
 (NP (D the) (N Soudan))))
 (, ,)
 (PP (P Whan)
 (CP-ADV (C 0)
 (IP-SUB (NP-SBJ (D the) (NS strangeres))
 (VBP speken)
 (PP (P to)
 (NP (PRO hym))))))
 (, ,)
 (NP-SBJ-RSP (PRO +tei))
 (BEP ben)
 (PP (P aboute)
 (NP (D the) (N souldan)))
 (PP (P with)
 (IP-ABS (IP-ABS (NP-SBJ (NS swerdes)
 (CONJP *ICH*-2))
 (VAN drawen)
 (CONJP-2 (CONJ &)
 (NS gysarmez) (CONJ &) (NS axes)))
 (, ,)
 (CONJP (IP-ABS (NP-SBJ (PRO\$ here) (NS armes))
 (VAN lift)
 (PP (RP vp) (P in)
 (ADJP (ADJ high)))
 (PP (P with)
 (NP (D +to) (NS wepenes)))
 (IP-INF-PRP (FOR for) (TO to)
 (VB smyte)
 (PP (P vpon)
 (NP (PRO hem)))
 (PP (P +gif)
 (CP-ADV (C 0)
 (IP-SUB (NP-SBJ (PRO +tei))
 (VBP seye)
 (NP-OB1 (Q ony) (N woord)
 (CP-REL (WNP-1 0)

T-1)

displesance)

to)

(NP (D the)

(N soudan))))))))))))))

(E_S .)) (ID CMMANDEV,24.593))

(C +tat)
(IP-SUB (NP-SBJ

(BEP is)
(NP-OB1 (N

(PP (P