A screenshot of a cell phone

Description automatically generated

**IN2110 Språkteknologiske metoder**

**Mandatory Assignment 2a**

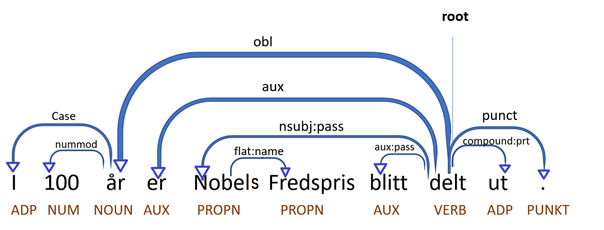
**Spring 2020**

**Lennart Nilsen**

**Majdi Omar Alali**

**Oppgave 1 – Dependensgrammatikk**

**a) Dependensgrafer**



**b) Transisjonsparsing**

Vi bruker *arc eager*-algoritmen for å parse setningen:

Merk at left-arc inholder reducing og right-arc innholder shift. Vi valgte derfor å ikke inkluderte denne i transisjonene.

**Configuration:** [root]**S** [I, 100, år, er, Nobel, Fredspris, blitt, delt, ut, .]**B**

[root, I] [ 100, år, er, Nobel, Fredspris, blitt, delt, ut, .] **shift**

[root, I, 100] [år, er, Nobel, Fredspris, blitt, delt, ut, .] **shift**

[root, I] [år, er, Nobel, Fredspris, blitt, delt, ut, .] **leftward-arc (numod)**

[root] [år, er, Nobel, Fredspris, blitt, delt, ut, .] **leftward-arc (case)**

[root, år] [ er, Nobel, Fredspris, blitt, delt, ut, .] **shift**

[root, år, er] [Nobel, Fredspris, blitt, delt, ut, .] **shift**

[root, år, er, Nobel] [Fredspris, blitt, delt, ut, .] **shift**

[root, år, er, Nobel, Fredspris] [blitt, delt, ut, .] **rightward-arc(flat:name) + shift**

[root, år, er, Nobel, Fredspris, blitt] [delt, ut, .] **shift**

[root, år, er, Nobel, Fredspris] [delt, ut, .] **leftward-arc( aux:pass)**

[root, år, er, Nobel] [delt, ut, .] **reduce**

[root, år, er] [delt, ut, .] **leftward-arc( nsubj:pass)**

[root, år] [delt, ut, .] **leftward-arc( aux)**

[root] [delt, ut, .] **leftward-arc(obl )**

[root, delt] [ ut, .] **rightward-arc(root)**

[root, delt, ut] [.] **rightward-arc(compound:prt)**

[root, delt] [.] **reduce**

[root, delt, .] [] **rightward -arc(punct)**

Forskjellen mellom arc eager og arc standard er at dependent slettes fra stacken i leftward-arc-operasjonen i arc eager. Dette gjøres ikke i arc standard sin leftward-arc-operasjon.

I arc standard slettes elementene fra stacken helt ned til hodet i rightward-arc-operasjonen, mens det i tilsvarende operasjon for arc eager legges til elementer på stacken.

I right-arc-operasjonen til arc standard kan ikke hodet brukes før alle dependentene til dette har blitt funnet. Derfor må roten legges til slutt i arc standard, men dette er ikke nødvendigvis tilfelle for arc eager.

I arc standard finnes det kun tre transisjoner: leftward arc, rightward arc, og shift, men ikke reduce. I arc eager finnes disse transisjonene og i tillegg reduce.

**Oppgave 2 – Trene modeller**

Vi trente modellen som beskrevet i oppgaven, med standard antall epochs (30). Det tok ca 20 minutter på en ifi-maskin over ssh.  
Modellen er vedlagt innleveringen, som model-best.zip

**Oppgave 3 – Evaluering**

**a) Attachment score**

All koden vi har benyttet finnes i innlevering.py. Vi har en løkke som går igjennom alle de parsede ordene, og inkrementerer en teller for hver riktig head. Hvis head er riktig sjekkes deprel. Tellerne deles så slik det er oppgitt i oppgaveteksten på totalt antall ord

|  |  |  |
| --- | --- | --- |
| **Datasett** | **UAS** | **LAS** |
| no\_bokmaal-ud-dev | 0.8943605818141824 | 0.8034039979103083 |

**b) Evaluering på andre teksttyper**

Her ser vi resultatene av parseren trent på bokmål, når den brukes til å parse nynorsk.

|  |  |  |
| --- | --- | --- |
| **Datasett** | **UAS** | **LAS** |
| no\_nynorsk-ud-dev | 0.683808 | 0.56256 |
| no\_nynorsklia-ud-dev | 0.49257109121322445 | 0.33139820919019974 |

Som vi ser oppnår modellen en signifikant lavere attachment score. Vi ser på de 10 første setningene i *no\_nynorsklia-ud-dev.conllu:*

1. vi spør først når dette her begynte for alvor og kva slags bil du hadde å køyre med .  
2. det første # det # i femogtjue .  
3. og da # kj- hadde eg Forden # eg hadde Forden da au .  
4. men den køyrde eg med ein månads seie så e .  
5. så vart den for liten så måtte eg bytte # og eg hadde masse bytta annakvart år # bilar .  
6. ja .  
7. for å få e # for å komme til noko større materiell .  
8. og dette her kj- fortsette vi med og køyrde # mjølk da leste på ein e # mellom åtti og hundre spann .  
9. om om dagen # som vi bar .  
10. utor mjølkekummen .

Det mest tydelige vi observerer her er at mange av ordene er på dialekt, noe bokmålsmodellen helt sikkert ikke er trent på. Det blir et veldig muntlig språk med en ordstilling som skiller seg fra bokmål. Enkelte av setningene er setningsfragmenter som mangler hode. Dette vil bryte reglene til arc eager. Arc eager håndterer ikke uformelt/flertydig (ambiguous) språk.