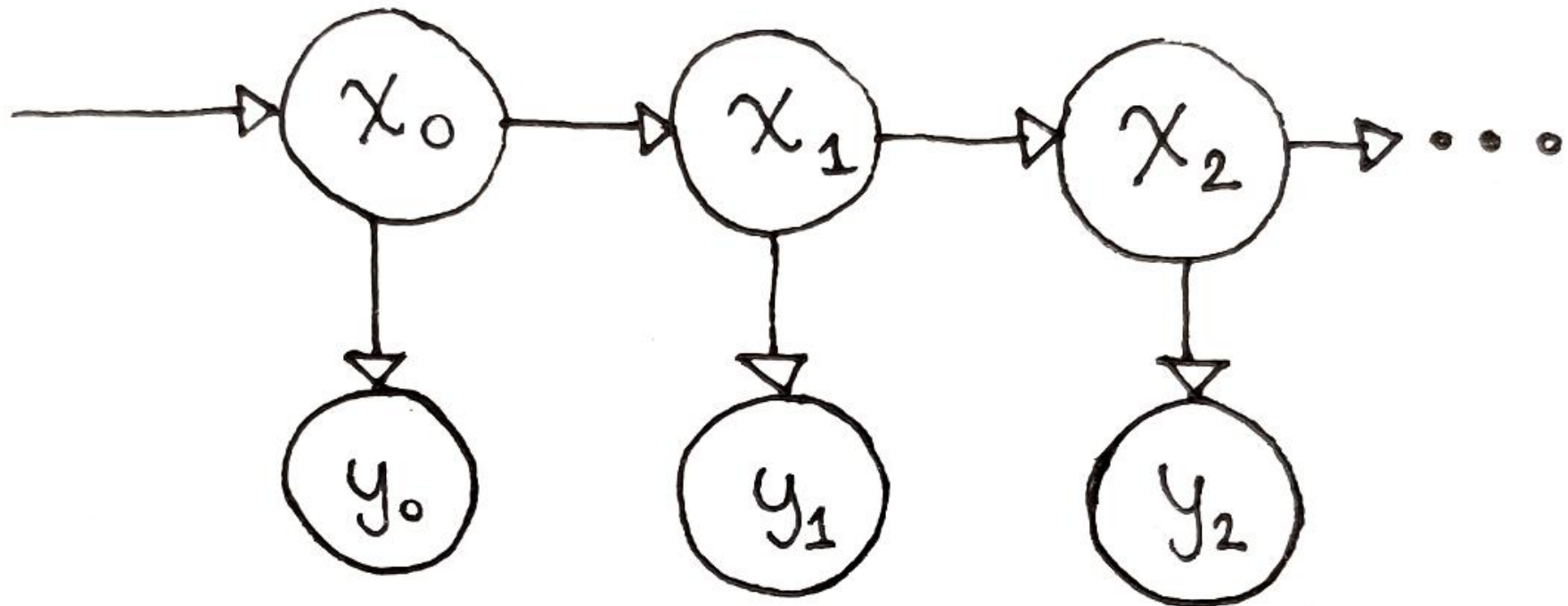
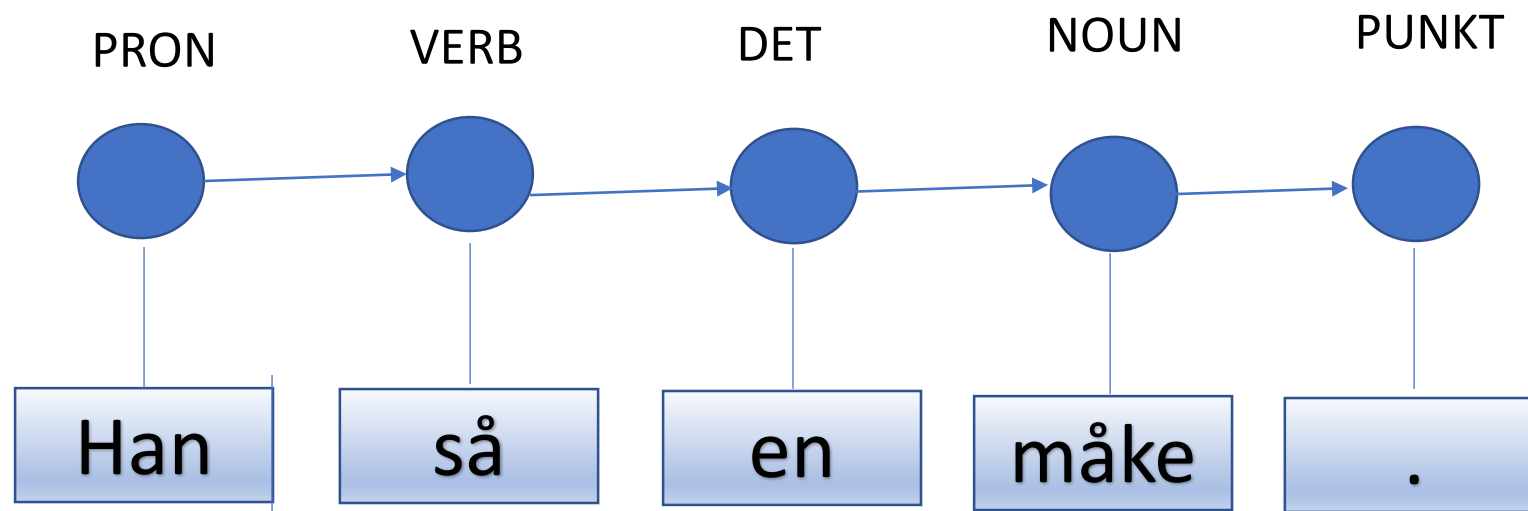




Sekvensmodeller





Ordklasser

- ADJ: adjective
- **ADP: adposition** (norsk har *preposisjoner*, andre språk *postposisjoner* eller *circumposisjoner*)
- ADV: adverb
- **AUX: auxiliary (hjelpeverb)**
- CONJ: coordinating conjunction
- DET: determiner
- INTJ: interjection
- NOUN: noun
- **NUM: numeral (tallord)**
- **PART: particle** (partikkel: ubøyelige småord, f eks *å*)
- PRON: pronoun
- PROPN: proper noun
- **PUNCT: punctuation**
- SCONJ: subordinating conjunction
- **SYM: symbol**
- VERB: verb
- **X: other**

BIO-system

BIO-skjema (kalles også IOB): Beginning - Inside - Outside -

- B = Begynnelsen på et navn
- I = Inne i et navn
- O = Utenfor et navn

B-PER	I-PER	O	O
Ole	Hansen	er	bra

Navnegjenkjenning

BIO-skjema (kalles også IOB): Beginning - Inside - Outside -

- B = Begynnelsen på et navn
- I = Inne i et navn
- O = Utenfor et navn

O	O	B-PER	B-PER
Faren	lover	Ole	Kari

Navnegjenkjenning

Alternativ til BIO-skjema: **BILOU** (kalles også BMEWO)

Beginning - **I**nside - **L**ast - **O**utside - **U**nit

- B = Begynnelsen på et navn
- I = Inne i et navn
- L = Siste ord i et navn
- O = Utenfor et navn
- U = Enhet, dvs ett-ords-navn

Navnegjenkjenning

Alternativ til BIO-skjema: **BILOU** (kalles også BMEWO)

Beginning - **I**nside - **L**ast - **O**utside - **U**nit

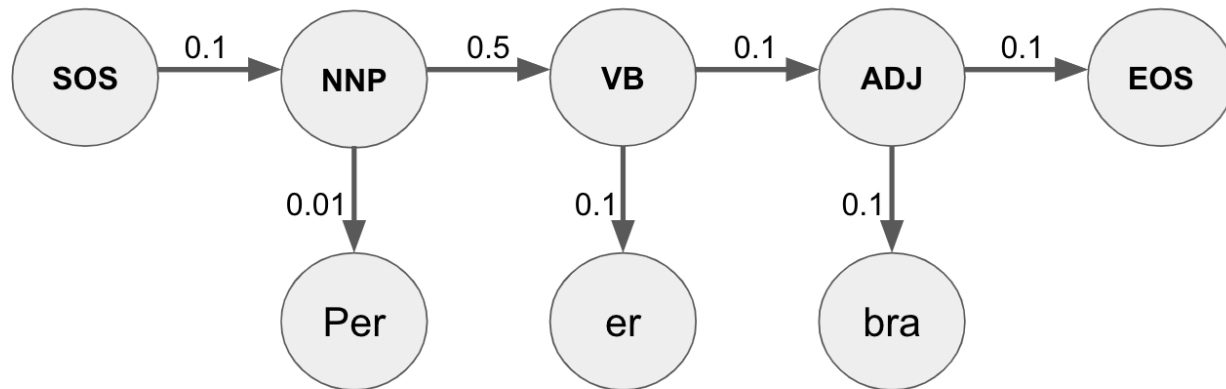
B-ORG	I-ORG	I-ORG	O	O	O	B-LOC	O
Universitet	i	Oslo	er	best	i	Norge	!
B-ORG	I-ORG	L-ORG	O	O	O	U-LOC	O
Universitet	i	Oslo	er	best	i	Norge	!

Navnegjenkjenning

Vanlige kategorier for navn er:

- PER: Person
- LOC: Sted
- ORG: Organisasjon
- MISC: Alt annet...

HMM



En HMM-modell defineres av 5 komponenter:

1. Et sett av mulige tilstander $\mathbf{Q} = \{s_1, s_2, \dots, s_N\}$
2. Et sett av mulige observasjoner $\mathbf{O} = \{o_1, o_2, \dots, o_O\}$
3. En transisjonsmodell \mathbf{A} med sansynnlighetene $P(s_{t+1}|s_t)$ av å bevege oss fra s_t til s_{t+1} for alle tilstandene
4. En sansynnlighetsfordeling $\pi = P(s_0)$ over starttilstander
5. En emisjonsmodell \mathbf{B} med sansynnlighetene $P(o_t|s_t)$ av å observere o_t hvis vi er i tilstanden s_t