

Treningsdata	
A	<i>eple smør sukker kanel</i>
A	<i>smør sukker kake</i>
B	<i>fotball leke kake</i>
Testdata	
?	<i>kake mel smør kake</i>

Korpusdata

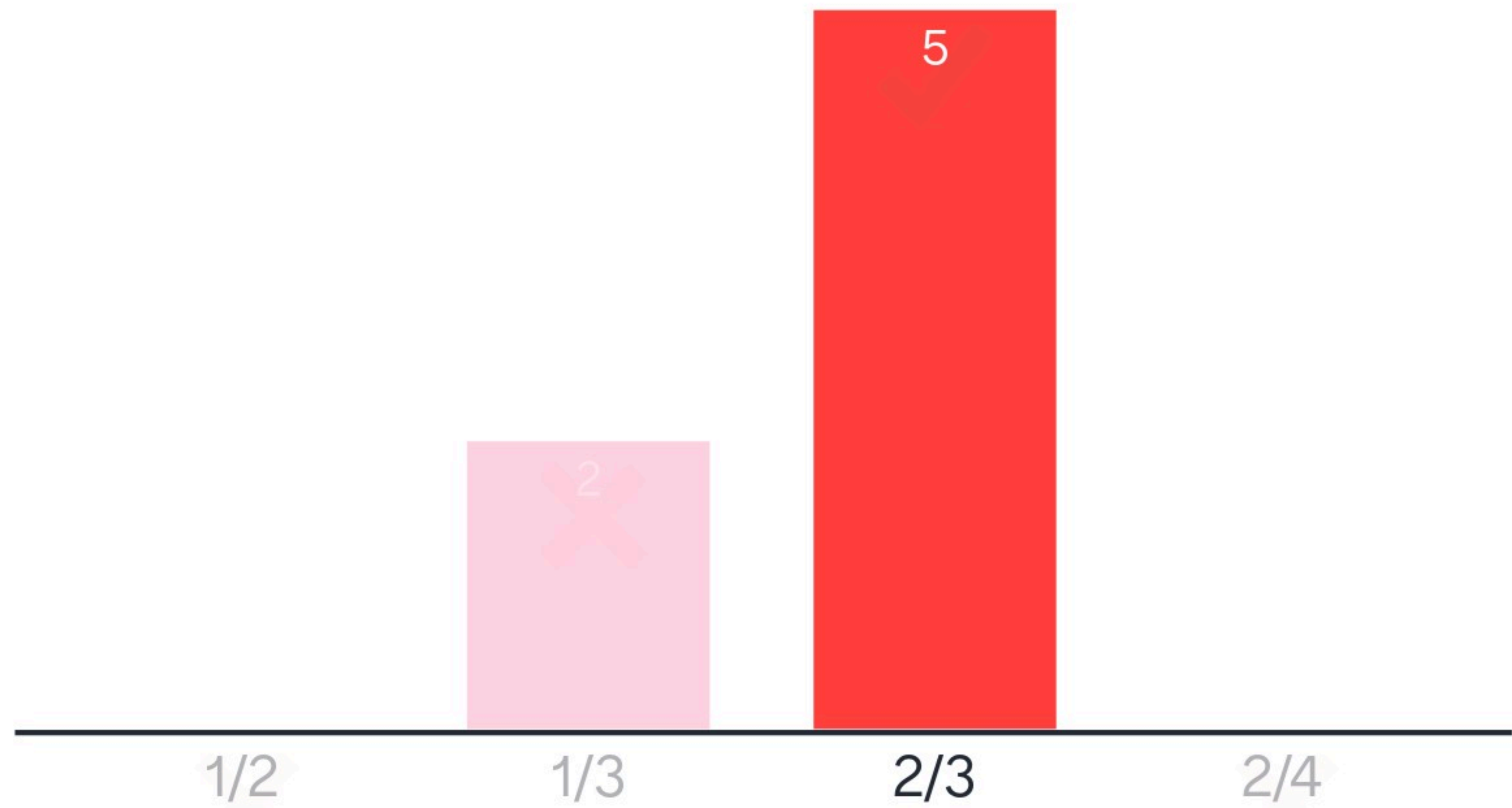


Hvilket ord ekskluderes sannsynligheten for?



Ordet "mel" forekommer ikke i noen av klassene i treningsdokumentene.

Hva er prior-sannsynligheten til klassen A?



2 av 3 treningsdokumenter har klassen A.

Hvor mange token forekommer i klassen A?

7



4x

4



2x

5



Det korrekte svaret er: 7




Hvor stort er ordforrådet i treningsdataen?

7 
3x

6 

9 

Eple, smør, sukker, kanel, kake 

Ordforrådet består av alle ordtypene. Størrelsen (eller lengden) av dette er 7.

Det korrekte svaret er: 7

Hva er sannsynligheten for at testdokumentet kategoriseres som klassen B hvis vi ikke bruker glatting?

0



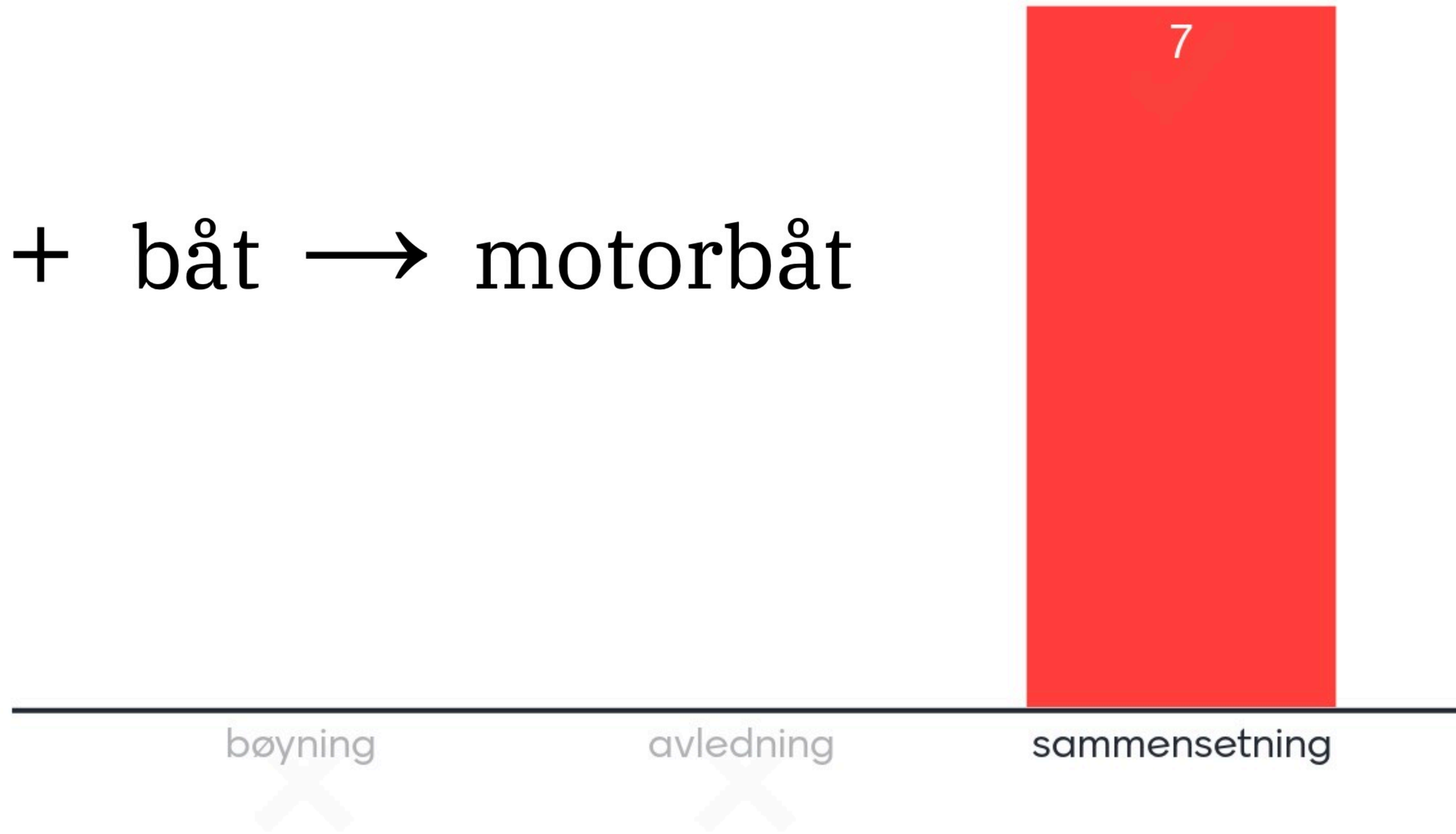
3x

$P(\text{smør} | B)$ blir 0 fordi "smør" ikke er i noe dokument med klassen B, da blir $P(\text{setning} | B)$ også 0 fordi alle sannsynlighetene multipliseres sammen.

Det korrekte svaret er: 0

Hva slags operasjon utføres her?

motor + båt → motorbåt

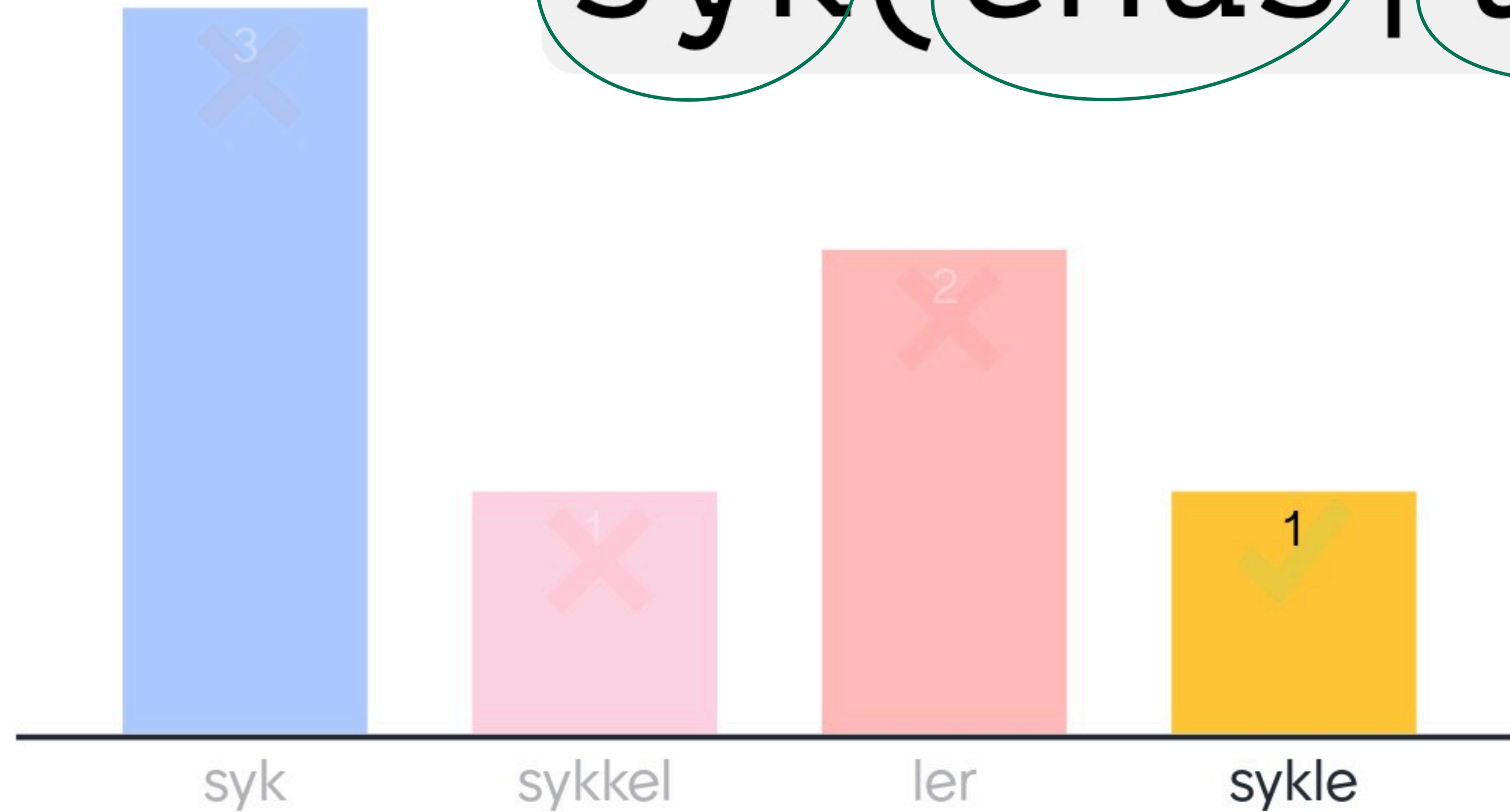


Hvilket ord gjenkjennes av det regulære uttrykket?

Må finne dette, og én av disse

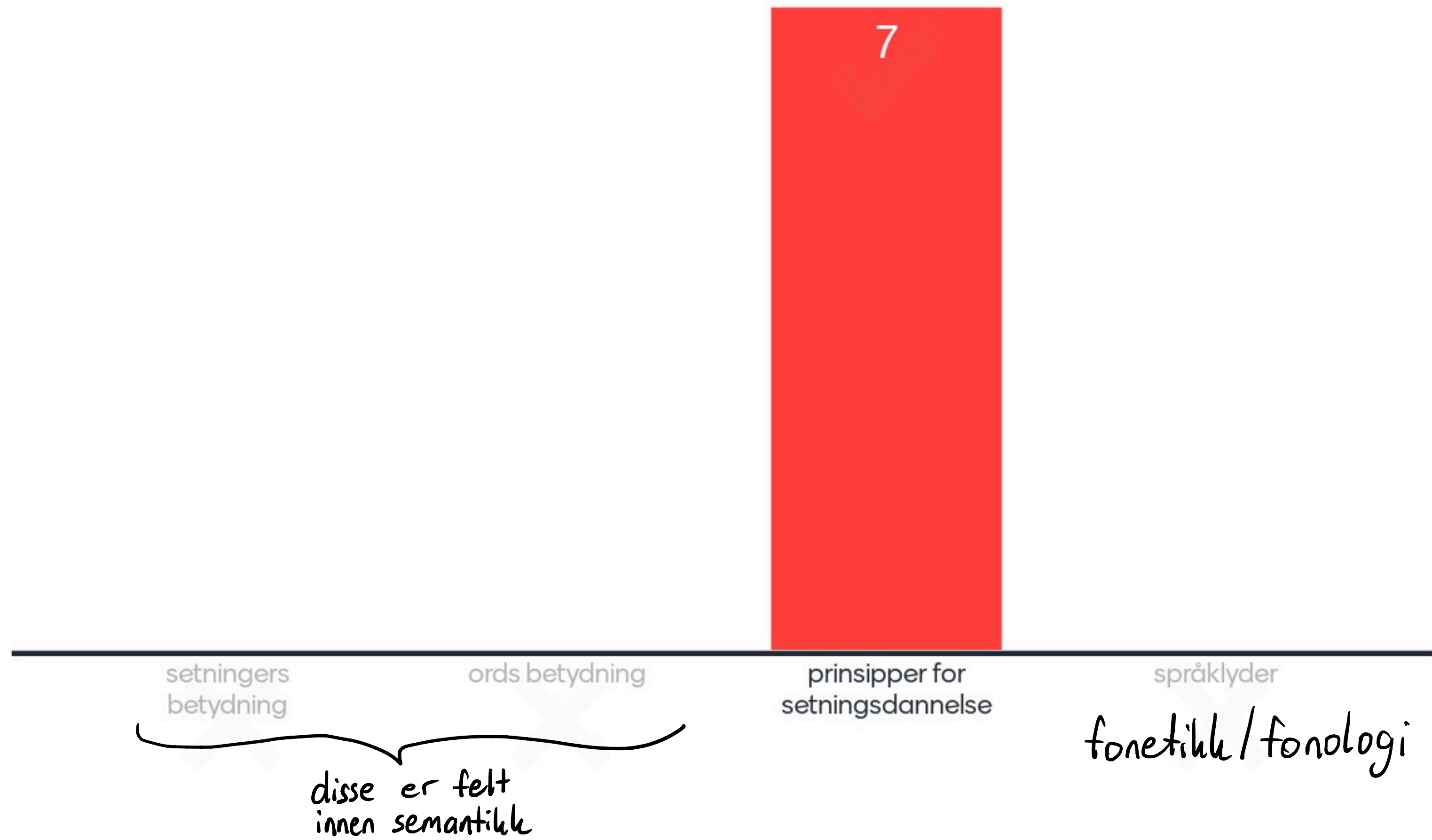
syk(ehus | ler?)

valgfri "r"

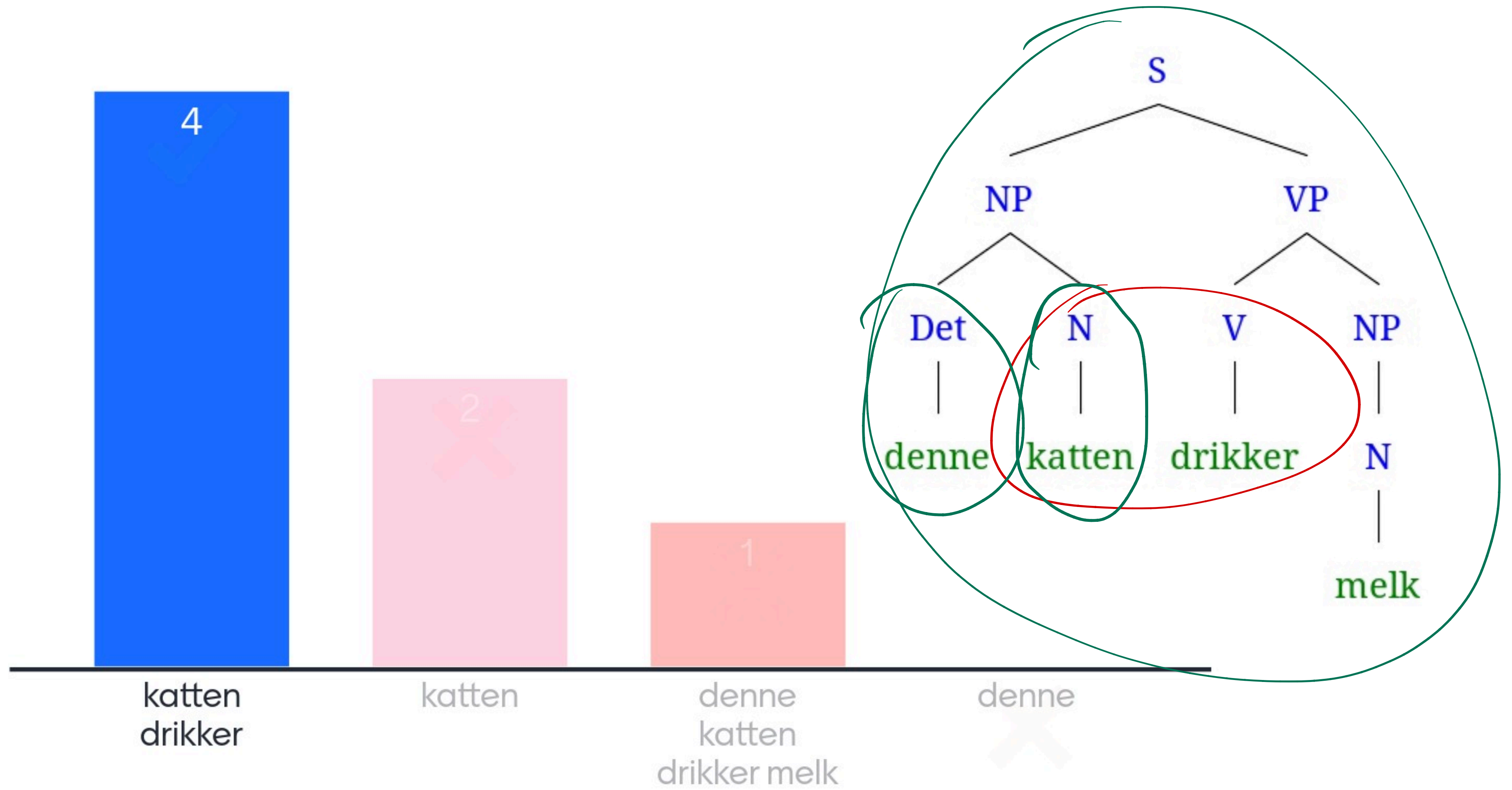


Andre ord som gjenkjennes: sykehus, sykler

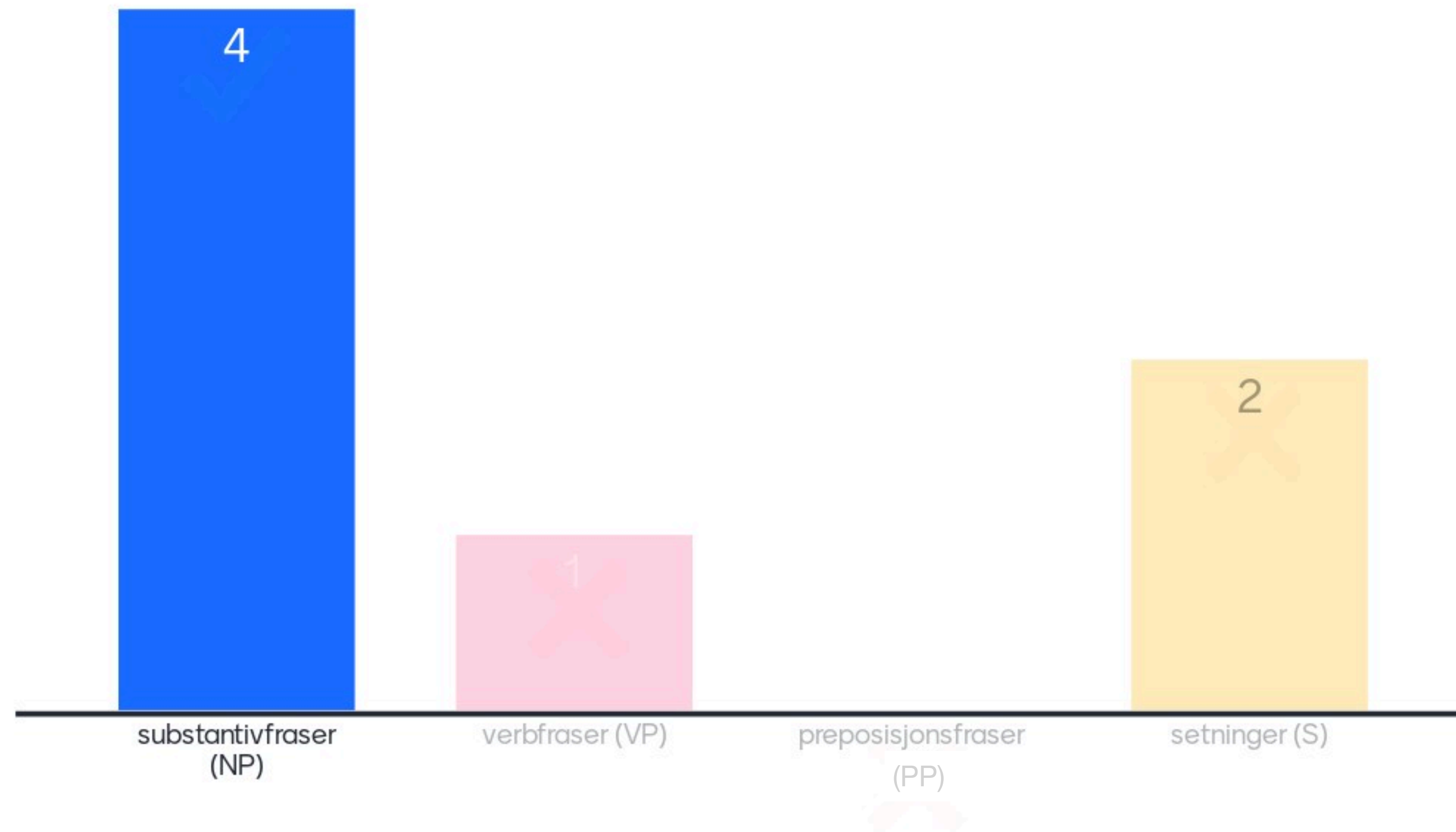
Hva handler syntaks om?



Hvilket alternativ er *ikke* en konstituent?

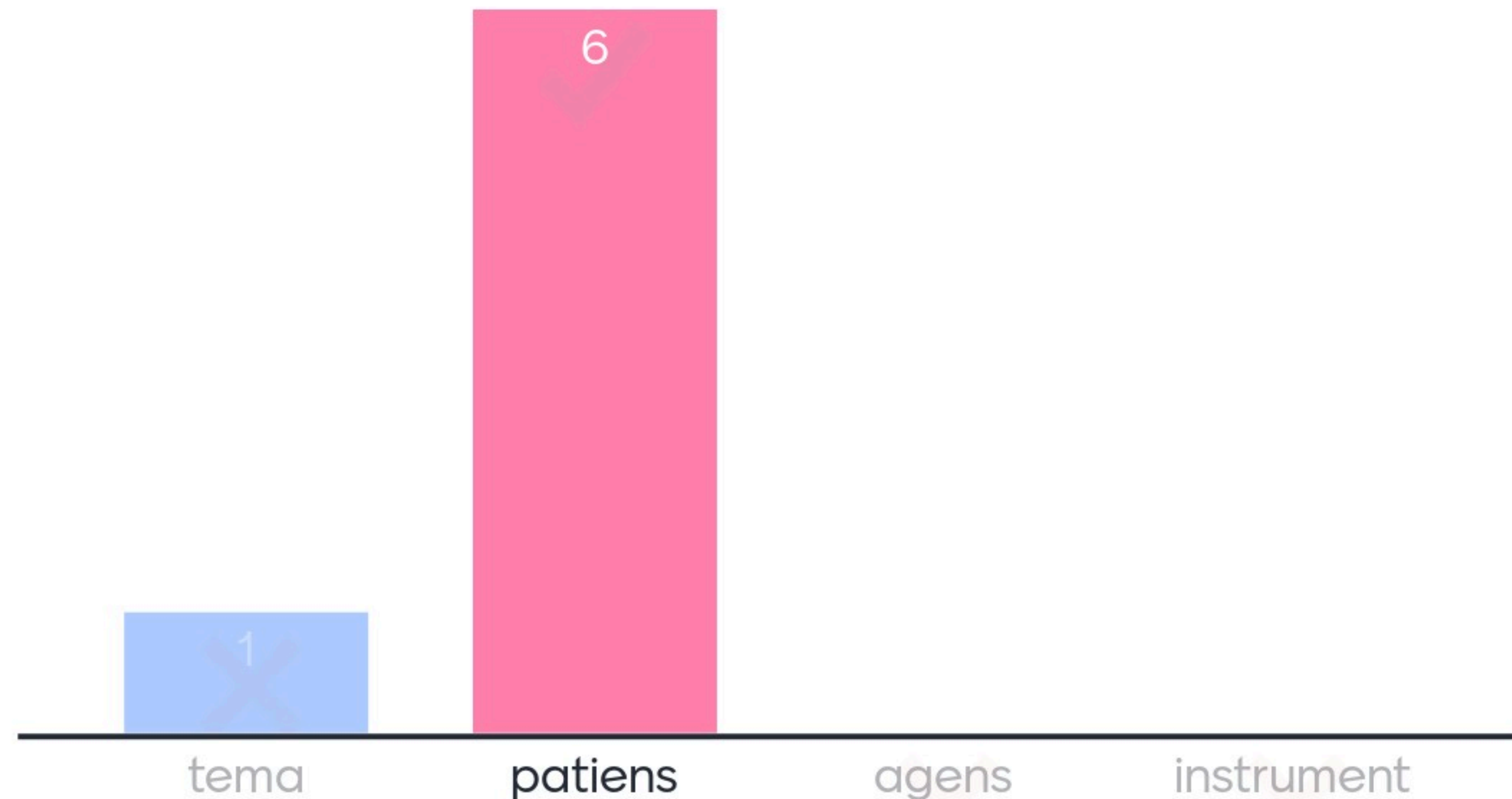


Hva slags fraser kan tildeles semantiske roller?



Stikkord: entiteter

Hvilken semantisk rolle har *et brev* i *Jørgen skrev et brev med en dårlig penn*?



gjennomgår forandring
eller endrer tilstand

Når et ord avledes, endres alltid ordets ...



Hvor mange bundne morfemer består ordet *banankakesmulen* av?

1 ✓
4x

3 ✗

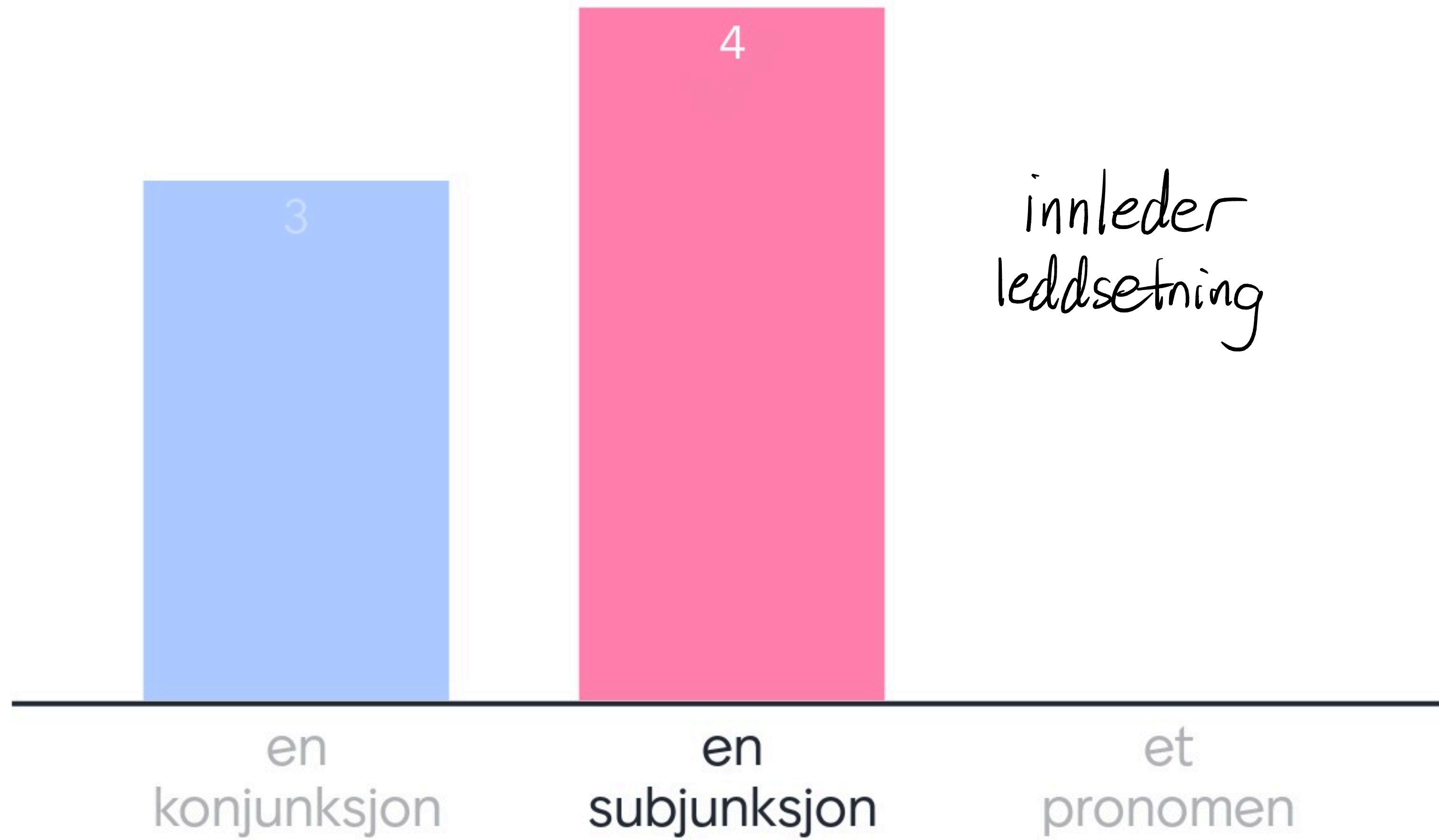
4 ✗

5 ✗

banan-kake-smule(n)

Det korrekte svaret er: 1

Ordet *som* er ...



Navnet *Lotte* er ...

