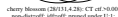


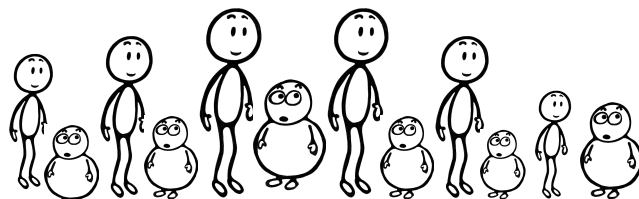
Hairball Effect!!!



Cooccurrence Weight

⇒ Bell curve!  ring!ring!

It allows us to divide words into 3 kinds: \Rightarrow high, middle, and low.



Cooccurrence weight word patterns
like two person's weights

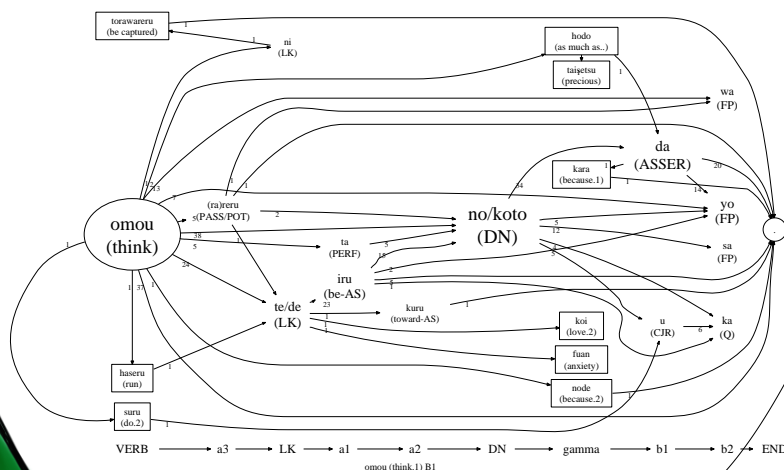
See what happens!

Over 1σ \Rightarrow **Content Tail**

Under $-1\sigma \Rightarrow$ **Function Tail**

Tall Head

Word distribution by frequency-rank



C. F. Gauss
(1777–1855)

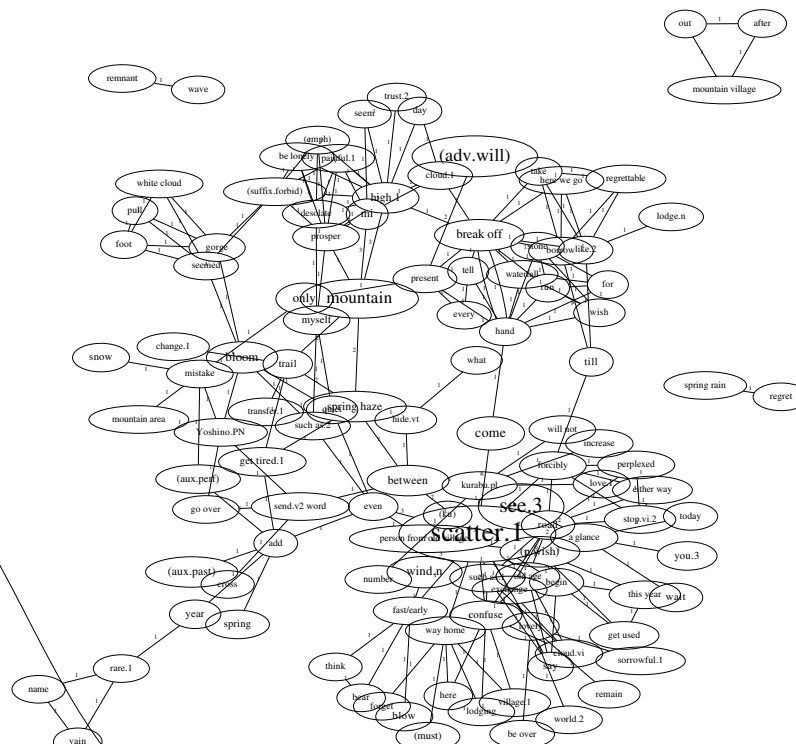
Function Tail

 $-\sigma$

0

C

Content Tail



cherry blossom (28/131,4.28): CT ctf>0.0

Long Tail Dinosaur

← more generic words

more specific words \rightarrow