

Mama

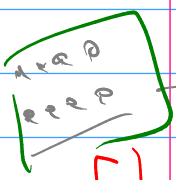
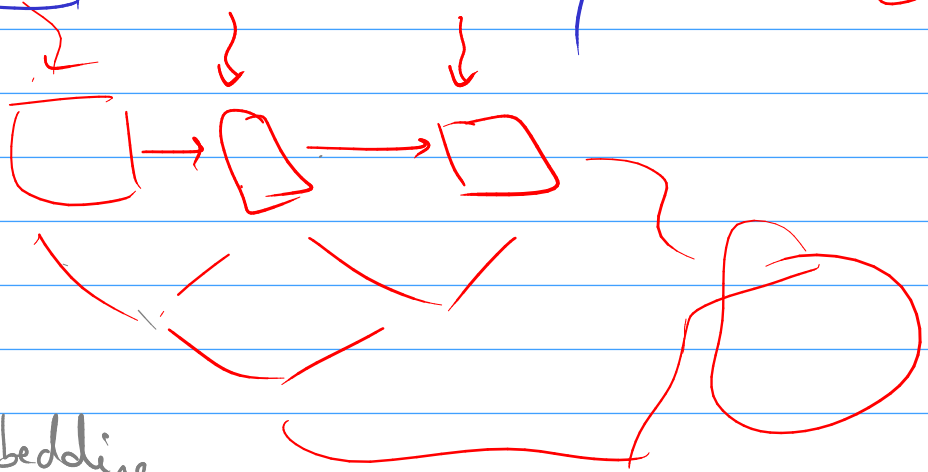
Mama Pamy

[... mama ...]  $\vec{0}_v$

↓ diff

[0 0 1 0 0]

x



embedding distributed representation

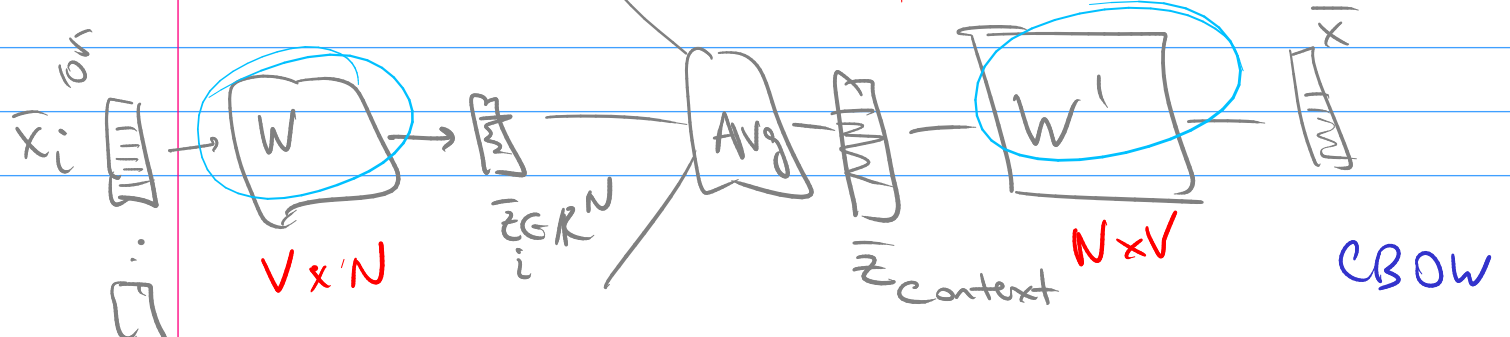
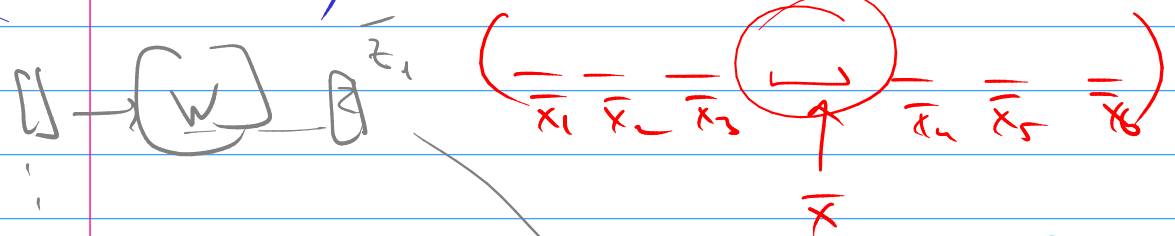
(Bengio, 2003; 2006)

$\vec{x} \in \mathbb{R}^{200}$

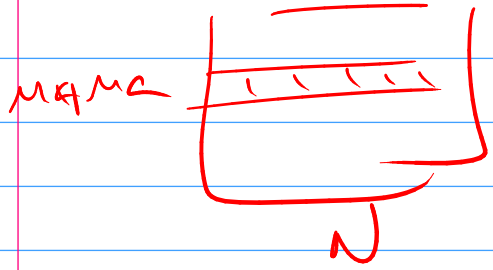
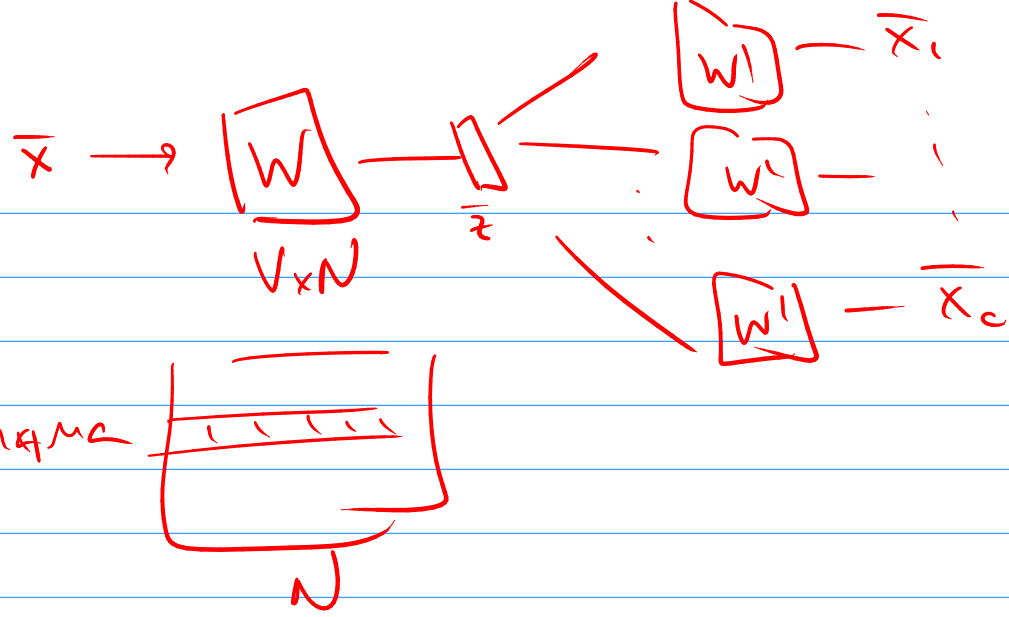
$\{e, t, f, o, r\}$

(Mikolov 2013 et al.)

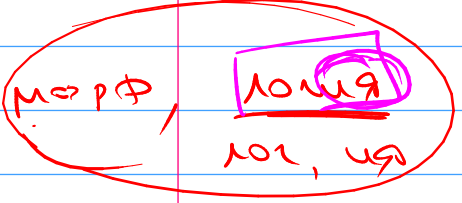
word2vec



skip-gram



morphology → tokens



bpe  
byte-pair encoding

(0) Vocab

(1) word → -, w, o, r, d, #

(2) pairs of adjacent characters

(3) merge the pairs

{#uo, мор, ..., us#}

u, o, p, f, ..., s → tokens

u, o, r, d

ur, or

#, ur, or, #

#uror#