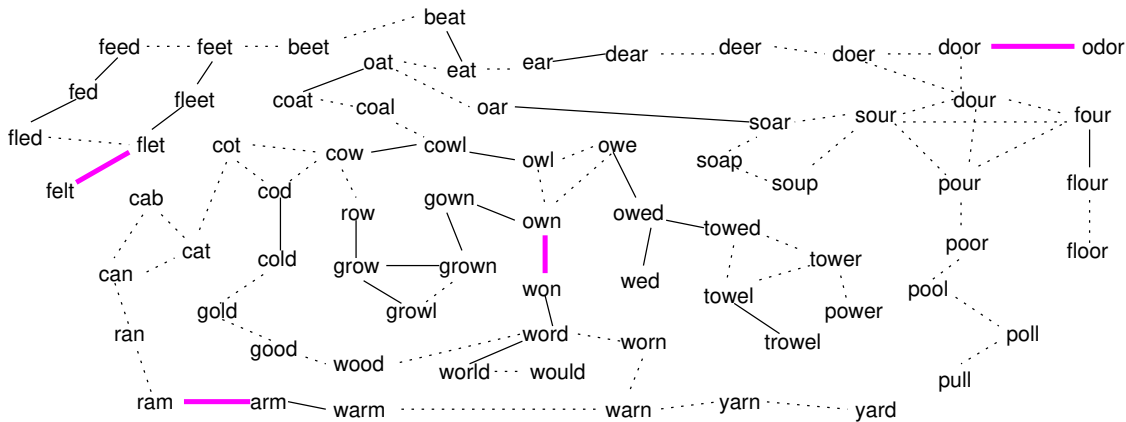


Edit distance and information theory units:

- ▶ The edit distance problem and algorithm (**today**)
- ▶ A quick tour of information theory (next week Wednesday)
- ▶ Lab: Autoregressive text generation (next week Friday)

Today:

- ▶ Perspective: Chatbots and the illusion of thinking
- ▶ The idea of edit distance
- ▶ The minimum edit distance problem
- ▶ The minimum edit distance algorithm

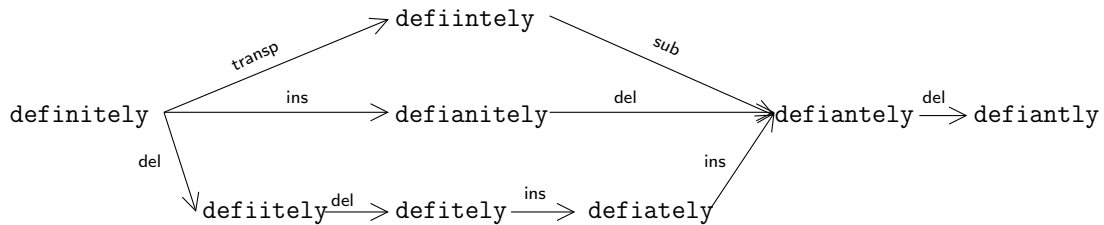


Versions of the minimum edit distance:

- ▶ Substitutions only: *Hamming distance*. (Richard Hamming, 1950)
- ▶ Insertions and deletions: *Longest common subsequence*.
- ▶ Substitutions, insertions, and deletions: *Levenshtein distance*. (Vladimir Levenshtein, 1966)
- ▶ Substitutions, insertions, deletions, and transpositions: *Damerau-Levenshtein distance*. (Fred Damerau, 1964)

recieve $\xrightarrow{\text{del}}$ receve $\xrightarrow{\text{ins}}$ receive versus recieve $\xrightarrow{\text{transp}}$ receive

seperate $\xrightarrow{\text{del}}$ seprate $\xrightarrow{\text{ins}}$ separate versus seperate $\xrightarrow{\text{sub}}$ separate



$$D[i][j] = \begin{cases} 0 & \text{if } i = j = 0 & \text{(Empty prefixes: do nothing)} \\ j \cdot C[0] & \text{if } i = 0 \text{ and } j > 0 & \text{(Empty prefix of } a: \text{ insert all the } b \text{ prefix)} \\ i \cdot C[1] & \text{if } i > 0 \text{ and } j = 0 & \text{(Empty prefix of } b: \text{ delete all the } a \text{ prefix)} \\ \min \left(\begin{array}{ll} C[0] + D[i-1][j] & \text{(insertion)} \\ C[1] + D[i][j-1] & \text{(deletion)} \\ C[2] + D[i-1][j-1] & \text{(substitution)} \\ C[3] + D[i-2][j-2] & \text{if } a[i-1] = b[j-2] \text{ and } a[i-2] = b[j-1] \text{ (transposition)} \\ C[4] + D[i-1][j-1] & \text{if } a[i-1] = b[j-1] \text{ (nop)} \end{array} \right) & \text{Otherwise} \end{cases}$$

n	6/ins-all	5/ins	4/ins	4/ins	3/ins	3/ins	2/nop	3/del
e	5/ins-all	4/ins	3/ins	3/ins	2/ins	2/sub	3/del	4/del
v	4/ins-all	3/ins	2/ins	2/ins	1/nop	2/del	3/del	4/del
a	3/ins-all	2/ins	1/nop	1/transp	2/del	3/del	4/del	5/del
r	2/ins-all	1/ins	1/sub	1/nop	2/del	3/del	4/del	5/del
c	1/ins-all	0/nop	1/del	2/del	3/del	4/del	5/del	6/del
	0/del-all	1/del-all	2/del-all	3/del-all	4/del-all	5/del-all	6/del-all	7/del-all
		c	a	r	v	i	n	g

c	a	r	v	i	n	g
nop	transp	nop	sub	nop	del	
c	r	a	v	e	n	

Coming up:

- ▶ Reading from J&M, Section 2.9 (Mon, Sept 8)
- ▶ Regular expressions assignment (Tues, Sept 9)
- ▶ Edit distance quiz (Tues, Sept 9)
- ▶ Reading: Stone (technical) or Doyle (fiction) (see Canvas) (Wed, Sept 10)
- ▶ Edit distance assignment (Mon, Sept 15)

Next time: A quick tour of information theory