Trigram Taggers

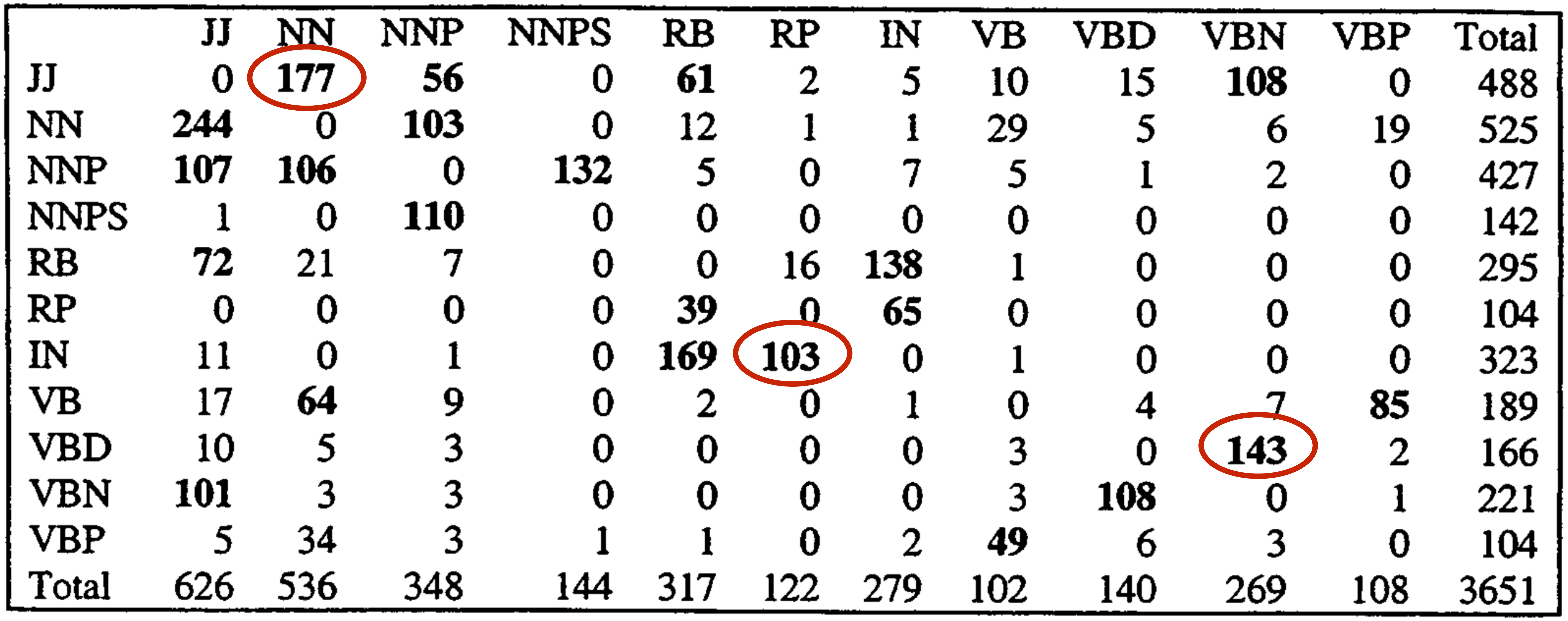
NNP VBZ NN NNS CD NN
Fed raises interest rates 0.5 percent

- Normal HMM “bigram” model: y_1 = NNP, y_2 = VBZ, …
- Trigram model: y_1 = (<S>, NNP), y_2 = (NNP, VBZ), …
- Probabilities now look like P((NNP, VBZ) | (<S>, NNP)) — more context! We know the verb is occurring two words after <S>
- Tradeoff between model capacity and data size — trigrams are a “sweet spot” for POS tagging

HMM POS Tagging

- Penn Treebank English POS tagging: 44 tags
- Baseline: assign each word its most frequent tag: ~90% accuracy
- Trigram HMM: ~95% accuracy / 55% on words not seen in train
- TnT tagger (Brants 1998, tuned HMM): 96.2% acc / 86.0% on unks
- MaxEnt tagger (Toutanova + Manning 2000): 96.9% / 87.0%
- State-of-the-art (BiLSTM-CRFs, BERT): 97.5% / 89%+

POS Errors



	JJ	NN	NNP	NNPS	RB	RP	IN	VB	VBD	VBN	VBP	Total
JJ	0	**177**	**56**	0	**61**	2	5	10	15	**108**	0	488
NN	**244**	0	**103**	0	12	1	1	29	5	6	19	525
NNP	**107**	**106**	0	**132**	5	0	7	5	1	2	0	427
NNPS	1	0	**110**	0	0	0	0	0	0	0	0	142
RB	**72**	21	7	0	0	16	**138**	1	0	0	0	295
RP	0	0	0	0	**39**	0	**65**	0	0	0	0	104
IN	11	0	1	0	**169**	**103**	0	1	0	0	0	323
VB	17	**64**	9	0	2	0	1	0	4	7	**85**	189
VBD	10	5	3	0	0	0	0	3	0	**143**	2	166
VBN	**101**	3	3	0	0	0	0	3	**108**	0	1	221
VBP	5	34	3	1	1	0	2	**49**	6	3	0	104
Total	626	536	348	144	317	122	279	102	140	269	108	3651

JJ/NN NN
official knowledge

VBD RP/IN DT NN
made up the story

RB VBD/VBN NNS
recently sold shares

(NN NN: tax cut, art gallery, …)

Toutanova + Manning (2000)

Remaining Errors

- Lexicon gap (word not seen with that tag in training): 4.5% of errors
- Unknown word: 4.5%
- Could get right: 16% (many of these involve parsing!)
- Difficult linguistics: 20%

VBD / VBP? (past or present?)

They set up absurd situations, detached from reality

- Underspecified / unclear, gold standard inconsistent / wrong: **58%**

adjective or verbal participle? JJ / VBN?

a $ 10 million fourth-quarter charge against discontinued operations

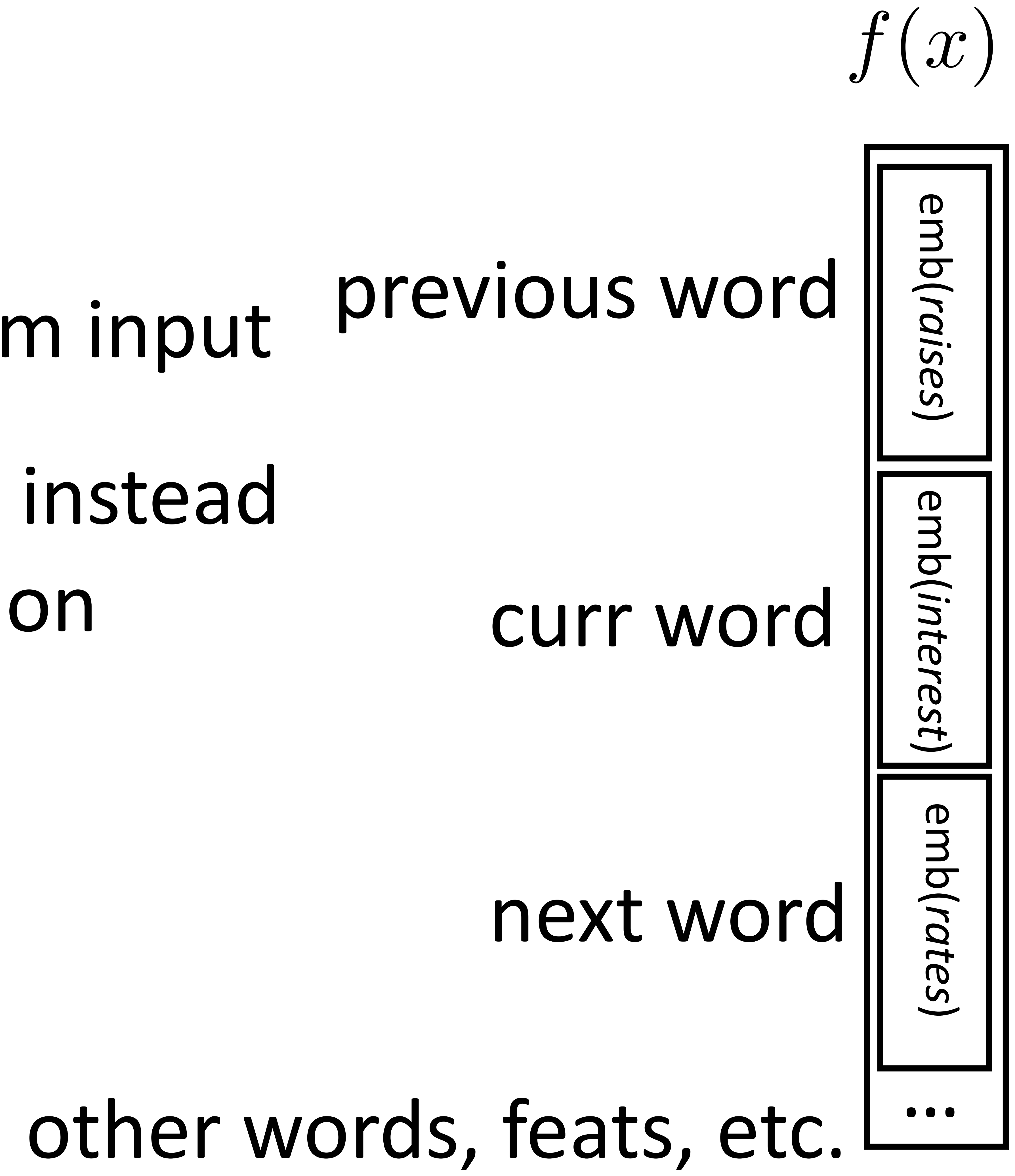
Manning 2011 “Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?”

POS with Feedforward Networks

??

Fed raises ***interest*** *rates in order to …*

- Word embeddings for each word form input
- *f*(*x*) doesn't look like a bag-of-words, instead captures position-sensitive information



$f(x)$

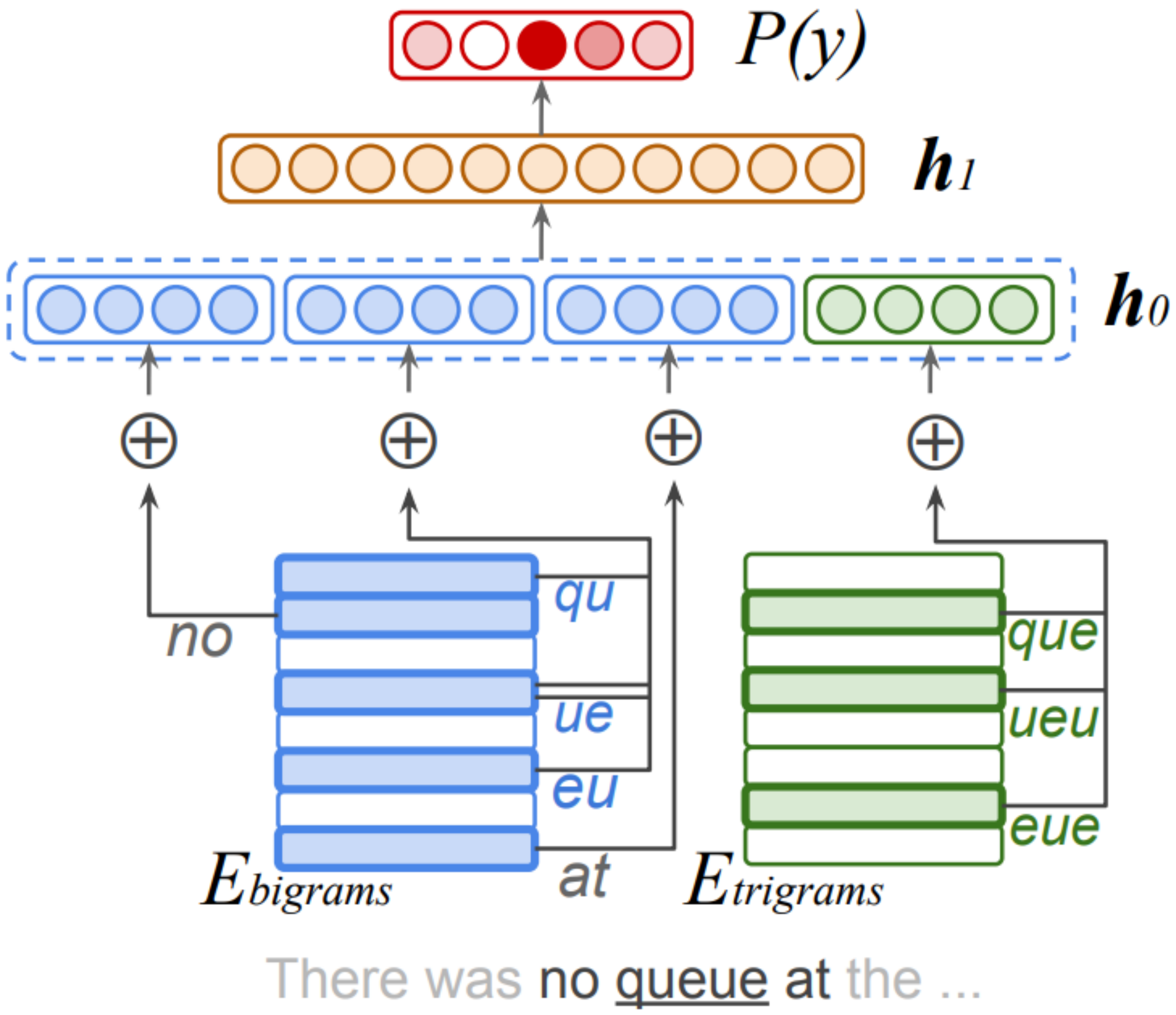
previous word

curr word

next word

other words, feats, etc.

POS with Feedforward Networks

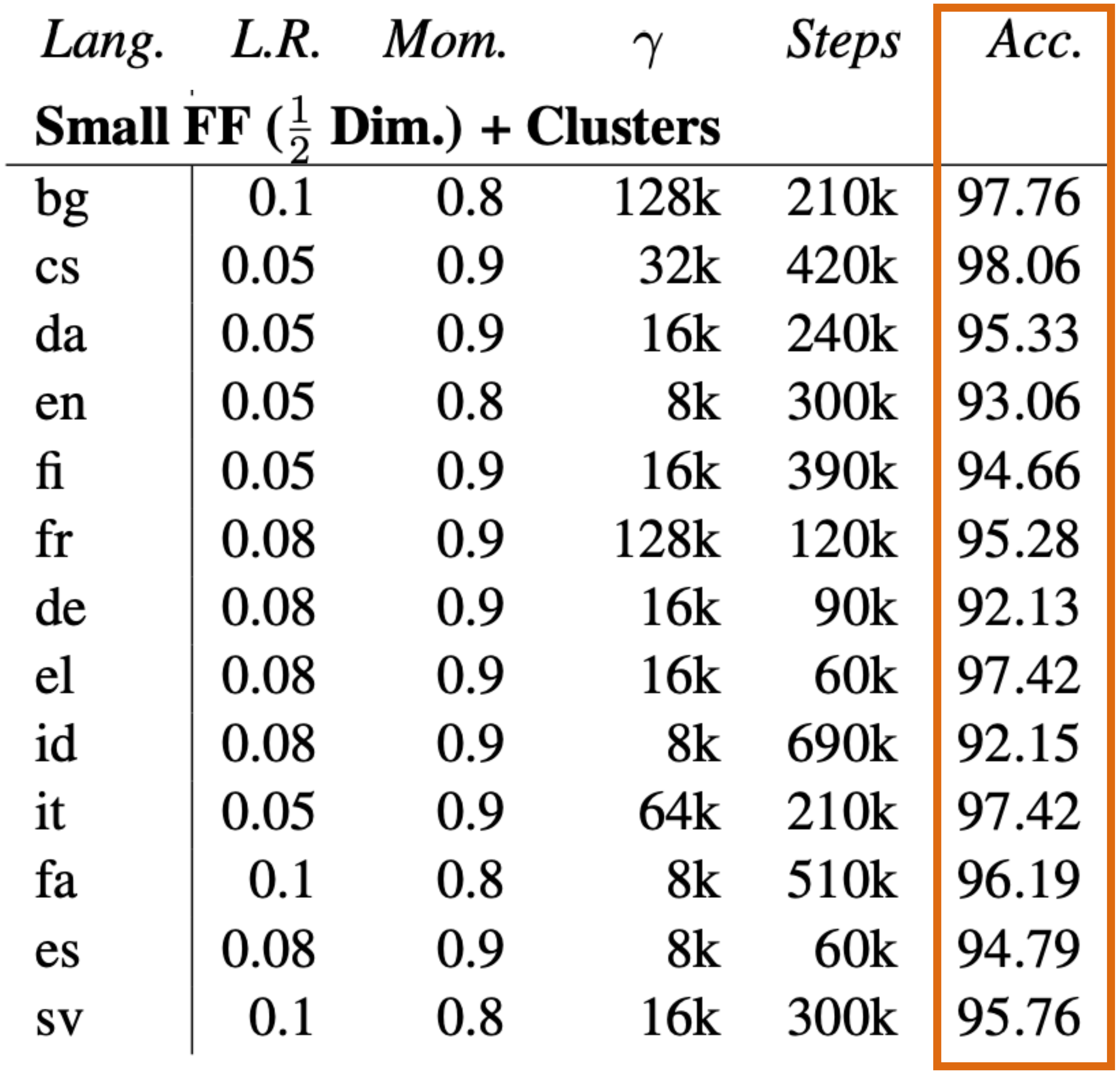


- Botha et al. (2017): small FFNNs for NLP tasks
- Use bag-of-character bigram + trigram embeddings for each word
- Hidden layer mixes these different signals and learns feature conjunctions

Botha et al. (2017)

POS with Feedforward Networks

- Works well on a range of languages
- Better than a RNN-based approach (Gillick et al., 2016)



Lang.	*L.R.*	*Mom.*	γ	*Steps*	*Acc.*
Small FF ($\frac{1}{2}$ Dim.) + Clusters					
bg	0.1	0.8	128k	210k	97.76
cs	0.05	0.9	32k	420k	98.06
da	0.05	0.9	16k	240k	95.33
en	0.05	0.8	8k	300k	93.06
fi	0.05	0.9	16k	390k	94.66
fr	0.08	0.9	128k	120k	95.28
de	0.08	0.9	16k	90k	92.13
el	0.08	0.9	16k	60k	97.42
id	0.08	0.9	8k	690k	92.15
it	0.05	0.9	64k	210k	97.42
fa	0.1	0.8	8k	510k	96.19
es	0.08	0.9	8k	60k	94.79
sv	0.1	0.8	16k	300k	95.76

Botha et al. (2017)