POS Tagging

HMM POS Tagging

- Baseline: assign each word its most frequent tag: ~90% accuracy

- Trigram HMM: ~95% accuracy / 55% on unknown words

Trigram Taggers

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

▶ Trigram model: y_1 = (<S>, NNP), y_2 = (NNP, VBZ), …

▶ P((VBZ, NN) | (NNP, VBZ)) — more context! Noun-verb-noun S-V-O

▶ Tradeoff between model capacity and data size — trigrams are a "sweet spot" for POS tagging

HMM POS Tagging

▸ Baseline: assign each word its most frequent tag: ~90% accuracy

▸ Trigram HMM: ~95% accuracy / 55% on unknown words

▸ 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): 97.5% / 89%+

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, …)

Slide credit: Dan Klein / 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%**
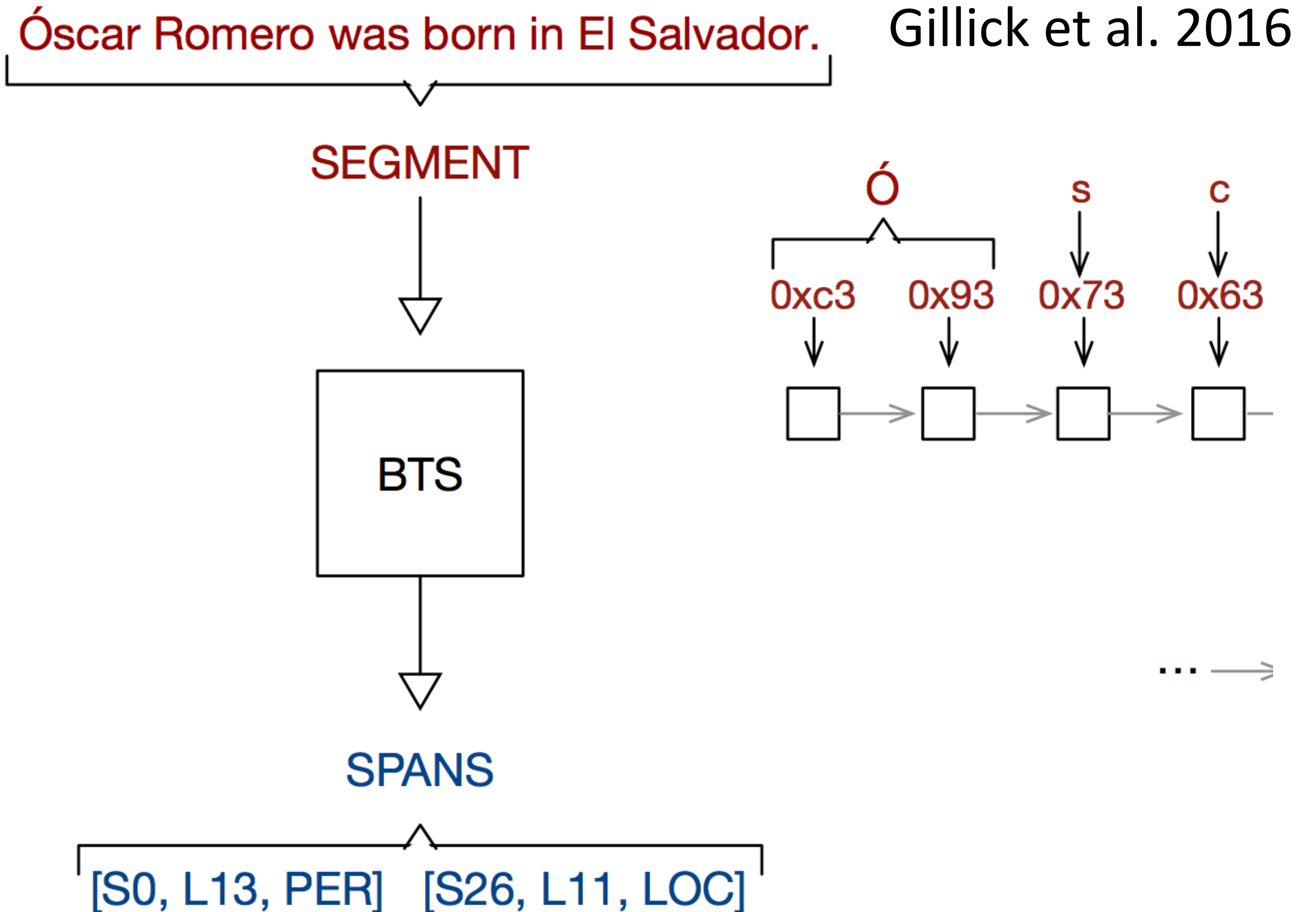
adjective or verbal participle? JJ / VBN?

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?"

Other Languages

Language	CRF+	CRF	BTS	BTS*
Bulgarian	97.97	97.00	97.84	97.02
Czech	98.38	98.00	98.50	98.44
Danish	95.93	95.06	95.52	92.45
German	93.08	91.99	92.87	92.34
Greek	97.72	97.21	97.39	96.64
English	95.11	94.51	93.87	94.00
Spanish	96.08	95.03	95.80	95.26
Farsi	96.59	96.25	96.82	96.76
Finnish	94.34	92.82	95.48	96.05
French	96.00	95.93	95.75	95.17
Indonesian	92.84	92.71	92.85	91.03
Italian	97.70	97.61	97.56	97.40
Swedish	96.81	96.15	95.57	93.17
AVERAGE	96.04	95.41	95.85	95.06

Gillick et al. 2016

Óscar Romero was born in El Salvador.

SEGMENT

Ó   s   c

0xc3   0x93   0x73   0x63

BTS

...

SPANS

[S0, L13, PER]   [S26, L11, LOC]

▸ Universal POS tagset (~12 tags), cross-lingual neural model works well