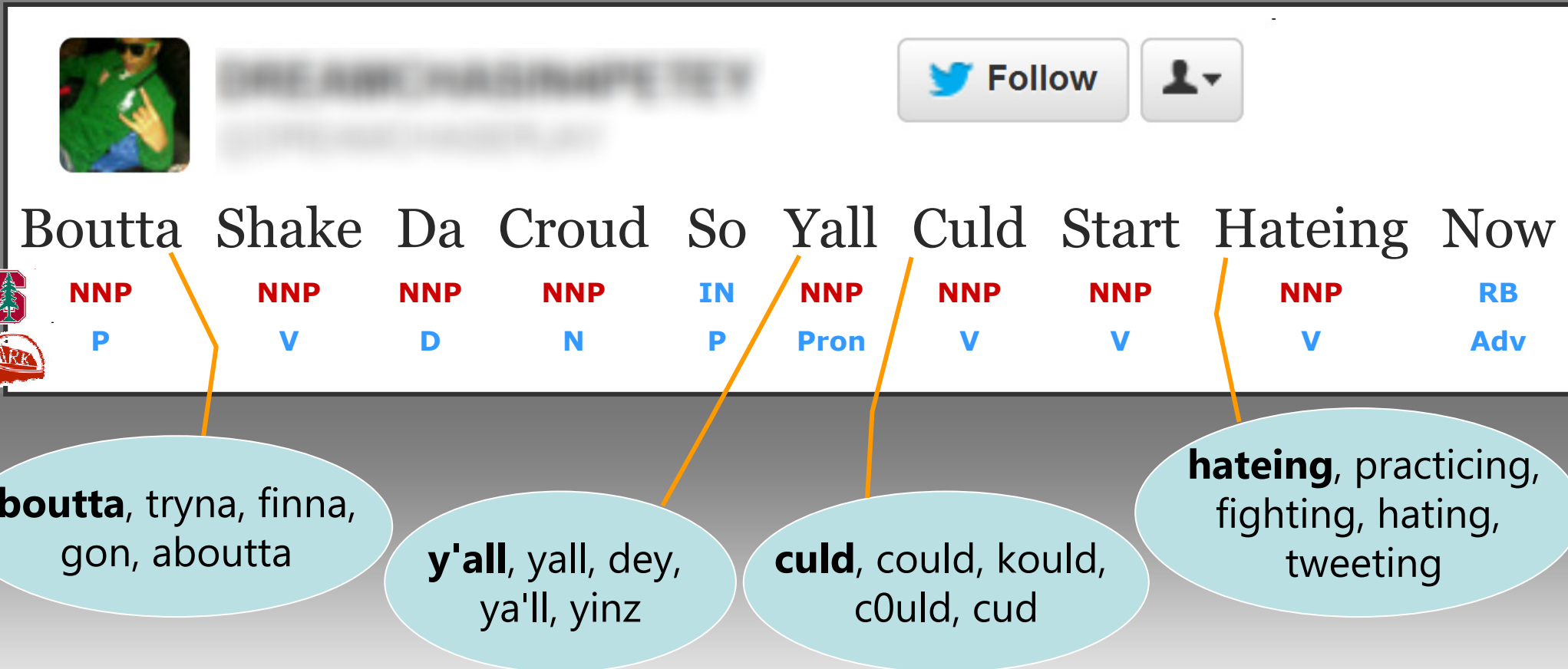


17. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters

Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, Noah A. Smith



The image shows a screenshot of a Twitter post with the text "Boutta Shake Da Croud So Yall Culd Start Hateing Now". Below the text, each word is annotated with a Part-of-Speech (POS) tag in red and a more specific tag in blue. Orange lines connect these tags to light blue ovals containing word clusters. On the left side of the tweet, there is a profile picture of a person in a green hoodie, a blurred name, and a "Follow" button. There are also small logos for "S" and "ARK" on the left.

Word	POS Tag	Specific Tag	Word Cluster
Boutta	NNP	P	boutta , tryna, finna, gon, aboutta
Shake	NNP	V	
Da	NNP	D	
Croud	NNP	N	
So	IN	P	y'all , yall, dey, ya'll, yinz
Yall	NNP	Pron	
Culd	NNP	V	culd , could, kould, c0uld, cud
Start	NNP	V	
Hateing	NNP	V	hateing , practicing, fighting, hating, tweeting
Now	RB	Adv	

Accuracy: **92.80%** on Twitter, **91.19%** on IRC

New **Twitter dataset** with updated annotation guidelines

Tagger + Data: www.ark.cs.cmu.edu/TweetNLP



Carnegie Mellon