

SlangNet: A WordNet like resource for English Slang

Shehzaad Dhuliawala, Diptesh Kanojia, Pushpak Bhattacharyya

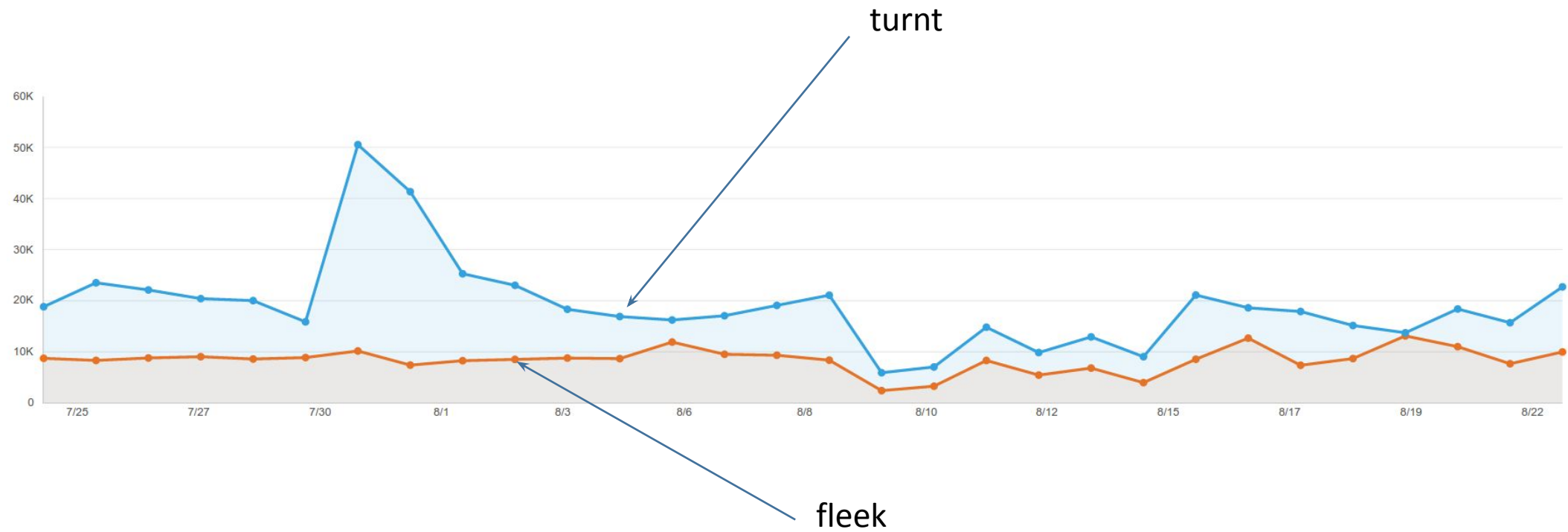
What's *cool* about slangs?

- For work dealing with informal text, usages of slangs has increased many fold.
- Several new words are coined on a daily basis and are used by several internet users who use microblogging services such as Twitter

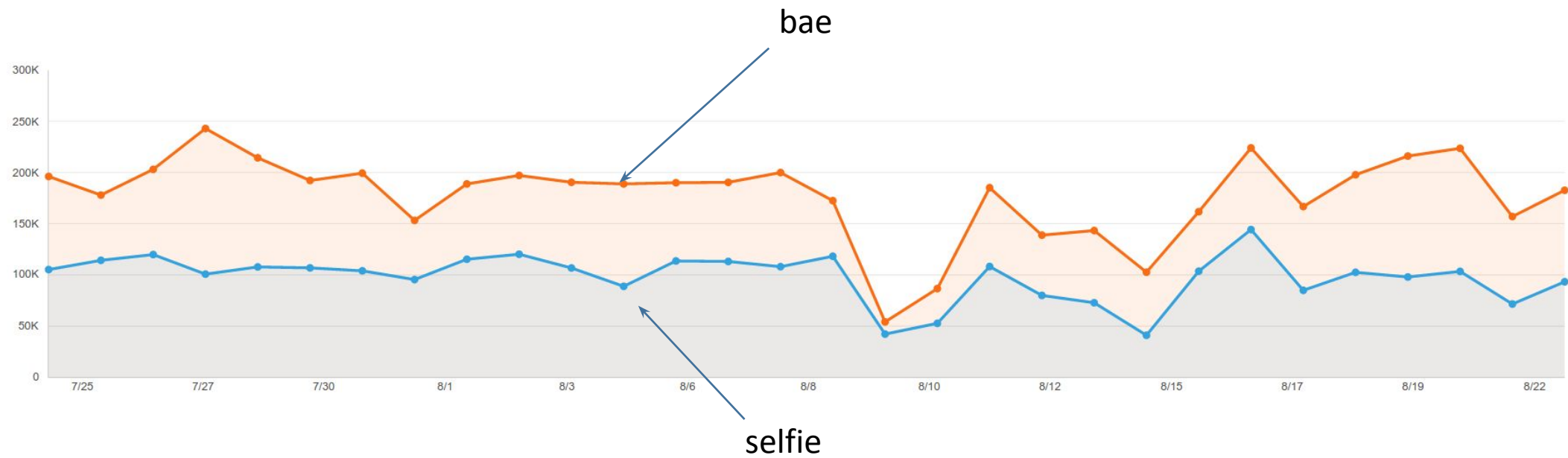
Urban slangs

- When you google "Top internet slangs", a varied list pops out.
- Some of them are words one has never heard of:
 - Fleek
 - selfie
 - Bae
 - Turnt
- Some of them are words one may have heard of
 - Ship
 - Feels
 - Basic
- Surprisingly (or, rather not) the meanings are nothing close to the traditional sense.

Usage statistics



Usage Statistics



Categorizing words

- We classify slang words into two categories:
 - **Newly created words** It includes words which don't presently exist in the English dictionary. These words are entirely new. Example: swag def.: style, turnt def.: drunk etc
 - **Newly created senses** This category includes words which do exist, but are commonly used with a different sense. Example: Cool (The best way to say something is neat, awesome, or swell) rather than the traditional sense: the quality of being at a refreshingly low temperature.

Creating a coarse resource

- For frequently occurring words we try to extract the relevant slang sense from Urban Dictionary.
- We do this by a simple bag-of-words approach of comparing the WordNet synsets and the key words from urbandictionary.com.

Eg: Green

TOP DEFINITION



green

yes... it is most definitely referring to the marijuana... and also, upon occasion, any substance with similar uses. too, properly, of course, it is a golfing term, but that;s really not very much fun, now is it?

'yo, where da green at?'

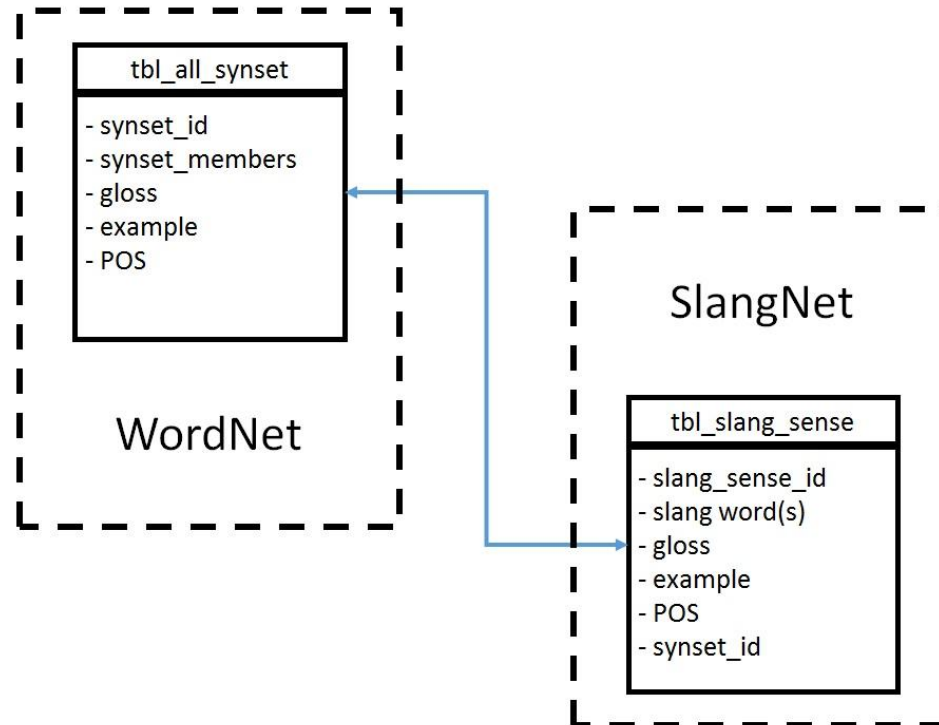
weed • marijuana • pot • bud
• ganja • grass • greens •
smoke • money • herb •
cannabis • blue • high •
red • mary jane • dank •
dope • chronic • bong • joint

Green -> Marijuana

Refining the Dictionary

- A lexicographer uses this coarse dictionary to further refine it
- For our green example, a new slang sense is added and linked to the WordNet synset [02953127]: [cannabis](#), **marijuana**, [marihuana](#), [ganja](#) (the most commonly used illicit drug; considered a soft drug, it consists of the dried leaves of the hemp plant; smoked or chewed for euphoric effect)
- If no pre-existing sense is found, a new sense is created for the word

UML Data Model diagram



Primary key: slang_sense_id

Foreign key: synset_id

Measuring the efficacy of our resource

- We use 2 datasets to measure the effectiveness of SlangNet:
 - In house created dataset where we sense annotate (WN senses + SlangNet sense) sentences from Reddit
 - Words: Cool (97%), Insane (88%), Own (56%)
 - On the Novel sense dataset provided by Cook et al. (2014)
 - Words: Follower, Poster, Text

Results (In-house dataset)

- Cool: Awesome
- Insane: Fantastic, terrific
- Own: Defeat someone in (mostly in an online game)

Word	Simple Lesk	Extended Lesk
Cool	0.15	0.17
Insane	0.38	0.54
Own	0.45	0.51

Results for Cook et al.'s dataset

- **Text:** A short message sent using a mobile device usually through protocol such as a short messaging service (SMS)
- **Poster:** A person who uploads a message or comment online on a social media
- **Follower:** Someone who follows your activity on a social media website or application

Word	Simple Lesk
Text	0.31
Poster	0.19
Follower	0.42

Conclusion and future work

- We show that the creation of a resource like SlangNet could be beneficial for detecting novel slang senses from informal text
- This lexical database could also be extended to hold sentiment information such as the SentiWordNet. This intuition comes about as several slangs are sentiment bearing