

STASH, Vol.14(2) - ☹️ Is that what they're calling marijuana now?

February 20, 2018

When people discuss substance use, they often use slang terms that can vary over time and from place to place. Healthcare professionals, public health practitioners, and researchers are faced with the challenge of keeping up to date with emerging drug and drug-related terms. This week, STASH reviews a study by Sean Simpson and his colleagues that demonstrates the usefulness of a synonym detection computer program to identify newly [emerging terms for drugs](#) by analyzing Twitter content.

What was the research question?

Can a synonym detection computer program discover new terms for marijuana people use on Twitter?

What did the researchers do?

The researchers used a series of computer programs to gather and process data from 82.6 million randomly selected tweets that were written in English and located in the U.S. The purpose of the programs was to identify words or phrases used as synonyms for the common slang terms “weed” and “ganja.” The authors then compared the terms identified by the synonym detection software with a list of slang terms compiled by experts in order to optimize the program’s precision¹ and recall.² This optimized computer program then generated a list of terms that human reviewers then double-checked.

What did they find?

The computer program identified 200 potential slang terms for marijuana. After reviewers vetted these terms, they determined that 50 referred to marijuana-related paraphernalia, and 65 referred to marijuana. When compared to the list of slang terms compiled by experts, 30 of the 65 marijuana terms were previously unknown. Interestingly, one of the new terms discovered was in fact a combination of emoji - ☹️.

Using Natural Language Processing to Discover New Slang Terms for Marijuana

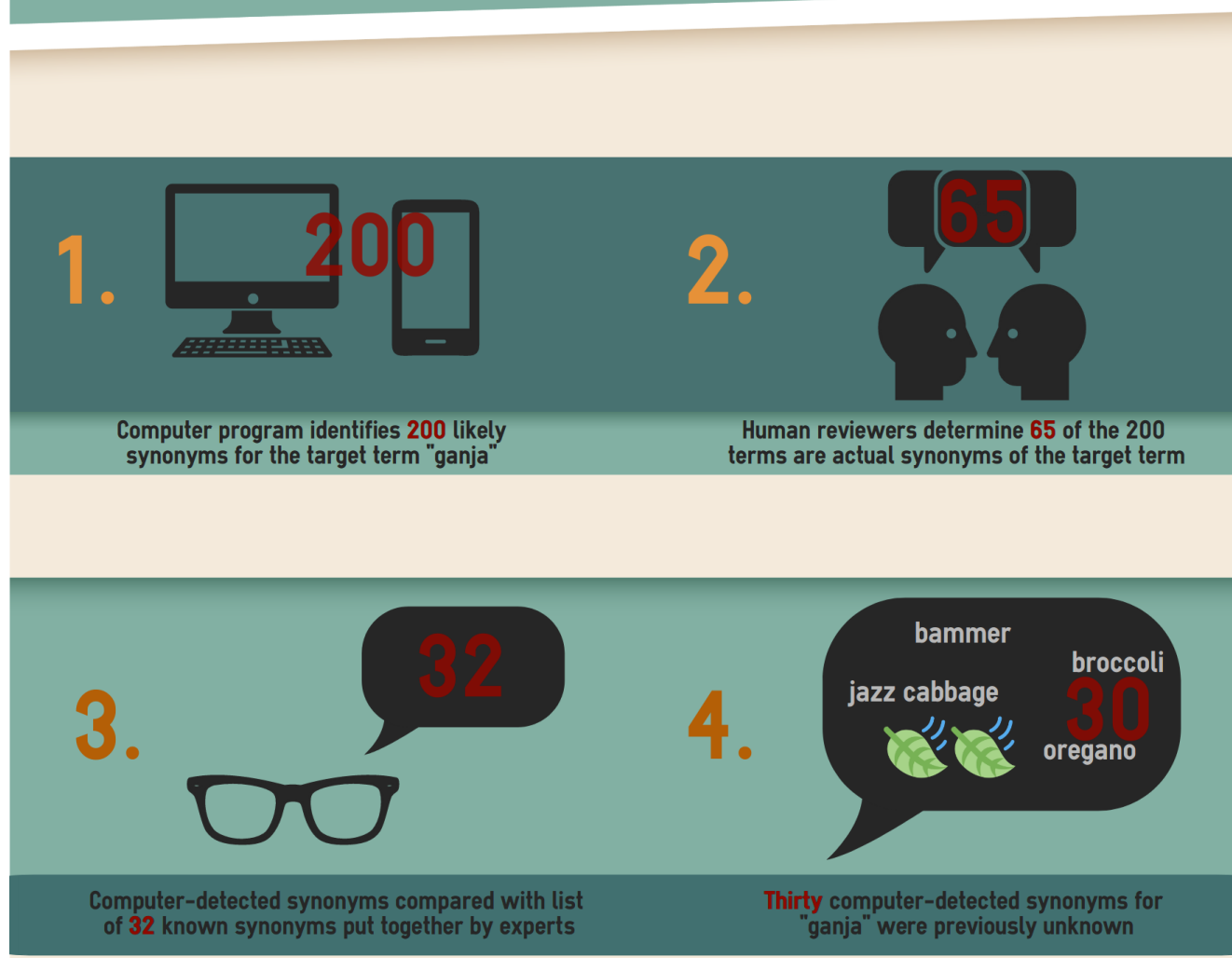


Figure. The author's process of identifying, validating, and evaluating potential new terms for marijuana. Figures in red indicate the number of terms considered at each stage of the process (modified from Simpson, Adams, Brugman & Connors, 2017). Click image to enlarge.

Why do these findings matter?

The findings indicate that this specific technique has promise as a means of discovering new drug terms as they emerge, though further development is needed to reduce the number of [false positives](#). A tool such as the one described could eventually be used without human oversight, speeding up the discovery and dissemination of not only new slang terms but new drug combinations and methods of ingestion.

Every study has limitations. What are the limitations in this study

In order to limit their sample of tweets to those emerging from the U.S., the authors limited their data to tweets that contained geotagged information. As geotagging a tweet is optional, this exclusion process might produce a [sampling bias](#). In other words, users who discuss especially taboo topics might not geotag their tweets to maintain more anonymity, resulting in a sample that is less taboo than the population it's intended to represent.

For more information

If you are concerned about your or a loved one's marijuana use, consider [Your First Steps to Change](#), a free anonymous self-help guide.

— Rhiannon Chou Wiley

What do you think? Please use the comment link below to provide feedback on this article.

[1] Minimizing identified terms that don't actually refer to marijuana

[2] Finding as many terms that truly refer to marijuana as possible