#1

Thank you very much for your comments and suggestions.

-To investigate the effects of different kinds of lexicons deeply, we conduct several experiments as follows:

-(also for #3) We also compared our methods with Duong (2016) and MultiCCA using their code on GitHub:

(More extensive comparisons will be conducted from now to better our camera-ready version.)

-Yes, it is true that in task 1 the annotators are not required to label the highly subjective social differences directly. However, we argue that since it is social media that the most frequently mentioned 20 terms shown to annotators came from, and thus the similarity between the two lists intrinsically indicates much “social elements”. Also, we are very interested in your suggestion of detecting differences between superficially synonyms in the future, in which ground truth would be harder to obtain. In this work, we focused on the named entities.

- As bilingual lexicon induction for slang is still a novel and open problem, we believe both explanation and direct mapping are beneficial to solve the problem. In task 2, we actually first focus on explanation task (given a slang term, find the most similar regular terms in another language), and then we take a step forward to do the direct mapping (find similar slang terms). The ground truth terms (“foolish” etc.) are used for the quantitative evaluation for the explanation task, so it’s natural they should be not slang.

- The word vectors used in ACS computation should come from a third-part pre-trained one (we used a public Glove pre-trained one) instead of from the systems being tested, since ACS is just to measure the similarity between two word lists. Thus, the terrible situation will not occur. Table6 is showing the sum of ACSs over 200 times of slang translation. (Thank you very much! We should have mentioned these two details in footnotes.)

- The “asymmetry”