
Anonymous Author(s)

Affiliation

Address

email

1 Thanks to all of you for your thoughtful reviews and very useful suggestions.

2 **Reviewer 1**

- 3 a) "it's good science to do more careful experimentation on existing techniques and new
4 combinations of existing techniques. I think this is a valuable contribution." **RESPONSE:**
5 We agree and believe systematic comparisons of unsupervised cross-lingual learning methods
6 are particularly important at a time where this area is getting very crowded.
- 7 b) "How was the 2% threshold for defining "failure" chosen?" **RESPONSE:** We follow previ-
8 ous work in using an absolute threshold, as well as maximum scores. (Artetxe et al., 2018,
9 uses 5%, for example.) In practice, performance for unsuccessful runs tends to be either >.1
10 or 0, so a different threshold would be unlikely to change results. We will include mean and
11 standard deviation in the revised version of the paper, but note that maxima highlight the
12 potential of methods.
- 13 c) We are also happy to include results on less difficult language pairs, but would like to point
14 out that unsupervised cross-lingual learning is *only relevant for low-resource languages*,
15 which tend to be typologically different from English/Spanish and therefore difficult.
- 16 d) "I found Section 4.2 difficult to understand." **RESPONSE:** Thanks for the suggestions,
17 which we will implement in the revised version.
- 18 e) "But this comparison fails to make a connection with Section 4.1 in two ways." **RESPONSE:**
19 Sorry if this was not clear: MUSE is the FAIR system consisting of GAN+Procrustes, so
20 GAN=C-MUSE. Both C-MUSE+Procrustes and C-MUSE+SBDI use cosine-based model
21 selection (csls).

22 **Reviewer 2**

- 23 a) We agree our paper presents a "detailed and fair comparison" and "show that combining
24 GANs with stochastic dictionary induction gives a new state of the art". We do not agree
25 this "is not enough for a full NeurIPS paper." This is a crowded area, with new methods
26 being proposed all the time. The world does not necessarily need more methods, but to
27 understand what works (when), and what does not.
- 28 b) You state that our paper needs "a new insight or method that improves the current perfor-
29 mances of unsupervised word translation methods". While this was not our main goal, we
30 do, as you say, "show that combining GANs with stochastic dictionary induction gives a new
31 state of the art". This, we believe, qualifies as an insight improving the current performance
32 of our methods.

33 **Reviewer 3**

- 34 a) We agree our main contribution is "to fairly compare many methods in a standardized
35 fashion", and that, in addition, we also present a new model selection criterion and establish
36 a new state of the art.
- 37 b) We like the idea of "aligning 3 or more languages in a shared embedding space", but this
38 goes well beyond the standard scenario explored in this paper.