Chinese Room Objectives

- Machine translation thirsts for bitext, especially in-domain.
- Low-resources languages are low on bitext, especially in-domain.

Primary Objectives

- Enable people to translate from low-resource language to English, even without any prior knowledge of source language.
- Build in-domain bitext for tuning, ideally some more for training.

Secondary Objective

- Support computational linguists in identifying challenges of a specific low-resource language.

Approach

- Reuse machine translation resources such as t-tables and special modules (e.g. for quantities, named entities) to build a glossing tool to support human translators.
- Allow user to explore alternative translations.
- Combine artificial intelligence and human intelligence.

Some Challenges

- Foreign scripts can present a massive cognitive barrier.
- Inconsistent spelling for many low-resource languages due to dialects, lack of spelling standards, lack of education.

Solution: Universal romanizer *uroman* (Ulf Hermjakob et al., ACL 2018).

- yaponie fukushima 1-yadro elektir istansisining toet genratorlar guruppisi

Experiments

- Built *Chinese Rooms* for Bengali, Hungarian, Kinyarwanda, Oromo, Sinhalese, Somali, Swahili, Tagalog, Tigrinya, Uyghur.
- Trained more than 20 people to use the *Chinese Room*.
- Successfully improved MT system for low-resource languages.