Improving SHARE translation verification

Yuri Pettinicchi, *Max-Planck-Institute for Social Law and Social Policy (MEA-MPISOC)*

Joint work with Yi-Chen Liu & Alexander M. Fraser, *Ludwig-Maximilians-Universität München (CIS, LMU Munich)*

BigSurv20 - 6th November 2020
Session: **Improving survey questions using machine learning and AI**

Improving SHARE translation verification

⚙ Improving

   ⚙ Adding an extra tool to the translator's working environment

⚙ SHARE

   ⚙ Survey of Health, Ageing and Retirement in Europe

⚙ translation

   ⚙ From English to European languages

⚙ verification

   ⚙ "Quality" control & management of translation procedures

Motivation

- Avoid avoidable mistakes

- High volume of work, time pressure, complexity

- Empowering translators with Machine Translation (MT) tools

- Competitive advantages

  - Keep humans working on what they do best

  - Let algorithms take care of simple and repetitive tasks

Exercise - Ingredients

- ⚙ Survey items
  - ⚙ Source text: English – Target Text: German
  - ⚙ Questions, Instructions, Response options: 242 items (11080 words)
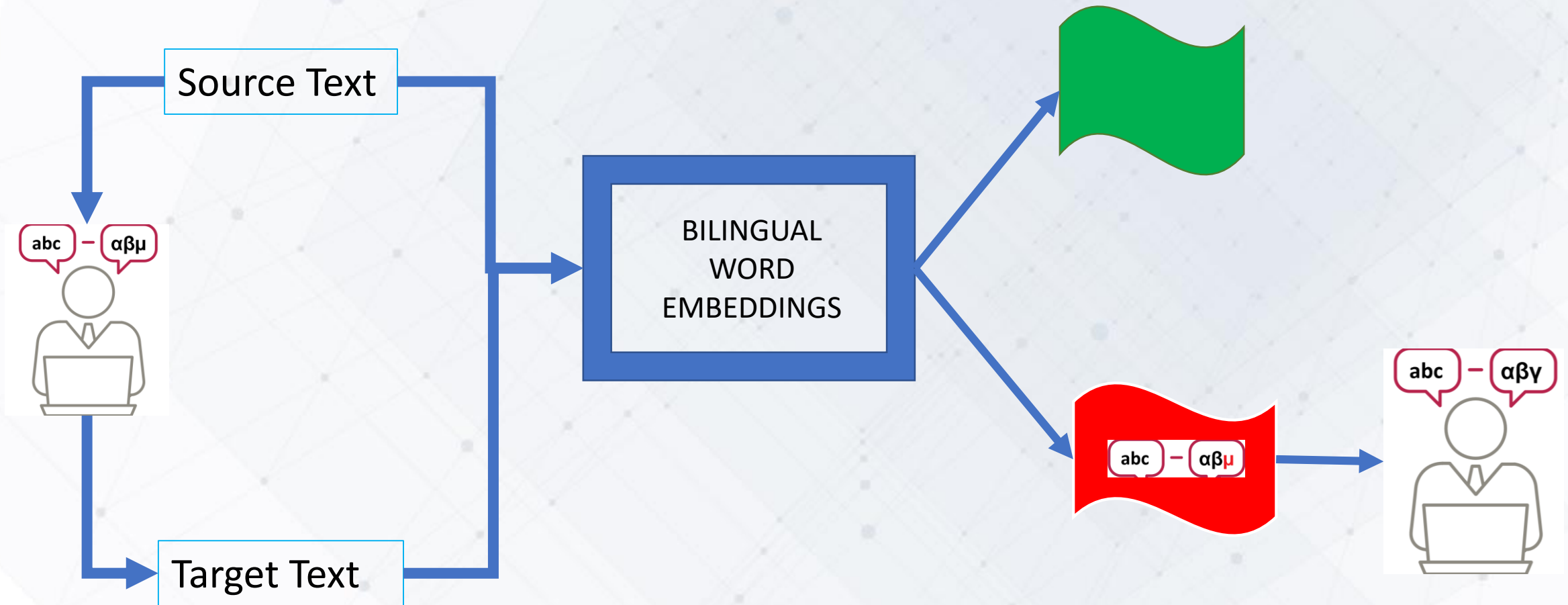 - ⚙ 4 country teams: DE, AT, LU, CH
- ⚙ Bilingual word embeddings
- ⚙ Flagging rules
  - ⚙ Computing the Translation Score
  - ⚙ Flag items when score is below a given threshold

Exercise - Workflow

BILINGUAL WORD EMBEDDINGS

Source Text

Target Text

BILINGUAL WORD EMBEDDINGS

- Using approach proposed by Artetxe et al. (2018)
- Training unsupervised bilingual word embeddings
  - trained on monolingual corpora in the source language and the target language
  - Corpus: general and domain-specific (survey questions) text
- Performing bilingual lexicon induction with trained DE-EN embeddings
  - match correct word translations
  - compute translation verification score

$$Translation\ Score = \frac{\#Matched\ Words}{\#Human\ Translation\ Words}$$

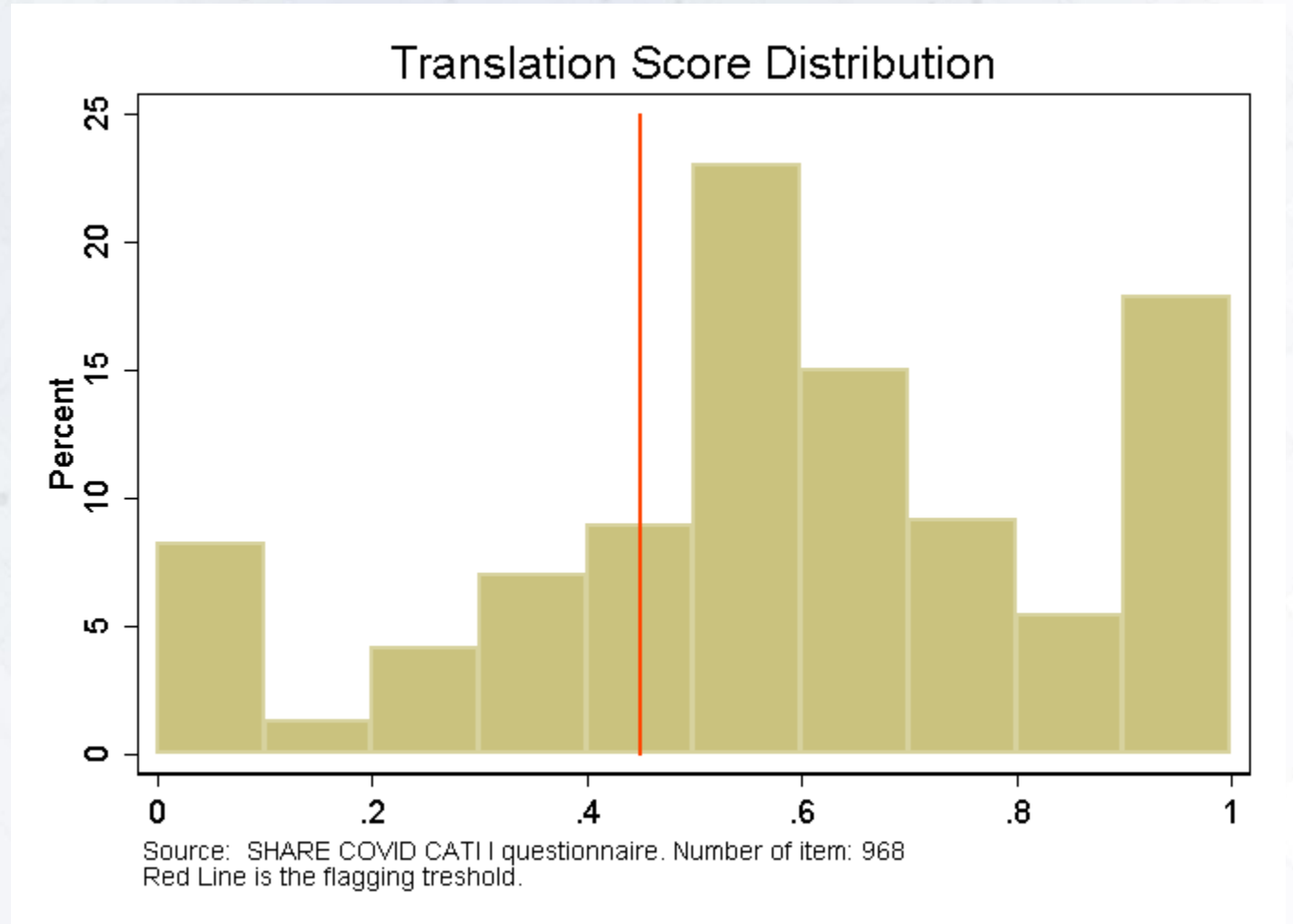# Results

⚙ **Translation Score**
  ⚙ Mean: 0.59 – SD: 0.28

⚙ **Flagged Items**
  ⚙ 255 items (26,3%)

⚙ **Outcome**
  ⚙ 73 actual positive (27,1%)
  ⚙ 182 false positive (72,9%)

Translation Score Distribution

Source: SHARE COVID CATI I questionnaire. Number of item: 968
Red Line is the flagging treshold.

Results – by country

⚙ Translation Score
  - ⚙ Mean: 0.59 – SD: 0.28

⚙ Flagged Items
  - ⚙ 255 items (26,3%)

⚙ Outcome
  - ⚙ 73 actual positive (27,1%)
  - ⚙ 182 false positive (72,9%)

	Austria	Germany	Luxembourg	Switzerland
N	242	242	242	242
Translation Score	.593 (.285)	.612 (.286)	.559 (.285)	.594 (.281)
Flagged Items	58 (24%)	59 (24%)	78 (32%)	60 (25%)
Outcome	0 (0%) 58 (100%)	4 (6%) 55 (94%)	56 (72%) 22 (28%)	9 (15%) 51 (85%)

Results – by type of text

- ❇ **Translation Score**
  - ❇ Mean: 0.59 – SD: 0.28

- ❇ **Flagged Items**
  - ❇ 255 items (26,3%)

- ❇ **Outcome**
  - ❇ 73 actual positive (27,1%)
  - ❇ 182 false positive (72,9%)

	Qtext	Qinstruction	Response Option
N	512	116	340
Translation Score	.61 (.230)	.470 (.201)	.59 (.362)
Flagged Items	82 (16%)	55 (47%)	118 (35%)
Outcome	28 (34%) 54 (66%)	17 (31%) 38 (69%)	24 (20%) 94 (80%)

Results – by length

⚙ Translation Score
- ⚙ Mean: 0.59 – SD: 0.28

⚙ Flagged Items
- ⚙ 255 items (26,3%)

⚙ Outcome
- ⚙ 73 actual positive (27,1%)
- ⚙ 182 false positive (72,9%)

#words	1-3	4-10	11-20	20+
N	318	242	254	154
Translation Score	.633 (.397)	.548 (.277)	.588 (.153)	.567 (.132)
Flagged Items	92 (29%)	92 (38%)	44 (17%)	27 (17%)
Outcome	23 (25%) 69 (75%)	21 (22%) 71 (78%)	13 (29%) 31 (71%)	12 (44%) 15 (56%)

Lessons learnt

⚙ Efficient process

⚙ High false positive rate

    ⚙ Higher for high quality translations
    ⚙ Higher for response options
    ⚙ Higher for short sentences

⚙ List of "unmatched" words – improving the training data

⚙ Next step:

    ⚙ Training bilingual phrase embeddings
    ⚙ Analysis of the changes made (actual positive): cosmetics, wording, fluidity,…

SSHOC
social sciences & humanities open cloud

SHARE
SURVEY OF HEALTH, AGEING
AND RETIREMENT IN EUROPE

Join our community

🌐 https://www.sshopencloud.eu

✉️ info@sshopencloud.eu

🐦 @SSHOpenCloud

in /in/sshopencloud

Back-up – bilingual word embeddings (BWE)
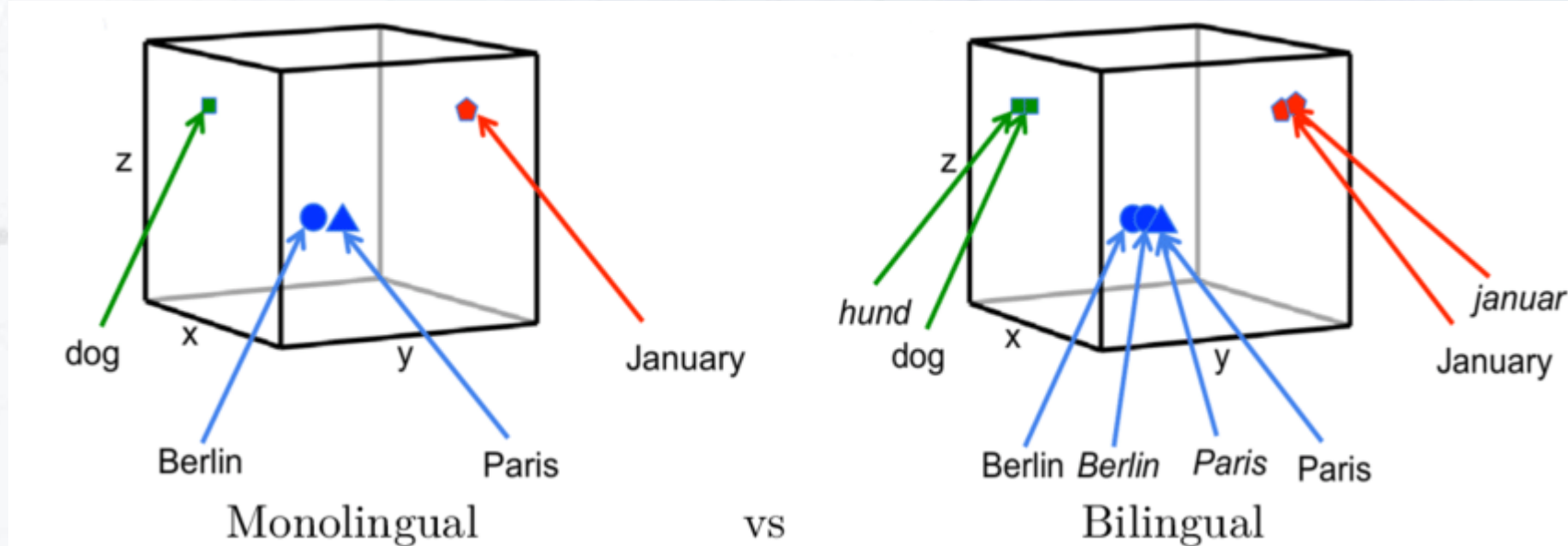
A toy 3D shared bilingual embedding space from Gouws et al. (2015): While in monolingual spaces words with similar meanings should have similar representations, in bilingual spaces words in two different languages with similar meanings should have similar representations (both mono-and cross-lingually).
Source: Vulić, Ivan & Moens, Marie-Francine. (2015). Bilingual Distributed Word Representations from Document-Aligned Comparable Data. Journal of Artificial Intelligence Research. 55. 10.1613/jair.4986.