OECD-GESIS Seminar on Translating and
Adapting Instruments in Large Scale Assessments

SHARE
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE

Automatic translation verification: can we rely on MOSES?

7-8 June 2018 – Paris

Yuri Pettinicchi
Jeny Tony Philip

HORIZON 2020
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536
Federal Ministry of Education and Research
NIH National Institute on Aging
mea MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY
Munich Center for the Economics of Aging
SHARE ERIC

Motivation

Why do we need to use machines to verify translation?

- General Aim
  - Avoiding avoidable mistakes
  - Improving data quality

Specific to SHARE:

SHARE - cross-national survey
- 28 national CAPIs
- 40 languages

- Translation verification:
  - National CAPI tools should correctly display translated text
  - Green flag to go on-field

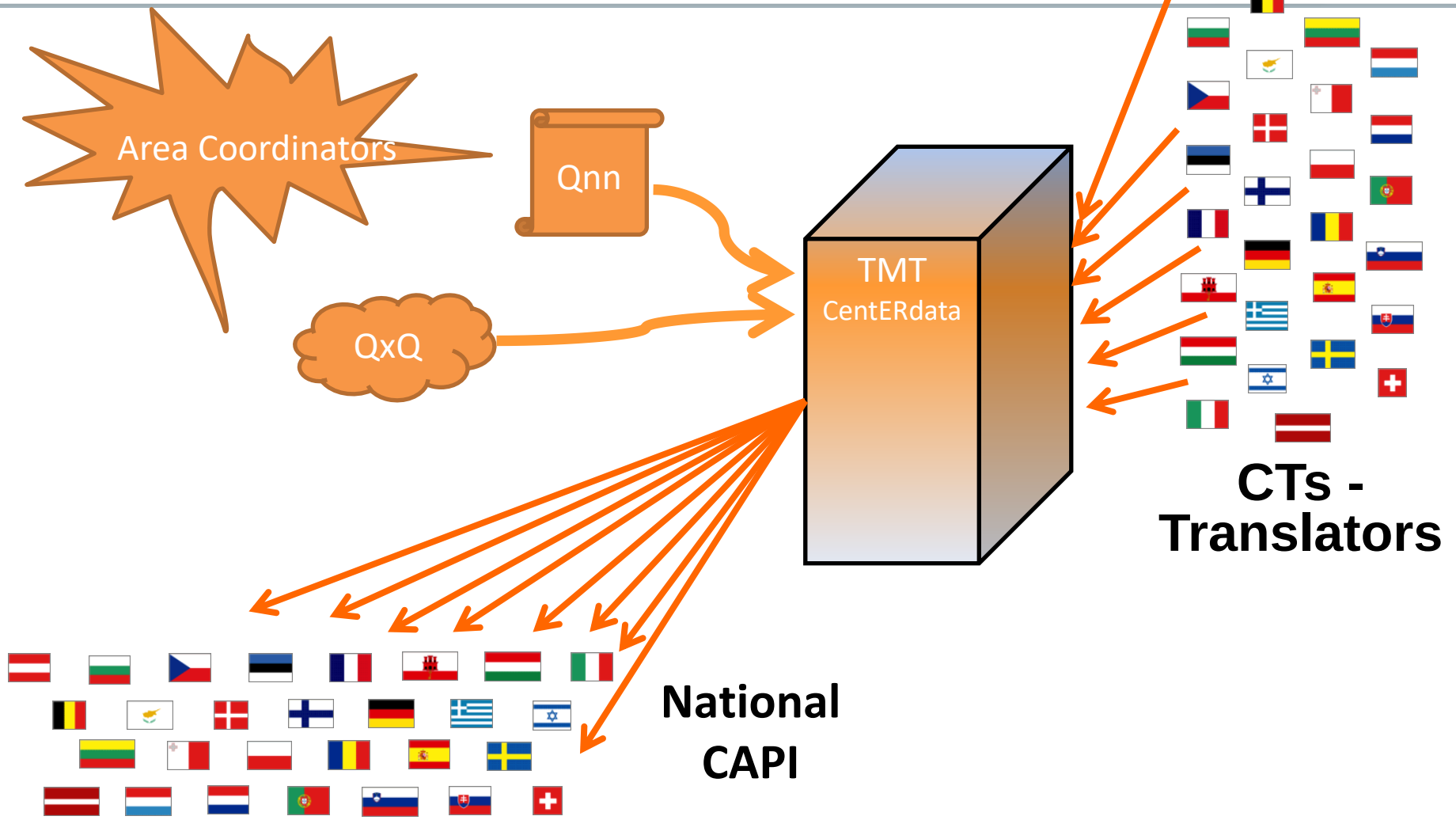
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Background

- Workflow
- Workload
  - Volume
  - Time
  - Costs
- Possible issues

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Workflow



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

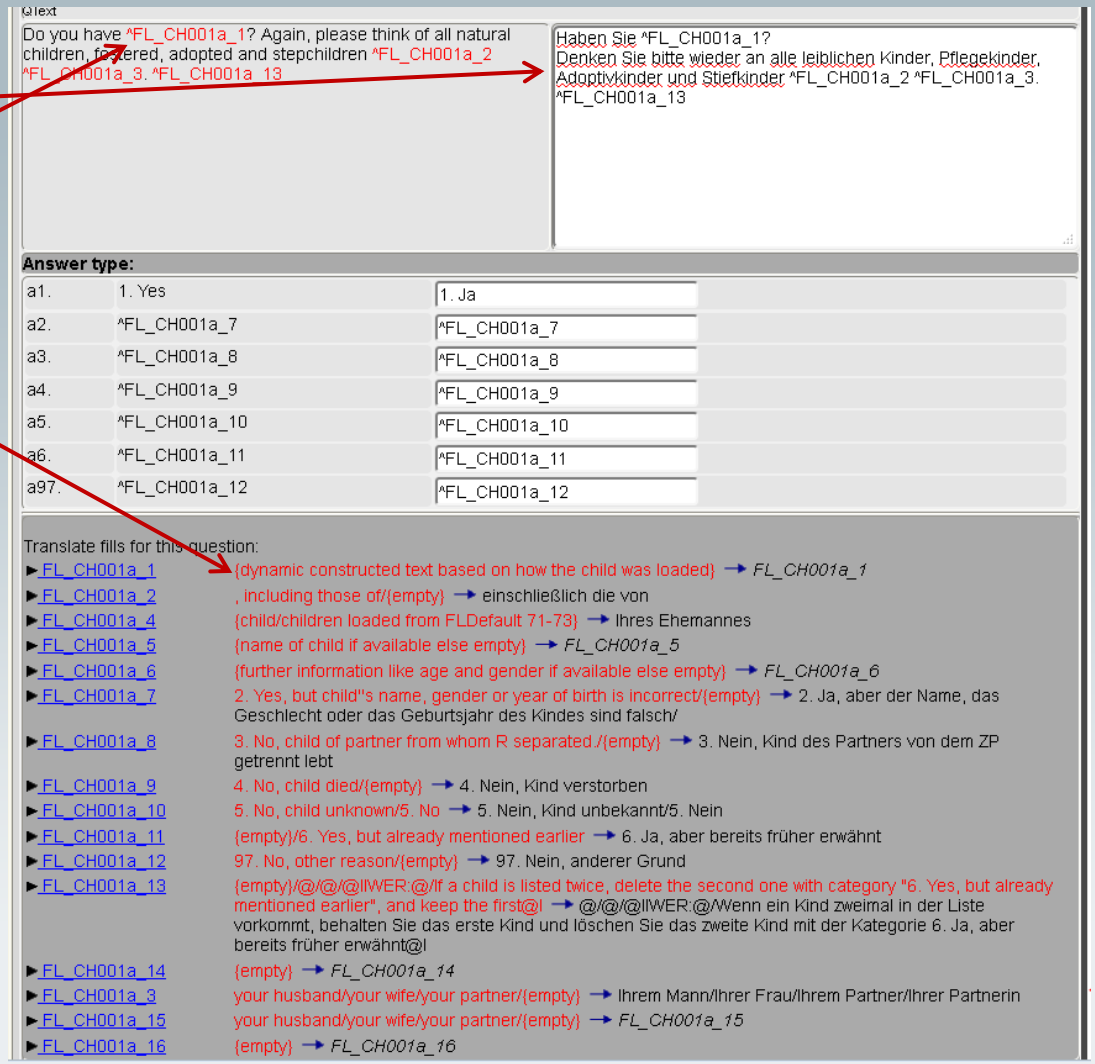
Workload - 1

Multiple fills in the translation

Dynamic fill with hidden text

Gender specific fills

Numbered answer options



QText

Do you have ^FL_CH001a_1? Again, please think of all natural children, fostered, adopted and stepchildren ^FL_CH001a_2 ^FL_CH001a_3. ^FL_CH001a_13

Haben Sie ^FL_CH001a_1? Denken Sie bitte wieder an alle leiblichen Kinder, Pflegekinder, Adoptivkinder und Stiefkinder ^FL_CH001a_2 ^FL_CH001a_3. ^FL_CH001a_13

Answer type:

a1.	1. Yes	1. Ja
a2.	^FL_CH001a_7	^FL_CH001a_7
a3.	^FL_CH001a_8	^FL_CH001a_8
a4.	^FL_CH001a_9	^FL_CH001a_9
a5.	^FL_CH001a_10	^FL_CH001a_10
a6.	^FL_CH001a_11	^FL_CH001a_11
a97.	^FL_CH001a_12	^FL_CH001a_12

Translate fills for this question:

- FL_CH001a_1 {dynamic constructed text based on how the child was loaded} → *FL_CH001a_1*
- FL_CH001a_2 , including those of/{empty} → einschließlich die von
- FL_CH001a_4 {child/children loaded from FLDefault 71-73} → Ihres Ehemannes
- FL_CH001a_5 {name of child if available else empty} → *FL_CH001a_5*
- FL_CH001a_6 {further information like age and gender if available else empty} → *FL_CH001a_6*
- FL_CH001a_7 2. Yes, but child"s name, gender or year of birth is incorrect/{empty} → 2. Ja, aber der Name, das Geschlecht oder das Geburtsjahr des Kindes sind falsch/
- FL_CH001a_8 3. No, child of partner from whom R separated./{empty} → 3. Nein, Kind des Partners von dem ZP getrennt lebt
- FL_CH001a_9 4. No, child died/{empty} → 4. Nein, Kind verstorben
- FL_CH001a_10 5. No, child unknown/5. No → 5. Nein, Kind unbekannt/5. Nein
- FL_CH001a_11 {empty}/6. Yes, but already mentioned earlier → 6. Ja, aber bereits früher erwähnt
- FL_CH001a_12 97. No, other reason/{empty} → 97. Nein, anderer Grund
- FL_CH001a_13 {empty}/@/@/@IWER:@/If a child is listed twice, delete the second one with category "6. Yes, but already mentioned earlier", and keep the first@I → @/@/@IWER:@/Wenn ein Kind zweimal in der Liste vorkommt, behalten Sie das erste Kind und löschen Sie das zweite Kind mit der Kategorie 6. Ja, aber bereits früher erwähnt@I
- FL_CH001a_14 {empty} → *FL_CH001a_14*
- FL_CH001a_3 your husband/your wife/your partner/{empty} → Ihrem Mann/Ihrer Frau/Ihrem Partner/Ihrer Partnerin
- FL_CH001a_15 your husband/your wife/your partner/{empty} → *FL_CH001a_15*
- FL_CH001a_16 {empty} → *FL_CH001a_16*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Federal Ministry of Education and Research

National Institute on Aging

MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY
Munich Center for the Economics of Aging

Workload - 2

- Volume

Lang	EN
Items	1099
Text	6837
Words	31000+

Lang	40
Items	43960
Text	273480
Words	1270000+

- Time 1 year

- Costs 12 PM per language (480PM)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Issues

- Misspelling a word (minor)
- Empty fields / missing sentence
  - (full sentence vs part of a sentence)
- *Flipped* translations (major)
  - Negative effects on questionnaire routing

GENERIC	NATIONAL
1. Employed	1. Arbeitslos
2. Unemployed	2. Abhängig

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Automatic Translation verification

- **Ideal environment**
  - High volume of data
  - Several languages
  - Available in the translator working environment
- **Expected results**
  - Improving quality of the translation
  - Reducing testing time

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Experiment design

1. Automatic back translation

- Automatic translation of French version back to English using web interfaces provided by third parties

2. Compute the BLEU index

- Calculate BLEU score per item using Python's NLTK library for symbolic and statistical natural language processing.
  - The Bilingual Evaluation Understudy Score(BLEU) is used to evaluate a generated sentence in comparison to a reference sentence.
  - A perfect match produces a score of 1.0, whereas a perfect mismatch produces a score of 0.0.

3. Compute the Similarity index

- Calculate (cosine) similarity score per item using SciPy open-source Python-based solution for mathematics, science and engineering.
  - Cosine similarity is computed by representing datasets as vectors(vectorization) and then comparing the angles formed by the vectors

4. Check flagged items with the help of human translators

HORIZON 2020
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536
Federal Ministry of Education and Research
NIH National Institute on Aging

MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY
Munich Center for the Economics of Aging
SHARE ERIC

Back-translation

Machine translation:

In-house solution	Market alternatives
Moses	DeepL
	Google Translate
	Bing

Trained on parallel corpora

In-house solution	Market alternatives
Domain specific	General corpora
Phrase-based	Neural network

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

The Moses Experience - 1

- Statistical machine translation system (SMT)
  - Trained on parallel corpora
    - **News-commentary**
    - Europarl v7.0
    - UN Corpus

- Open source software: http://www.statmt.org/moses/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

The Moses Experience - 2

MOSES - Statistical translation machine

Specs	
Language Model:	KenLM
Operating System:	Ubuntu-16.04.2
Corpus	News-commentary

Training:

Specs	
Training time:	5h 30m
Tuning time:	1 h 20 m
Binarization time:	5m
Testing time	20m

Performance

BLEU (Bilingual Evaluation Understudy) score: 22.95

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

The SHARE questionnaire

Sample	N	%
All	2465	100
Plain	849	32,47
Fills	1616	61,8
NaN	150	5,7
Short (<17)	1325	50,67
Long (>17)	1140	43,59

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Evaluations

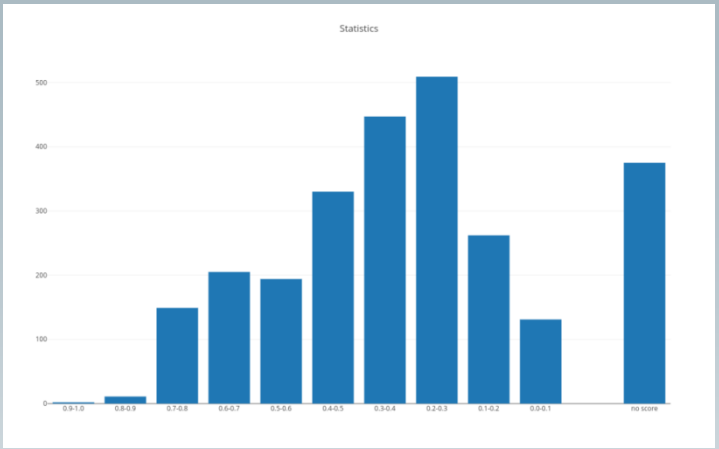
Sample		Deepl	Google	Bing	Moses	Moses (!=0)
All	BLEU	0.77	0.74	0.62	0.06	0.47
	Similarity	0.87	0.88	0.87	0.67	0,69
Plain	BLEU	0.76	0.76	0.72	0.03	0,70
	Similarity	0.85	0.86	0.85	0.63	0,67
Fills	BLEU	0.78	0.70	0.46	0.17	0,4
	Similarity	0.89	0.89	0.88	0.69	0,7
Short (<17)	BLEU	0.81	0.78	0.68	0.06	0,39
	Similarity	0.84	0.85	0.83	0.60	0,64
Long (>17)	BLEU	0.24	0.30	0.29	0.23	0,47
	Similarity	0.91	0.91	0.90	0.75	0,85

HORIZON 2020
Union's Horizon 2020 research and innovation programme under grant agreement No 676536
and Research
on Aging

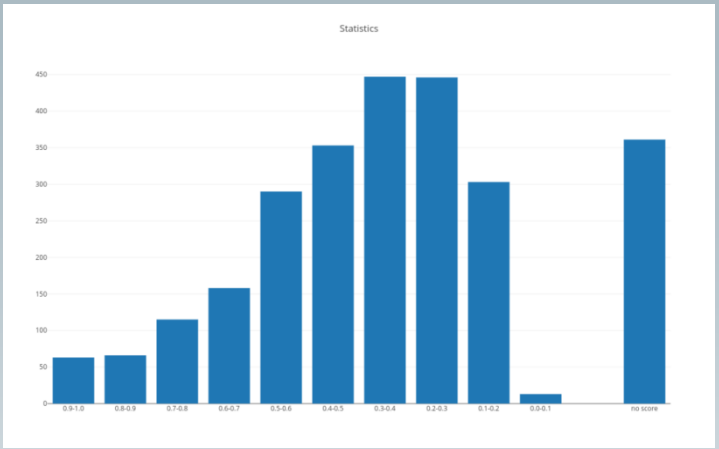
SOCIAL LAW AND SOCIAL POLICY
Munich Center for the Economics of Aging
SHARE ERIC

Results – BLEU Score

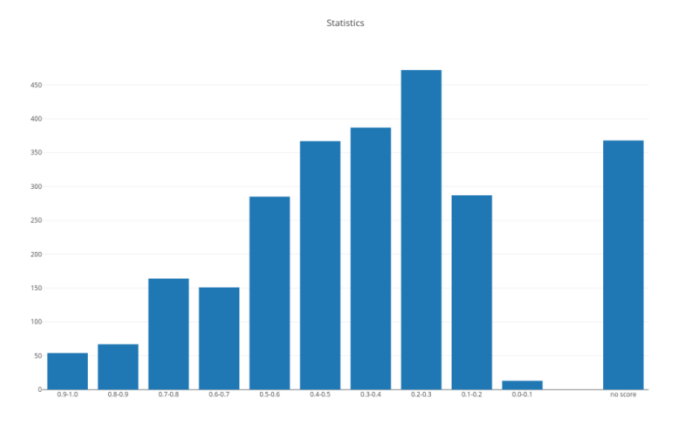
Moses



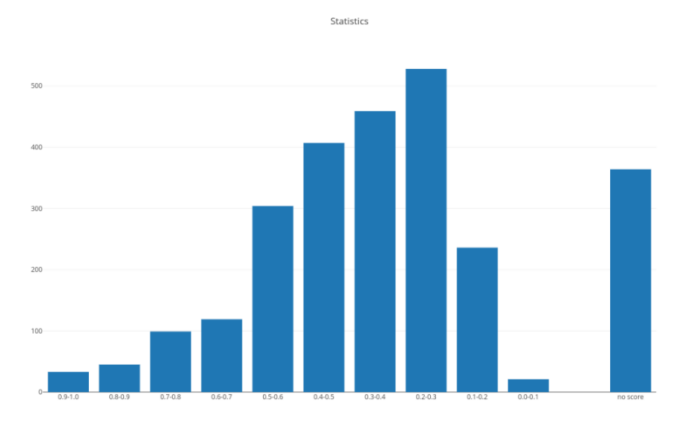
DeepL



Google



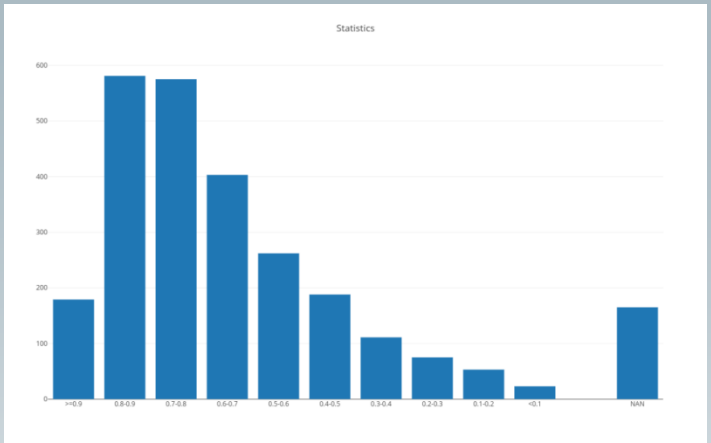
Bing



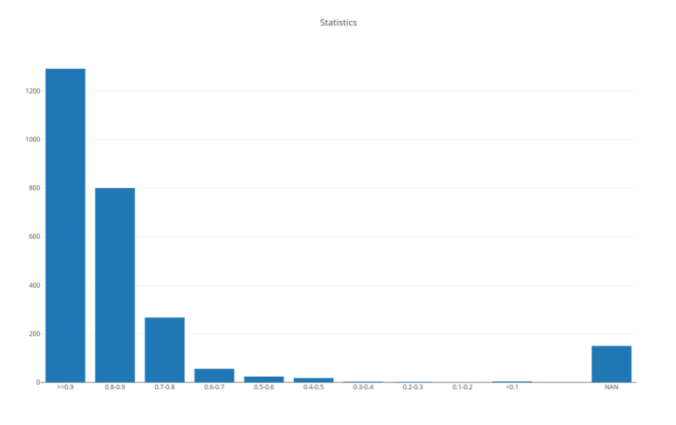
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Results – Similarity Score

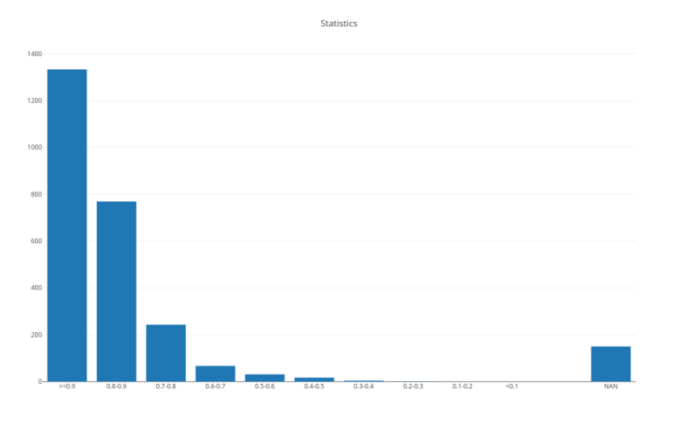
Moses



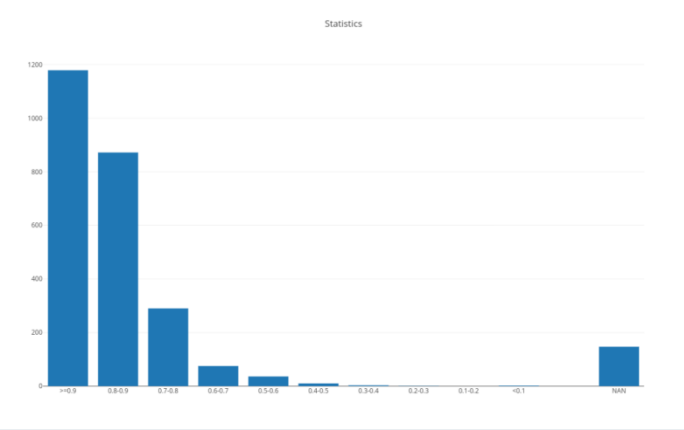
DeepL



Google



Bing



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Flagged

Sample		Deepl	Google	Bing	Moses
All	Flagged	9	5	6	283
	False Negative	7	3	4	281
Plain	Flagged	9	5	6	153
	False Negative	7	3	4	151
Fills	Flagged	0	0	0	130 (19)
	False Negative	n.a.	n.a.	n.a.	130 (19)
Short (<17)	Flagged	9	5	6	262
	False Negative	7	3	4	260
Long (>17)	Flagged	0	0	0	21
	False Negative	n.a.	n.a.	n.a.	21

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Conclusions

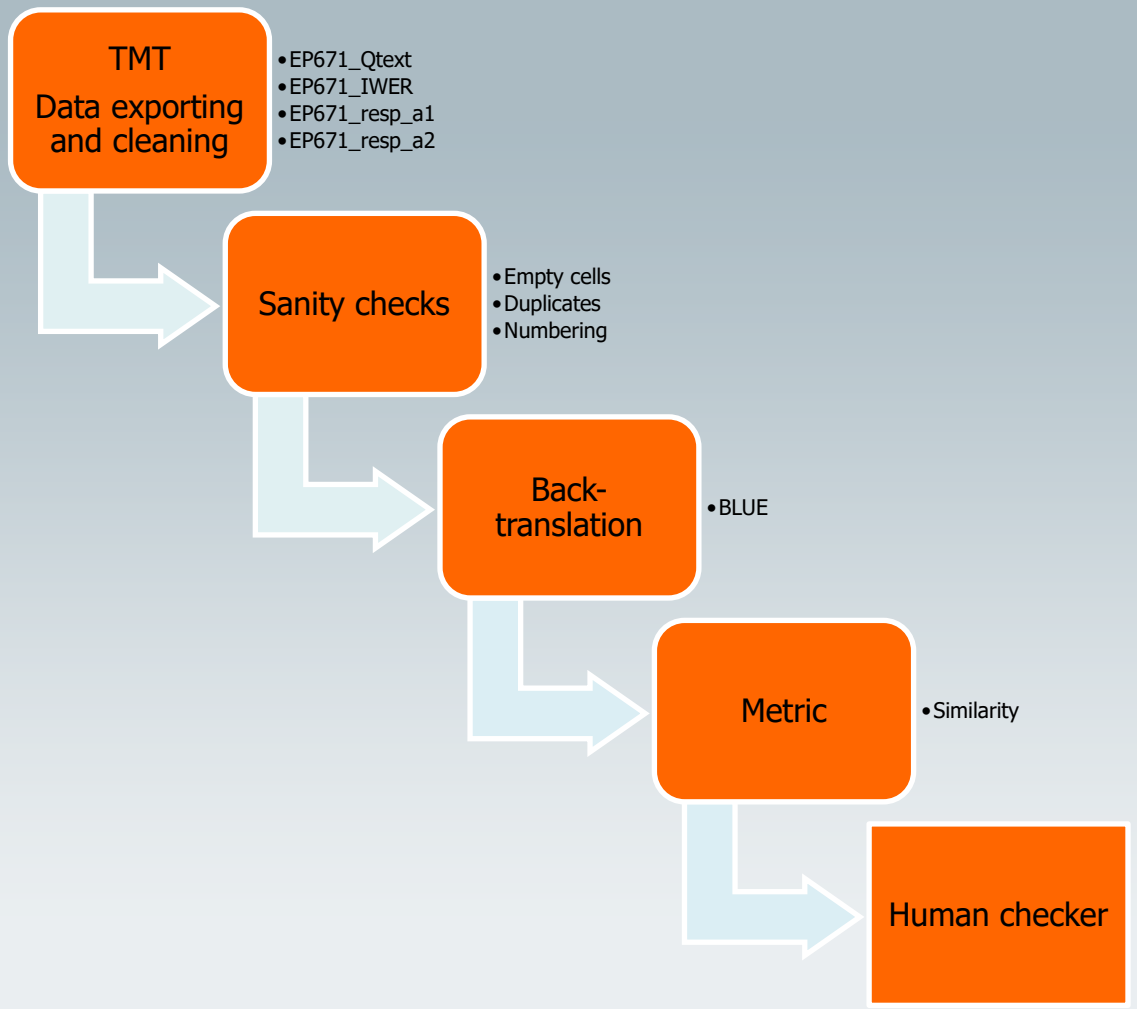
- Lessons Learnt:
  - „In-house" solutions require a heavy investment of time, manpower and IT infrastructure.
  - Market solutions are more user friendly but are limited by their „one size fits all" design.
  - Overall, market solutions are more effective.

- Scope for collaboration across cross-national surveys
  - Compilation of a domain specific parallel corpora.
  - Joint effort to build a machine translation tool.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Questions?

Translation verification in SHARE



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676536

Overview of the third-party technology

Setting and Performance details

VER 1

- What was used:
  - web interfaces
  - ScyPy
  - NLTK

  Word2vec

- Measurements taken:
  - Cosine similarity (or similarity)
  - BLEU score