



CHIST-ERA
Challenge Call Topic
“Science in Your Language”

François Yvon, LIMSI





Topic Description

- ❑ This challenge is fully automated translation for scientific materials from and into English
 - ✓ Into English: to reduce publication barriers and efforts for scholars from non-English countries; to increase the value of scientific writing in languages other than English
 - ✓ From English: to increase access to scientific materials for the largest possible population of stakeholders (scholars, engineers, policy makers, citizens)

- ❑ Main objective:
 - ✓ acceptance at a top tier conference of an automatically translated paper, for several target domains of interest



- ❑ A challenge for the AI / Machine Learning / Natural Language Processing / Machine Translation community
- ❑ Scientific challenges for Machine translation (MT):
 - ✓ MT of scientific papers: translating complex documents (graphs, images, tables, references) with intricate cross-references
 - ✓ MT in technical domains: terminology and phraseology issues
 - ✓ Document level MT: enforcing global constraints in translation
 - ✓ MT in multiple domains and languages, with little, if any, domain adaptation data
- ❑ This is a new challenge for MT:
 - ✓ existing evaluation campaigns mostly focus on the translation of isolated sentences
- ❑ Potential to attract attention from translation studies, text mining, for scientific texts, scientometrics, and scholars from multiple disciplines



- ❑ Improvement of translation quality for scientific texts would also
 - ✓ improve document-level MT for other types of texts
 - ✓ improve Natural language Processing for complex document types
 - ✓ develop automated understanding / text-mining / visualization technologies for scientific content in languages other than English

- ❑ Two types of end users :
 - ✓ non-English writers of scientific material
 - ✓ non-English readers of scientific papers and derived materials (protocols, surveys, funding schemes, scientific programmes, teaching material, etc)

- ❑ The focus on publishable translation
 - ✓ increase the focus on high-quality translation
 - ✓ create new (train / test) resources for translating scientific texts
 - ✓ requires to develop new metrics for document level translation



Evaluation Campaign Scenario

- ❑ Evaluations is based on several rounds of:
 - a) Organizer announce test domain(s) and languages, provide development data
 - b) Organizer collect actual papers for translation from and into English
 - c) Participant automatically translate test set
 - d) Organizer automatically evaluate results
 - e) (Optional: translated texts are revised by professional translators)
 - f) Open discussion of results and methods

- ❑ Step (a)-(d) are typical of evaluations campaigns in MT
 - ✓ used in many projects / shared tasks : WMT, WAT, IWSLT, etc.
 - ✓ accept submissions from teams outside the challenge
 - ✓ can be done completely online based on existing infrastructure

- ❑ Possible organizers in Europe: LNE, ELDA/ELRA, labs. involved in the organization of WMT, IWSLT and similar campaigns



- ❑ Multiple available evaluation metrics for MT
 - a) automatic comparison with a reference (BLEU, METEOR, TER, and many more) on a per sentence basis
 - b) automatic evaluation of document-level processing issues (NEW)
 - c) crowdsourced human judgements (direct assessment of fluency, adequation)
 - d) human post-editing effort (hTER, etc)

- ❑ Metrics for (a) and (c-d) are well defined
 - ✓ Could be readily used for the first round

- ❑ Open questions remain:
 - ✓ How to evaluate at the document-level ?
 - ✓ Design evaluation protocol that would incentivize participation (submission of papers) from scholars from multiple domains



Costs of Evaluation Campaigns

- ❑ Main costs for the organizers
 - ✓ data collection and preparation, including professional quality of texts
 - ✓ development costs for submission / human evaluation platform,
 - ✓ development cost for new metrics
 - ✓ organization of workshops to present / discuss results



❑ Facilitating factors:

- ✓ many past shared tasks in NLP / MT, lots of existing work on evaluation design, software, metrics
- ✓ experience of MT for medical abstracts (WMT shared task since 2014); for texts in the medical domain (H2020 project HIML, MLIA –covid19 evaluation campaign); for scientific documents (new shared tasks at WAT 2020)
- ✓ huge open archives of open scientific material in English (arxiv, biorxiv, medrxiv, HAL, openAire, etc), as well as other languages
- ✓ national conferences accepting publications in languages other than English in many domains and countries
- ✓ growing interest for NLP in scientific domains (eg sciBERT from OpenAI, claim checking in scientific texts, improved scientometrics, arxiv labs, etc)