Human Factors in Neural Machine Translation

Joss Moorkens

15th March 2019

Université Paris Diderot

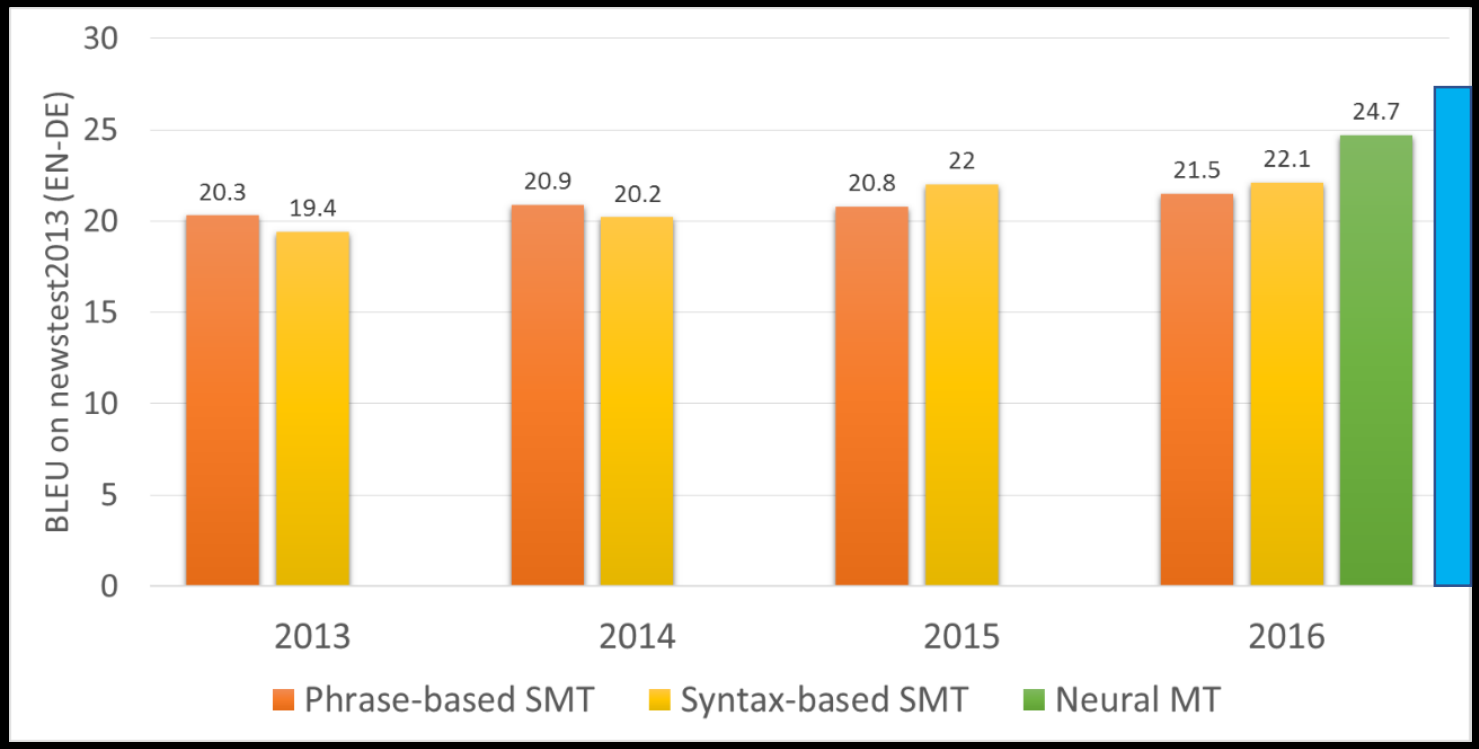
Contents

- Neural Machine Translation (NMT)
- Human Factors in NMT
- Translation industry context
- Reasons to be positive

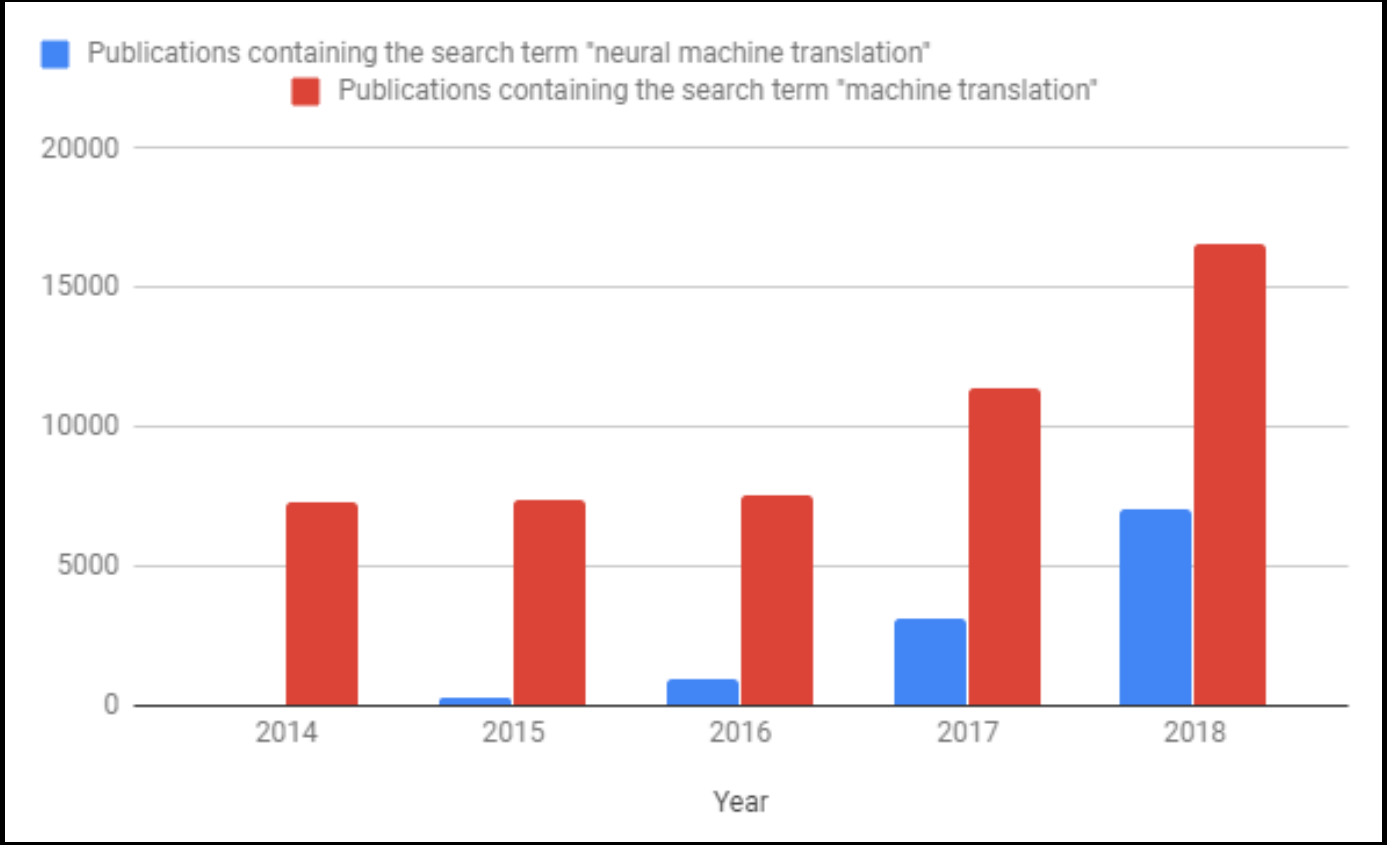
Machine translation for users

- **Assimilation**: gist translation
- **Dissemination**: "machine translated as an intermediate step in production" (Forcada 2010)

What is the State of the Art for Machine Translation?

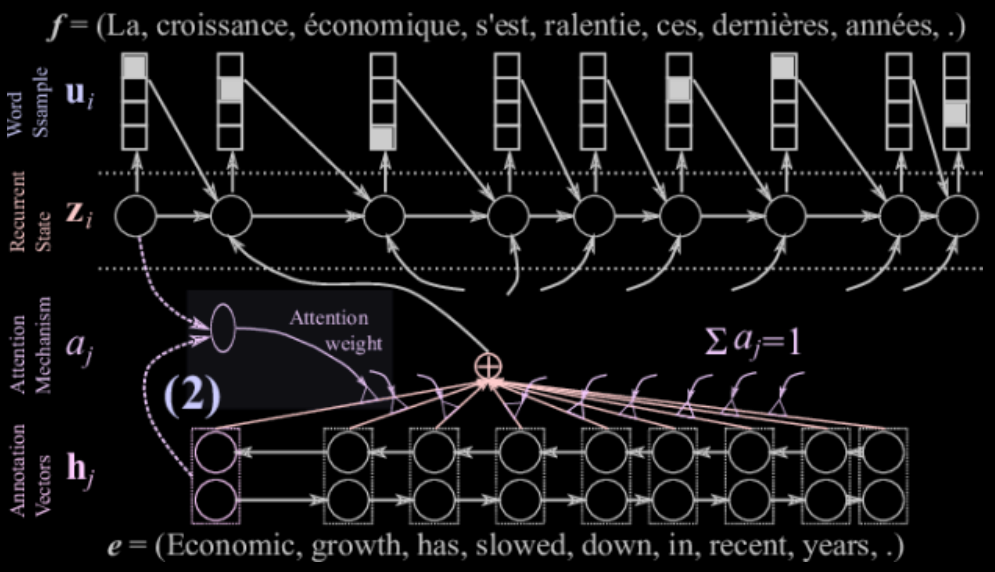


Google Scholar



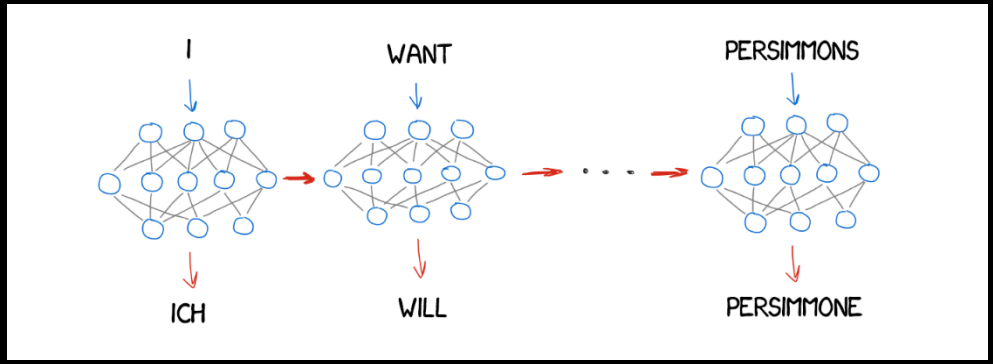
Neural Machine Translation (NMT)

- **SMT** – *translation model* predicts most likely translation, *language model* predicts probability of seeing that target segment in target language, plus many small sub-components that are tuned separately
- **NMT** - build and train a **single, large neural network** that reads a sentence and outputs a correct translation (Bahdanau, Cho, Bengio 2015)



Neural Machine Translation (NMT)

- Uses a Recurrent Neural Network (RNN) to deal with variable segment length
- NMT predicts a target word based on the context associated with source and previously generated target words
- A small neural network, called an *attention mechanism* analyses context for every source word



source: https://medium.freecodecamp.org/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5

NMT: Pros and Cons

- Main strength of NMT…
  - is grammatical improvements, but possible degradation in lexical transfer (Neubig, Morishita, Nakamura 2015)
- Some problems
  - Computationally expensive
  - Networks have fixed vocabulary → poor translation of rare/unknown words
  - Poor domain adaptation
  - Problems with long sentences
  - Models are trained on parallel data; how do we use monolingual data?
  - Difficult to incorporate terminology
    - (see also Six Challenges for NMT by Koehn and Knowles 2017)

Neural versus Phrase-Based Machine Translation Quality: a Case Study (Bentivogli et al. 2016)

Results show that NMT system outperforms all other approaches.

- Post-edit effort lower (-26%) on all sentence lengths
- Fewer morphology errors (-19%), lexical errors (-17%), and word order errors (-50%)
- Improvement in placement of verbs (-70% errors)

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation (Wu et al. 2016)

Results show that NMT system strongly outperforms other approaches

- Improved translation quality for morphologically rich languages
- Human evaluation ratings closer to HT than PBSMT
- Additional tweaks required for NMT to perform well on "real data"

Evaluating MT for Massive Open Online Courses: A multifaceted comparison between PBSMT and NMT systems (Castilho et al. 2018)

4 datasets (250 segments) from real EN MOOC data translated into German, Greek, Portuguese, and Russian using TraMOOC engines

PB-SMT/NMT mixed, random task order

2-4 professional translators

MT engines trained on same data: open corpora plus educational data from Coursera, Qatar Educational Domain, EU Teachers' Corner

Castilho et al. 2018 (2)

- Comparative ranking of 100 randomised translations
- Post-editing using PET (Aziz, Castilho, Specia 2012)
  - Temporal effort – time spent post-editing (Krings 2001)
  - Technical effort – edit count
  - Cognitive effort – pause-word-ratio (Lacruz, Denkowski, Lavie 2014)
- Rating of fluency and adequacy
- Error annotation
  - Inflectional morphology, Word order, Omission, Mistranslation, Addition

Ratings of adequacy: mixed results

How much of the meaning expressed in the source fragment appears in the translation fragment?

1. None of it
2. Little of it
3. Most of it
4. All of it

	EN-DE		EN-EL		EN-PT		EN-RU	
% scores with 3-4 adequacy value (SMT, NMT)	**73.5**	66.4	**89**	**89**	94.7	**97.1**	72.8	**77.5**
% scores with 1-2 adequacy value (SMT, NMT)	26.5	33.6	11	11	5.3	2.9	27.2	22.5

Post-editing: temporal effort

Words per second (all PEs)	SMT	NMT
German	0.21	0.22
Greek	0.22	0.24
Portuguese	0.29	0.30
Russian	0.14	0.14

SMT, NMT	German		Greek		Portuguese		Russian	
POST-EDITED SENTENCES (CHANGED)	**940**	**813**	**928**	**863**	**874**	**844**	**930**	**848**
UNCHANGED SMT, NMT	**60**	**187**	**72**	**137**	**126**	**156**	**70**	**152**

Previous work by Moorkens & O'Brien (2015) found an average speed of 0.39 WPS for EN-DE professional PE.

Technical Post-Editing Effort (keystrokes/segment)

Language	System	Mean	Std. Deviation
DE	**SMT** **NMT**	**5.8** **3.9**	**1.84** **1.63**
EL	**SMT** **NMT**	**13.9** **12.5**	**0.16** **1.31**
PT	**SMT** **NMT**	**3.8** **3.6**	**1.68** **1.91**
RU	**SMT** **NMT**	**7.5** **7.2**	**4.99** **5.80**

Some examples

- ST: It's about copy-paste from pdf to wiki card.
- NMT: É sobre copiar-pasta de pdf para wiki card.
- SMT: Trata-se de copiar e colar de pdf para cartão wiki.

- ST: Would you send just 10 materials that are the most suitable.
- SMT: Würden Sie nur 10 Materialien, die am besten geeignet sind.
- NMT: Schicken Sie einfach 10 Materialien, die am besten geeignet sind.

Some examples

- ST: I am just making sure that I understand this correctly.
- SMT: Estou só para ter a certeza que entendi corretamente.
- NMT: Eu estou apenas me certificando de que eu entendo isso corretamente.

- ST: was webinar live today?
- HT: O webinar foi ao vivo hoje?
- NMT: Será que o webinar vive hoje?
- SMT: Foi webinar vivem hoje?

More examples

- EN: We begin our exploration today by looking at a particular ad that appeared in American magazines in recent years.
- PBSMT: Heute beginnen wir unsere Erforschung von einem bestimmten Ad anschaue, die auf amerikanischen Zeitschriften erschienen in den letzten Jahren.
- NMT: Wir beginnen unsere Forschung heute mit einer bestimmten Werbung, die in den letzten Jahren in amerikanischen Zeitschriften veröffentlicht wurde.

What can we expect from NMT?

- Fluency is improved, word order errors are fewer using NMT
- Fewer segments require editing using NMT
- NMT produces fewer morphological errors
- No clear improvement for omission or mistranslation using NMT
- NMT for dissemination: no great improvement in post-editing throughput, effort
    - "Errors are more difficult to spot"

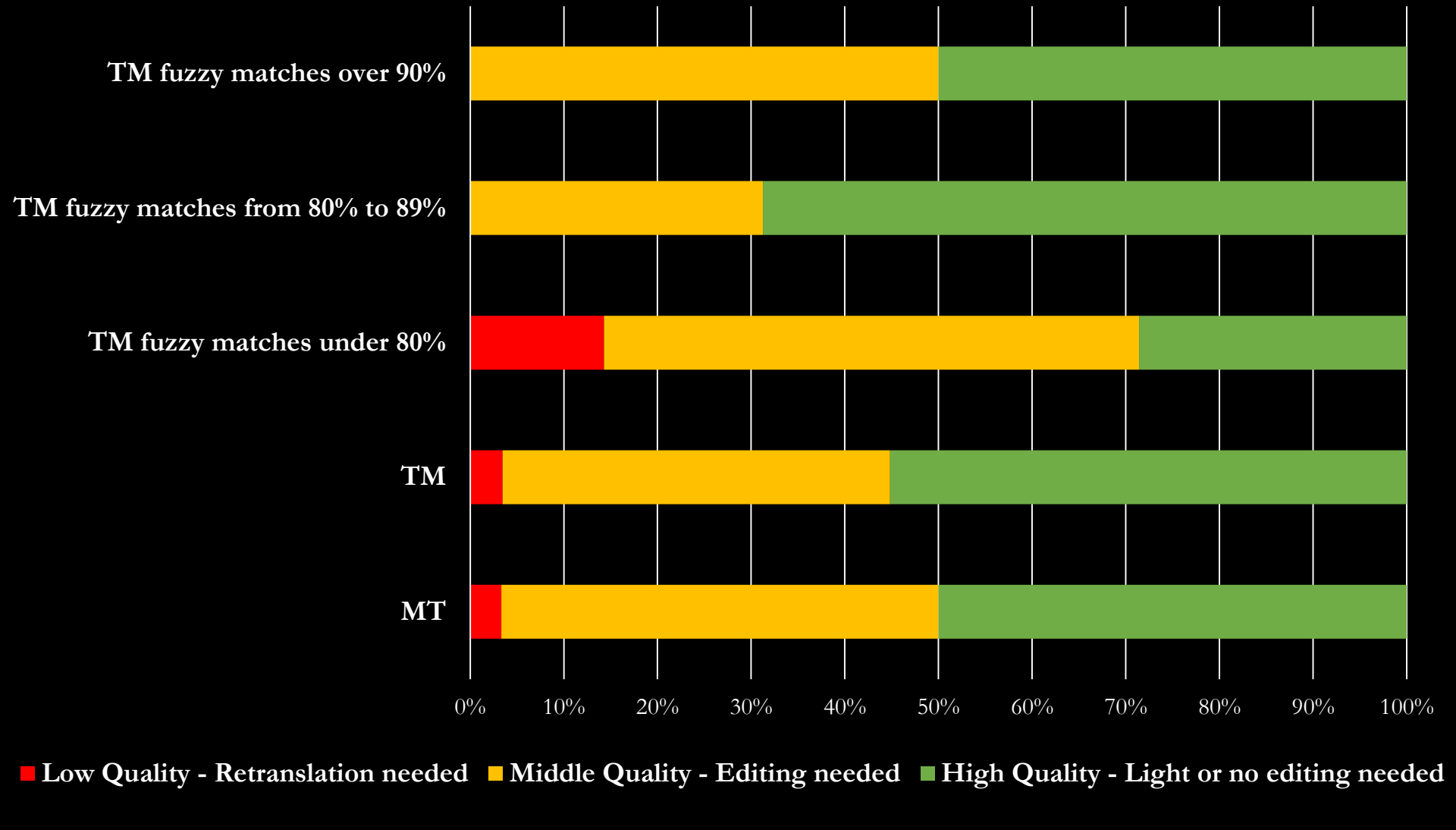
Interactions with NMT (2019)

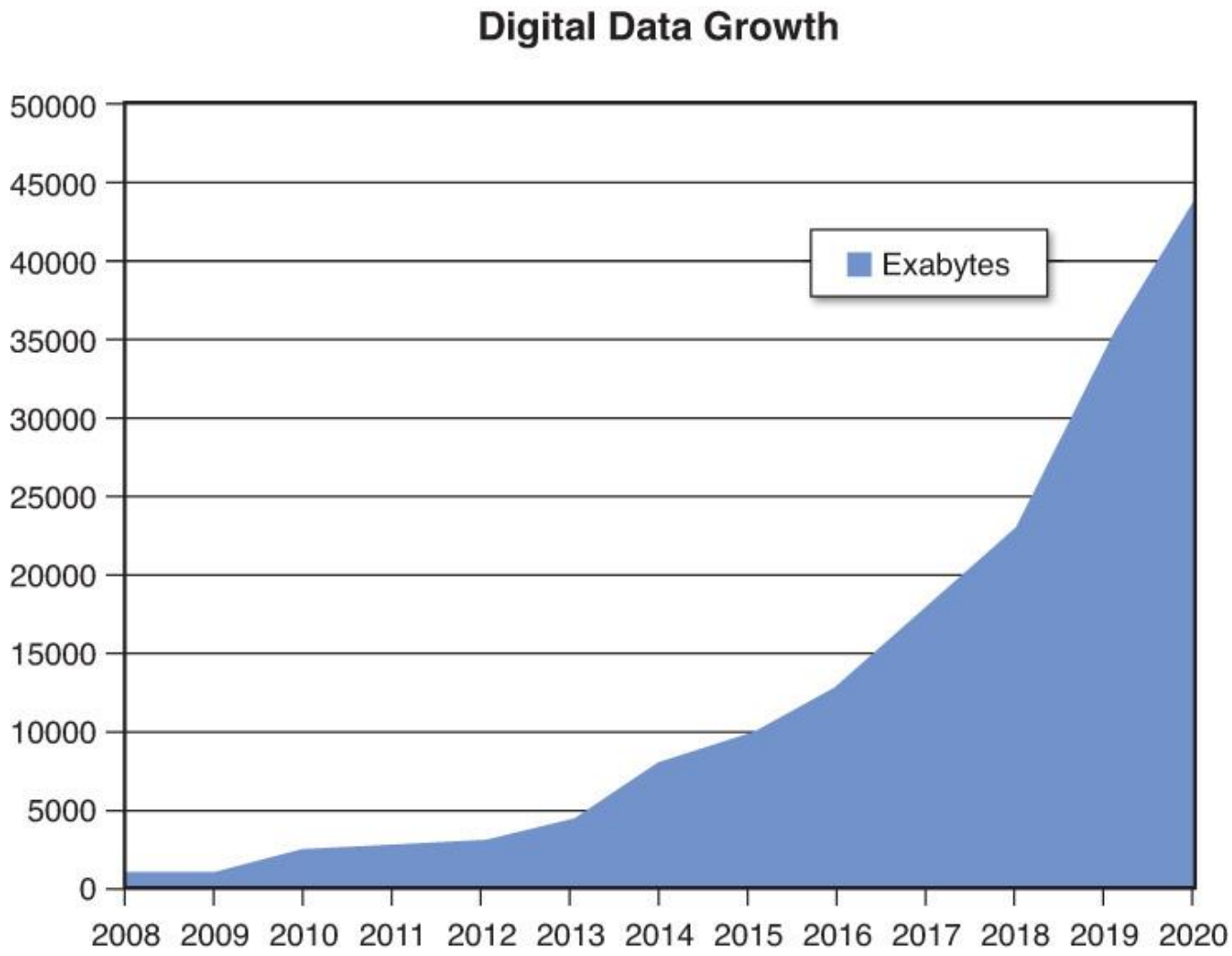
- **Little or no productivity gain in different domains & language pairs (EN-ES, ES-DE, EN-NL…)**
- **Where there are productivity gains they are minor, not statistically significant. Often productivity disimprovement in longer sentences**
- **NMT rated highly for adequacy and fluency, automatic metrics show positive results for NMT, and technical effort is less for NMT post-editing.**
- **Adaptive NMT: editing time is not improved when compared with adaptive SMT, with the caveat that the number of participants is small. Users tend to prefer working with the NMT output.**

Segment editing times (Sánchez-Gijón, Moorkens, Way 2019)

	NMT	TM
All segments	**22.53**	**20.47**
Segments of under 10 words	13.47	15.08
Segments from 10 to 19 words	20.63	16.54
Segments of over 20 words	33.24	35.40

Perceived quality (Sánchez-Gijón, Moorkens, Way 2019)





Source: International Data Corporation (IDC) 2017

STAFFING INDUSTRY ANALYSTS

SIA is the Global Advisor on Staffing and Workforce Solutions

REGION ▶ Americas

Home | Events | Memberships | Research | Editorial

DAILY NEWS

VIEW ALL NEWS

UNIVERSITY LISTS TOP 10 EMERGING CAREERS FOR COLLEGE GRADS IN THE US; INTERPRETERS TOP LIST

November 30 2017

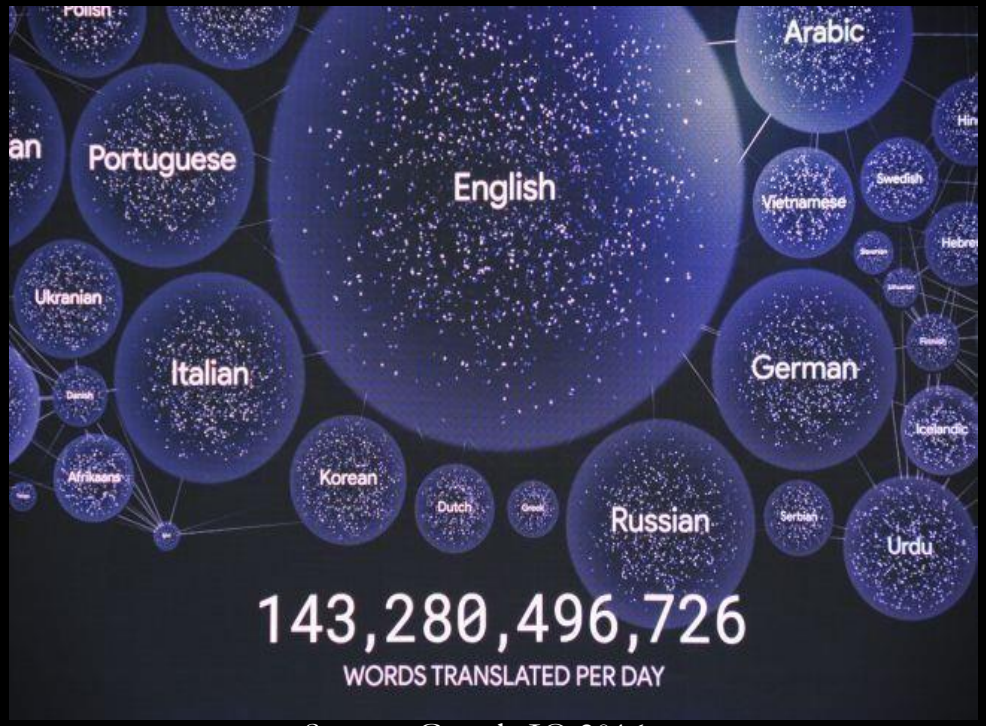
Email

The University of California San Diego Extension Center for Research on the Regional Economy released its annual list of the top 10 emerging careers for college grads in the US. Interpreters and translators topped the list.

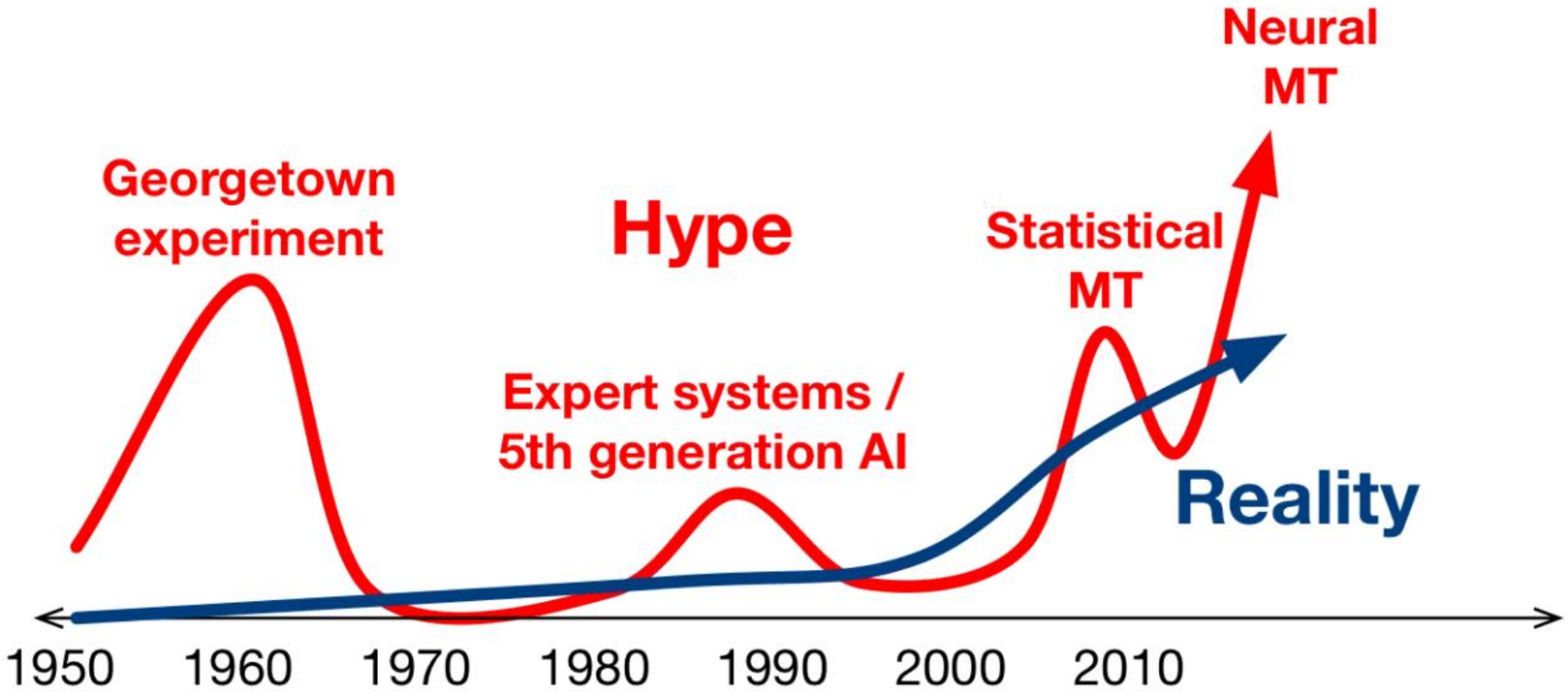
The list aimed to identify the top 10 occupations by combining the highest projected growth rates and the most online job postings using information from labor-market data firm Emsi. Researchers focused on jobs that required a bachelor's degree with less than five years of work experience.

The top 10 emerging careers in the US were:

1. Interpreters and translators



Source: Google IO 2016


Neural
MT
Georgetown
experiment
Hype
Statistical
MT
Expert systems /
5th generation AI
Reality
1950
1960
1970
1980
1990
2000
2010

Koehn (2017)

Electric Translator Is Latest 'Miracle'

The age of miracles is only just dawning, it appeared today, with the introduction of an electronic "brain" which can translate from one language to another.

The awe-inspiring machine was developed by Georgetown University language experts and the International Business Machine Corp. It was demonstrated yesterday at the I. B. M. offices, 590 Madison Ave., Manhattan.

Fifty observers watched the demonstration of what is believed to be the first mechanical translator ever developed. They saw Russian sentences fed into the machine and their English translation emerge almost simultaneously.

Mrs. Lynn Polle, 25, sat at a keyboard of the machine and began typing.

"Mi pyeryedayem mislyi posryedstvom ryechiyi," she punched.

"We transmit thoughts by means of speech," the brain know a word of Russian herself, then typed out:

"Vyelyichyina ugla opryedyelyayetsya otnoshyenyiyem dlyini dugi k radyiusu."

Colored lights blinked rapidly over the "face" of the electronic "brain" and it responded: "Magnitude of angle is determined by the relations of length of arc to radius."

The translator consists of 11 complicated electronic units which look on the outside as simple as kitchen ranges. To teach it Russian, linguists "fed" it Russian words along with their English equivalents and these were stored on a magnetic drum.

The "brain" also was "taught" rules of grammar so it could change the order of words when necessary to make a smooth intelligent English sentence.

Prof. Leon Dostert and Dr. Paul Garvin, who directed the 18-month job of creating the translator, said it has a vocabu-

Google Translate update makes it pretty much as good as a human translator

By Emma Boyle September 28, 2016 Computing

Getting chatty around the globe just got easier

GOOGLE | WEB | APPS

Google's AI translation system is approaching human-level accuracy

But there's still significant work to be done

by Nick Statt | @nickstatt | Sep 27, 2016, 2:07pm EDT

SHARE TWEET LINKEDIN

Newest GNMT: time to witness the miracle of Google Translate

TAUS Blog

Jaap van der Meer Wednesday, 24 Februa

The Future Does Not Need Translators*

slator

Language Industry Intelligence

England's Top Judge Predicts 'the End of Interpreters'

by Esther Bond on June 11, 2018

Microsoft Research
@MSFTResearch

Follow

In a major milestone, Microsoft creates a #MachineTranslation system that translates a test set of news articles from Chinese to English as accurately as a human: msft.social/W2eoUD

Microsoft researchers match human levels in translating news from Chinese to English

Microsoft researchers are making progress in translating text using deep neural networks AI training techniques.

By Mary Jo Foley for All About Microsoft | March 14, 2018 -- 15:06 GMT (15:06 GMT) | Topic: Artificial Intelligence

- Increasingly complex workflows aiming to maximise leverage and quality
- Use of raw or PE MT
- MT incorporated in batch translation, suggestion below fuzzy threshold, sub-segment matching, interactive/adaptive MT.

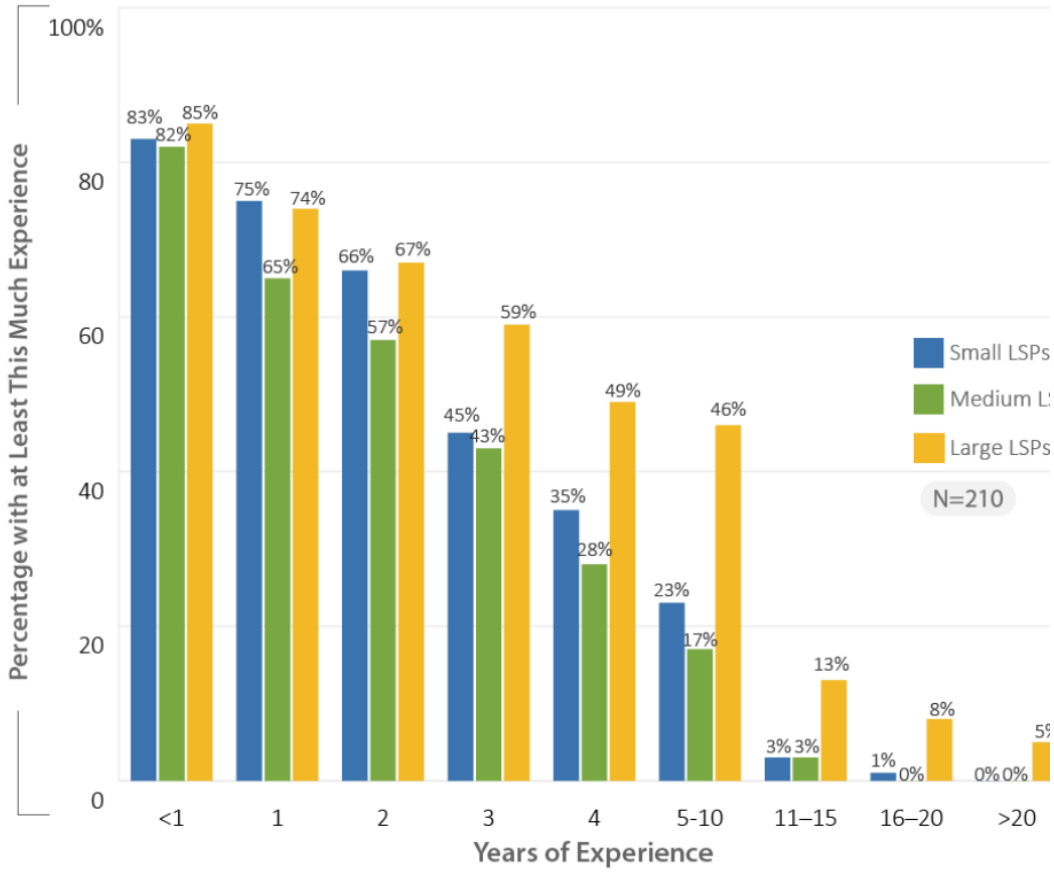
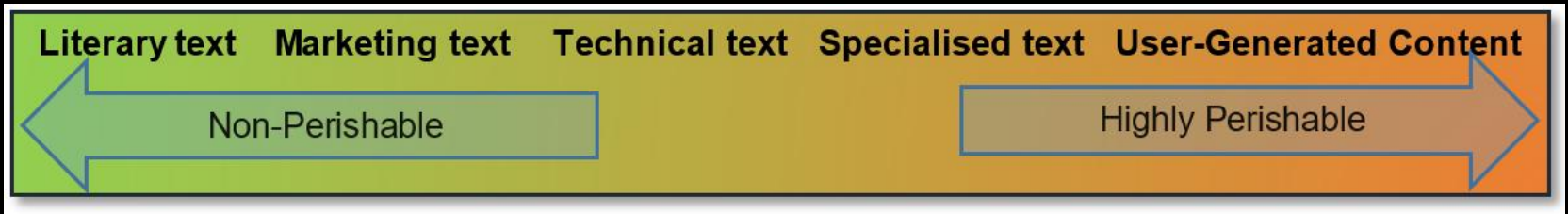
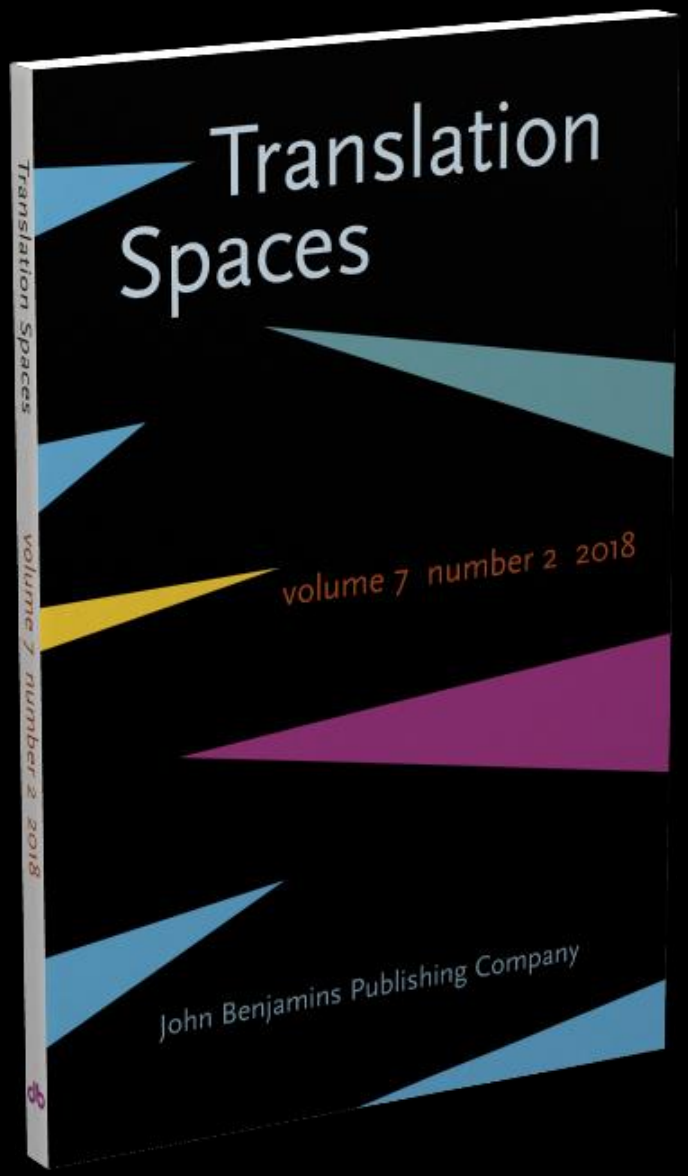


Figure 1: Large LSPs Lead in PEMT Experience
Source: Common Sense Advisory, Inc.

Use cases for MT



- Way (2013) "the degree of human involvement required – or warranted – in a particular translation scenario will depend on the purpose, value and shelf-life of the content."
- Pressure on cost and MT hype is pushing MT into action in use cases that were not previously considered suitable
- **Raw & post-edited MT for Microsoft online support and documentation (Schmidtke 2016)**
- **Raw & post-edited MT for Microsoft user interface translation (Guerberof 2018/19)**
- **Neural post-editing**
- **MT for subtitling (SUMAT 2011-2013, Ortiz-Boix & Matamala 2017)**
- **Patent translation (Iconic Translation)**
- **MT for Literary Texts (Toral, Wieling, Castilho, Moorkens and Way 2018)**



NMT for literature

Why?

- Training data available from ebooks
- Measure 'AI' translation on difficult task

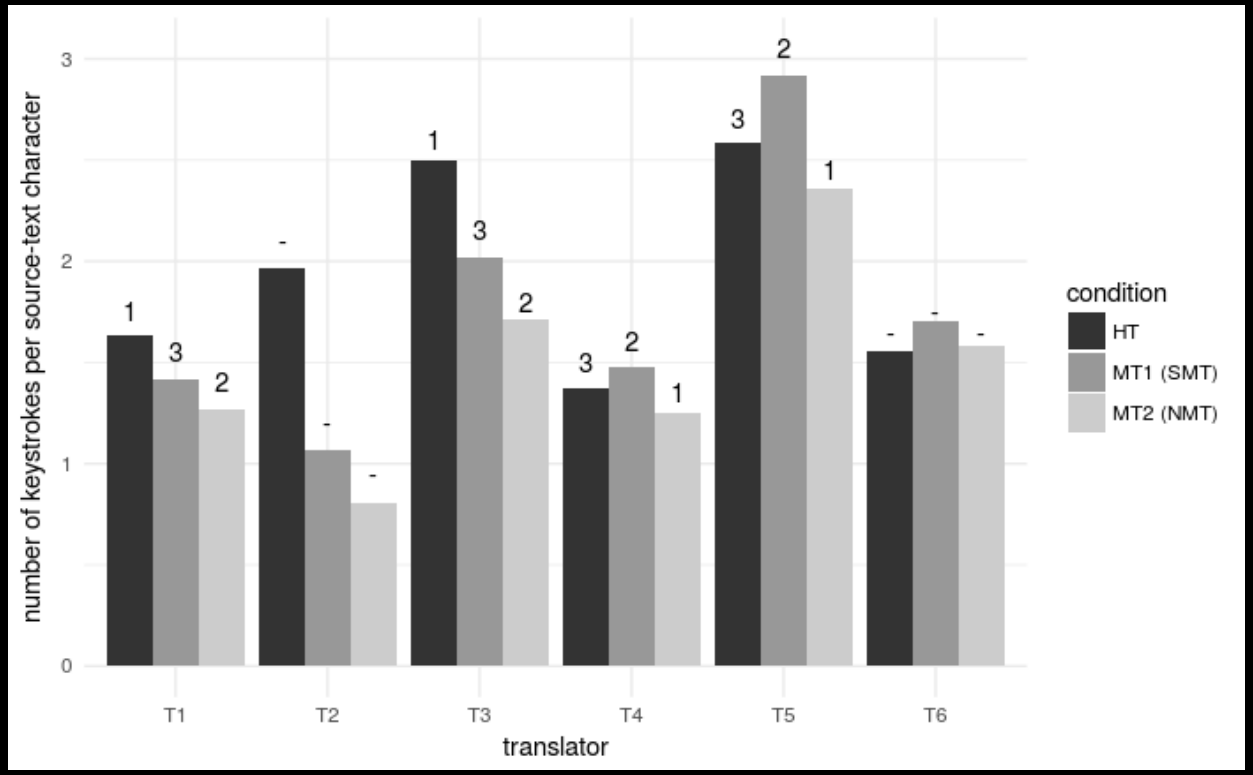
Literary translation under three conditions:

- Translation from scratch
- SMT post-editing
- NMT post-editing

Translators' Perceptions of Literary Post-editing using Statistical and Neural Machine Translation (2018)

- 6 translators, avg 10.5 years literary translation experience
- Mostly not well disposed towards MT
- 330 sentences split into 33 tasks: from scratch, SMTPE, NMTPE
- Retain MT output where possible
- Pre- and post-task questionnaires and interviews

NMT for literature



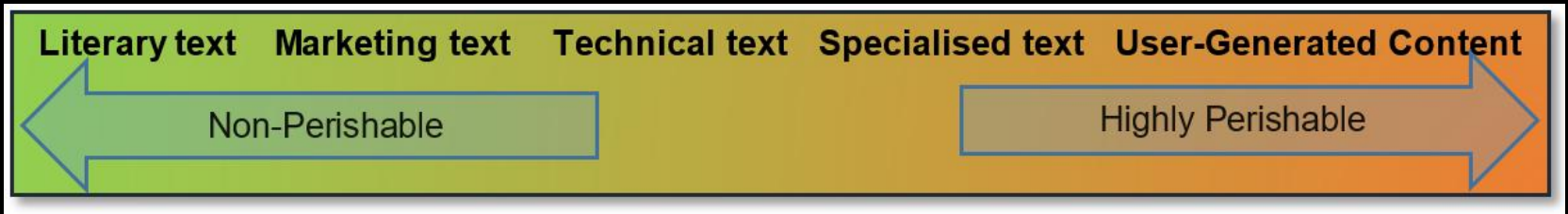
"MT conditions you"

"PE is like pre-cooked food that always tastes the same."

"It gives you a draft translation [that] you just need to fix and polish."

"A translator has a global view of the text, MT has a fragmented view."

Translation Continuums



- Other continuums:
  Low to high risk (Canfora & Ottman 2018)

Risks (Canfora & Ottmann 2018)

- Communication impaired or impossible
- Loss of reputation
- Financial damage
- Legal consequences
- Damage to property
- Injury or death

Emmanuel Macron
@EmmanuelMacron

Mes chers compatriotes, vous avez choisi de m'accorder votre confiance et je tiens à vous exprimer ma profonde gratitude.

Translated from French by bing

My fellow Americans, you chose to give me your trust, and I would like to express my deep gratitude.

07/05/2017, 22:09

7,215 RETWEETS **16.5K** LIKES

@evanmassenhove

The Guardian
International edition

Facebook

Facebook translates 'good morning' into 'attack them', leading to arrest

Palestinian man questioned by Israeli police after embarrassing mistranslation of caption under photo of him leaning against bulldozer

Alex Hern
@alexhern
Tue 24 Oct 2017 12.24 BST

▲ Facebook's machine translation mix-up sees man questioned over innocuous post confused with attack threat. Photograph: Thibault Camus/AP

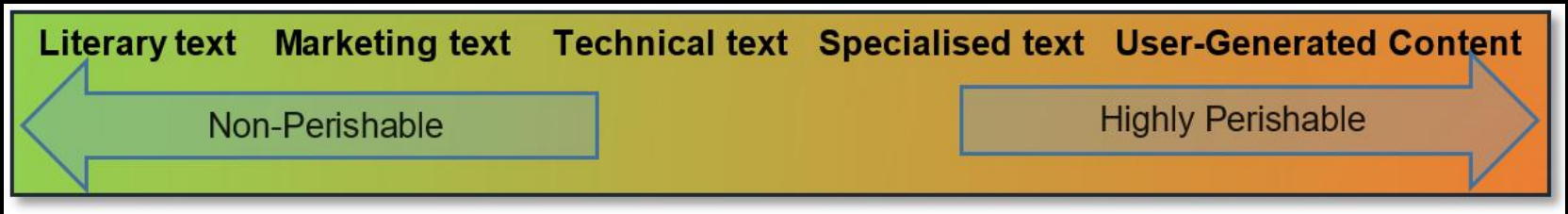
Risks

Communication impaired or impossible:

"producers and consumers… share the overall objective of providing sustainable access to reliable multilingual information" (Moorkens & Lewis 2019)

- Lack of translators
- Translation data anticommons

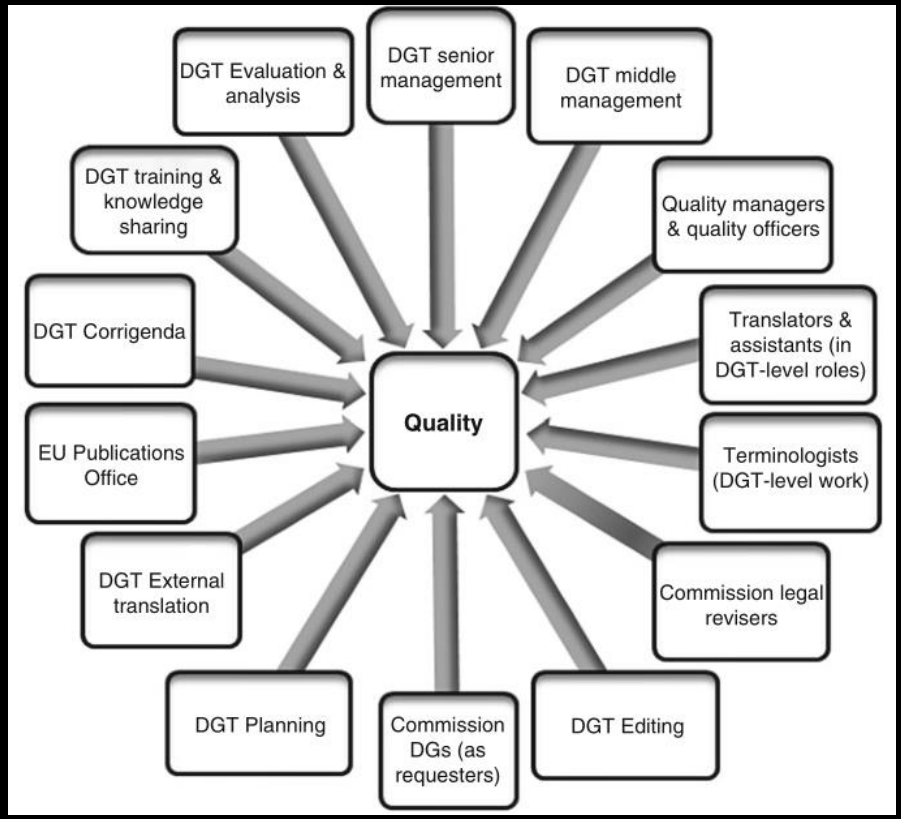
Translation Continuums



- Other continuums:
  'Bulk' to 'premium' market (Jemielity 2018)
  - Value
  - Quality expectations?

Quality

- ASTM definition: "A quality translation demonstrates required accuracy and fluency for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs"



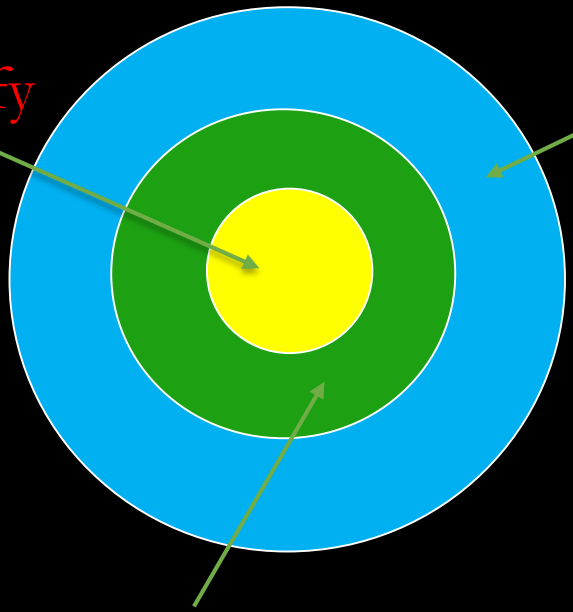
Quality management at DGT level: potential contributors (Drugan, Strandvik, Vuorinen 2018)

“Translation is done not only by the brain, but also by complex systems, systems which include people, their specific social and physical environments and all their cultural artefacts.” (Risku 2002)

3D Quality applicable to process (Abdallah)

PRODUCT QUALITY

Does the product satisfy the stated needs



SOCIAL QUALITY

- **Interaction** and **relations** of the actors **(human and non-human)** in the production network.
- Affects process quality which affects product quality.

Who/what does, under what sort of circumstances

PROCESS QUALITY

How is the work done

What is the source text like

(Abdallah 2007, 2012, 2014, see also Jääskeläinen 2016)

Post-Editing: A Contentious Topic

Does post-editing within a CAT tool spark joy?

- Survey found dissatisfaction with existing interfaces (Moorkens & O'Brien 2017)
- Making usable interfaces – and usable MT – is difficult to specify
- Computers struggle with poorly-defined tasks
- ROI for translator usability difficult to predict

Translator sustainability

2016 Survey:

The nature of translations is changing/evolving and so too are documents and CAT Tools…

There will be a bright future in Ireland for Irish translation.

I have had to increase my turnaround continuously since 2010 and it's showing no signs of decreasing.

Translator sustainability

2016 Survey: Isolation at work: 2.71

- “I've been actively moving out of the career for some time now”.
- “Isolation of self-employed working from home [was] literally killing me, as was RSI and stress related to tight deadlines and the cut throat market.”

Translator sustainability

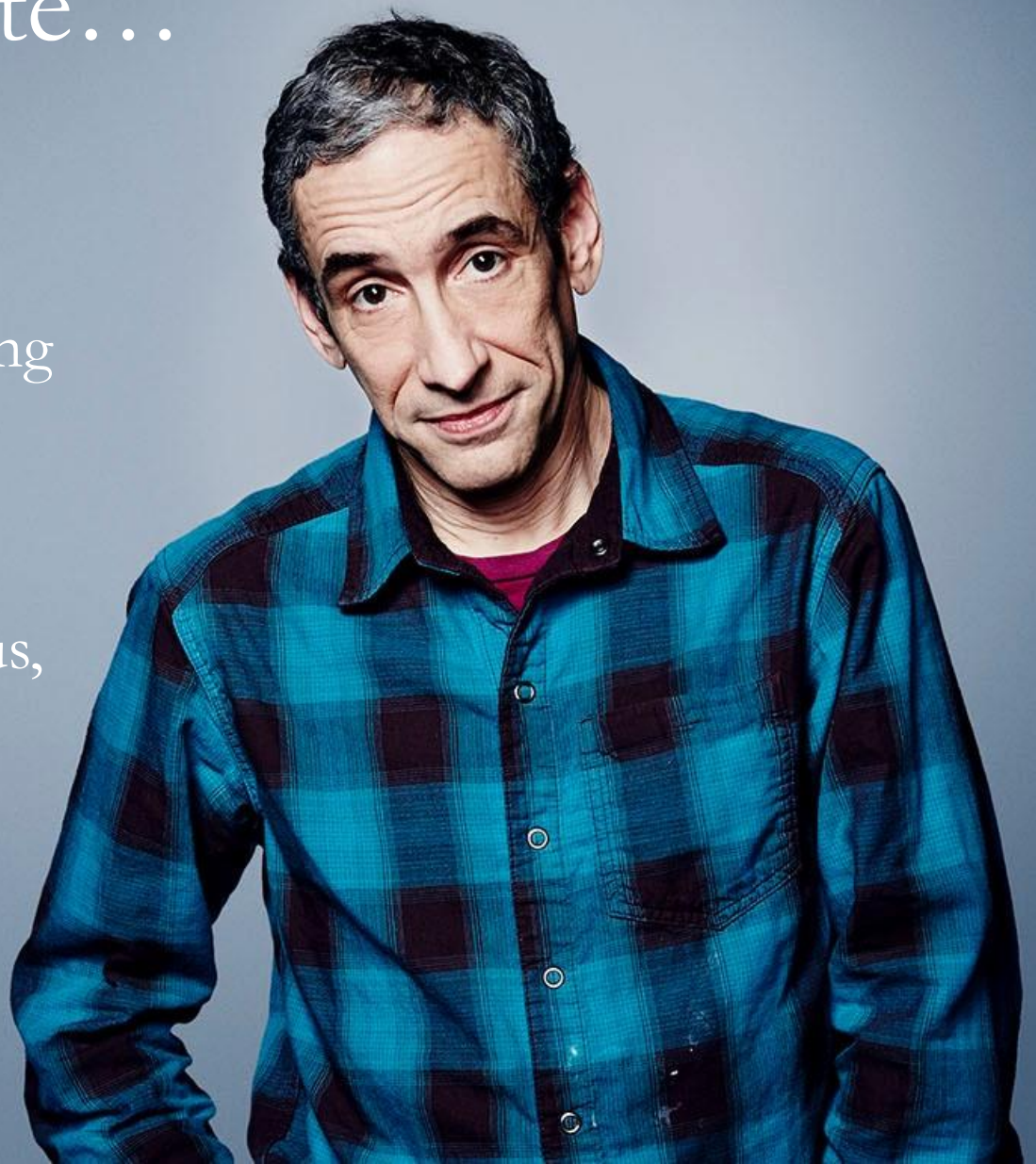
What do you consider the main threat to the profession of translation?

- "Race to the bottom in terms of unit rate and quality"
- "agencies aggressively lowering rates and winning huge contracts so they can corner the market and pressure translators into accepting ridiculous rates"
- "Increasing improvements in MT will eventually reduce the role of the translator to that of editor/proof-reader"
- "I don't see much of a future for human translators"

Removing waste…

… means cost-saving by automating or rationalising workflow steps

- "We use our social connections to orient us, to derive meaning and purpose" (Rushkoff 2019)



What does sustainable or ethical MT look like?

- Value – longer-term thinking
- Trust – relationship building or distributed ledgers?
- Professionalisation – defined, with a career path?
- Copyright – is there a balanced approach? Translation commons?
- Data ownership – personal data and translation data

… or is the current trajectory perfectly sustainable?

Can small wins become big wins?

- Council of Ministers in EU agree to collective bargaining by freelance workers
- Collective bargaining agreements by Medicaid translators, Italian literary translators
- Agreements for royalty payments when work is rebroadcast by Dutch audiovisual translators
- Broader conversation about ethics and machine learning

New roles in language services industry

Advisory – workflows, use of automation, training data
(Moorkens 2017)

Experts in multilingual and multimedia communication
(Massey 2018)

Online social media translation
(Desjardins 2017)

Email: joss.moorkens@dcu.ie

References

- Abdallah, K. 2010 Translator's agency in production networks. In T. Kinnunen & K. Koskinen (Eds.), Translator's agency (pp. 11–46). Tampere: Tampere University Press.
- Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Way, A., Georgakopolou, P. 2018. Evaluating MT for Massive Open Online Courses: A multifaceted comparison between PBSMT and NMT systems. Machine Translation 32:3, p255-278. doi: 10.1007/s10590-018-9221-y
- Koehn, P., Knowles, R. 2017 Six Challenges for neural machine translation. In: Proceedings of the 1st workshop on neural machine translation, Vancouver, BC, Canada, pp 28–39
- Moorkens, J., Lewis, D. 2019 (accepted). Research Questions and a Proposal for the Future Governance of Translation Data. Journal of Specialised Translation 32.
- Moorkens, J. 2019 (accepted). Comparative satisfaction among freelance and directly-employed Irish-language translators. Translation & Interpreting 11:2.
- Moorkens, J., Toral, A., Castilho, S., Way, A. 2018. Perceptions of Literary Post-editing using Statistical and Neural Machine Translation. Translation Spaces 7:2, p240-262.
- Moorkens, J., O'Brien, S. 2017. Assessing User Interface Needs of Post-Editors of Machine Translation. IN: Kenny, D. (Ed.), Human Issues in Translation Technology: The IATIS Yearbook. Oxford, UK: Routledge, p109-130.
- Moorkens, J. 2017. Under pressure: Translation in times of austerity. Perspectives: Studies in Translation Theory and Practice, 25(3), p464-477, doi: 10.1080/0907676X.2017.1285331
- Neubig, G., Morishita, M., Nakamura, S. 2015 Neural reranking improves subjective quality of machine translation: NAIST at WAT2015. arXiv preprint arXiv:1510.05203
- Risku, H. 2002 Situatedness in translation studies. Cognitive Systems Research 3, 523-533.
- Sanchez-Gijon, P., Moorkens, J., Way, A. 2019 (accepted). Post-Editing Neural Machine Translation vs. Translation Memory Segments. Machine Translation.
- Wu Y. et al. 2016 Google's neural machine translation system: bridging the gap between human and machine translation. arXiv:1609.08144