

# Is Neural Machine Translation the New State of the Art?

Sheila Castilho

[sheila.castilho@adaptcentre.ie](mailto:sheila.castilho@adaptcentre.ie)

Joss Moorkens  
Iacer Calixto  
Andy Way

Federico Gaspari  
John Tinsley

- MT and the hype
- Use cases
  - NMT for E-Commerce
  - NMT for Patents
  - NMT for MOOC
- Conclusion



- Great excitement and anticipation each new wave of MT
- NMT:
  - “bridging the gap between **human** and MT”



Technology

## “Nearly Indistinguishable Translation”—Google

by [Florian Faes](#) on September 27, 2016



## Google Translate update n

### much a

By Emma Boyle

Getting ch



GOOGLE WEB APPS

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

*But there's still significant work to be done*

by [N](#) **Intelligent Machines**



## Google's New Service Translates Languages Almost as Well as Humans Can

A jump in the fluency of Google's language software will help efforts to make chatbots less lame.

Home > Sci-Tech

**A major breakthrough in the field of Translation -- Google translation integration neural network, the translation quality is close to the manual translation**

© 2016-09-28 18:59:03    90 °C

# miracle of Google Translate

Published on October 4, 2016



## Is Neural MT Really as Good as Human Translation?

[www.common senseadvisory.com/default.aspx?Contenttype=ArticleDetAD...](http://www.common senseadvisory.com/default.aspx?Contenttype=ArticleDetAD...) ▼

Jan 18, 2017 - By now, most language professionals have seen that claims that neural machine translation (NMT) is delivering results as good – or almost as ...

## Can Neural Machine Translation Compete with Human Translation ...

<https://www.motionpoint.com/.../can-neural-machine-translation-compete-with-huma...> ▼

Feb 8, 2017 - MotionPoint has long advocated using human translators to craft this resonant online content. (And we still do; more on this near the end of this ...)

## Will Neural Machine Translation Replace Human Translation ...

[contentech.com](http://contentech.com) › blog ▼

Mar 26, 2017 - Will Neural Machine Translation Replace Human Translation? ... likelihood, NMT will be free or very inexpensive in the near future as ...



## Will the interpreter, translator job die out due to improving machine ...

<https://www.quora.com/Will-the-interpreter-translator-job-die-out-due-to-improving-ma...>

Apr 16, 2016 - Yes, definitely. And I think this will be tied to a computer passing the Turing test. The moment a computer gets to the point of understanding normal ...

## Tech is removing language barriers – but will jobs be lost in translation ...

<https://www.theguardian.com> › Education › Languages ▼

Sep 19, 2014 - When a translator's email reply landed in their inbox in 2008, the Welsh ... Experts divide translation technology in two distinct categories: machine translation (MT) ... actually is – that good translators will never be out of a job.

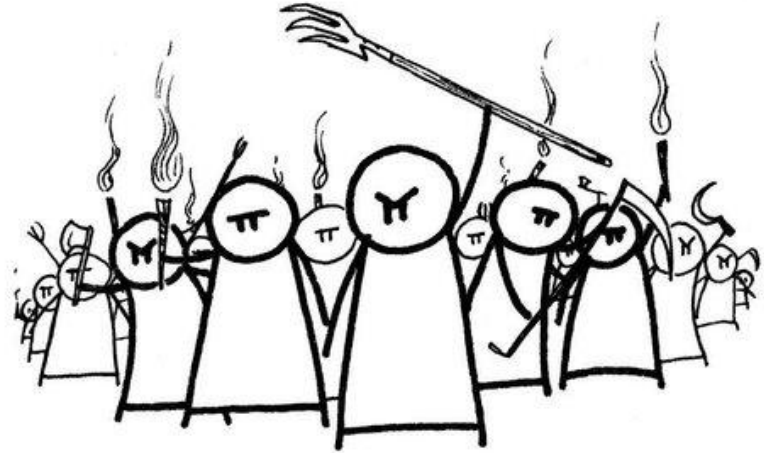
## What Does The Future Hold For Translators? | Coaching For Translators

<https://coachingfortranslators.com/2015/05/.../what-does-the-future-hold-for-translato...> ▼

May 10, 2015 - Although I'm not a fan of automatic translation, I was surprised by some ... and how things didn't turn out that way – a familiar experience for many of us. ... I've seen technology evolve over my 15-year career as a translator, ...

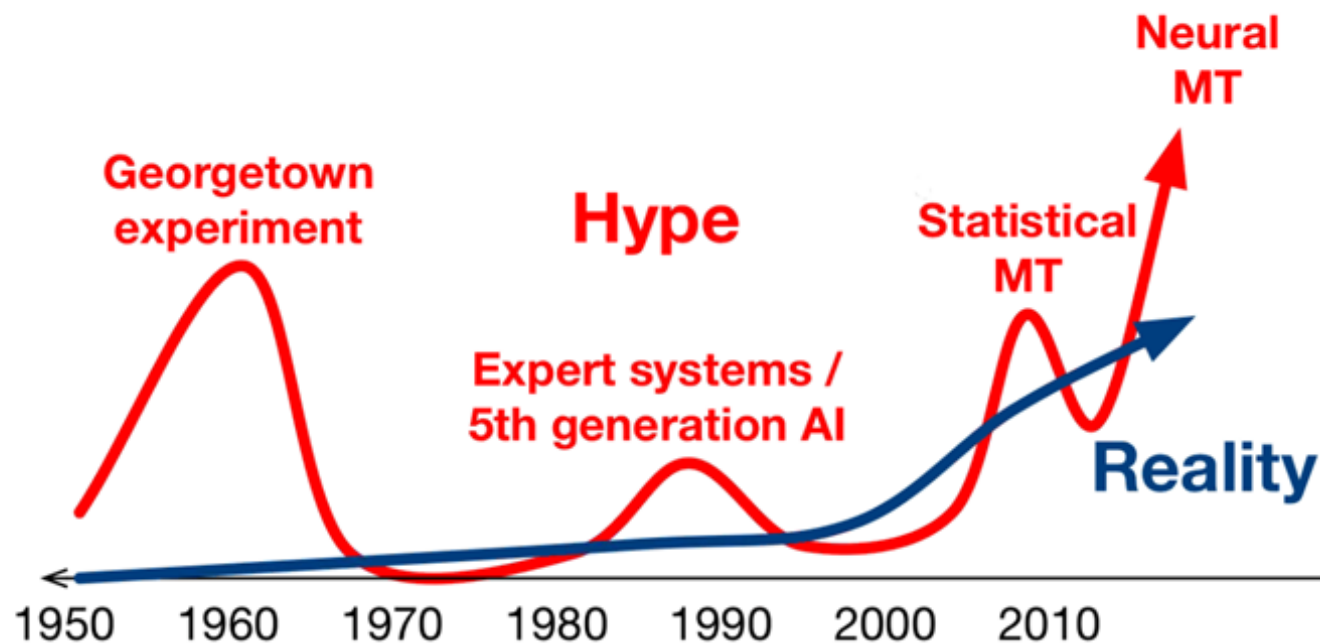


- Us vs them



- “MT will steal translators' jobs”
- “translators will be merely post-editors of MT”
- “MT is a threat”

## Hype and Reality



(Philipp Koehn, Omniscien Webinar 2017)

# But is NMT really that good?

- Use cases
- different domains
- different set of language pairs







- Systems (Calixto et al. 2017):
  - (1) a PBSMT baseline model built with the Moses SMT Toolkit
  - (2) a text-only NMTt model
  - (3) a multi-modal NMT model (NMTm)
- English into German
- Data set: 24k parallel product titles + images
- Validation/test data: 480/444 tuples
- 18 German native speakers
- Ranking
  - Translations from the 3 systems + product image
- Adequacy (Likert scale 1- All of it to 4- None of it)
  - Source + translation + product image





Model	BLEU4 $\uparrow$	METEOR $\uparrow$	TER $\downarrow$	chrF3 $\uparrow$	Adequacy $\downarrow$
NMT <sub>t</sub>	22.5	40.0	58.0	56.7	2.71 $\pm$ .48
NMT <sub>m</sub>	25.1 $\uparrow$	42.6 $\uparrow$	55.5 $\uparrow$	58.6	2.36 $\pm$ .47
PBSMT	27.4 $\uparrow\ddagger$	45.8 $\uparrow\ddagger$	55.4 $\uparrow$	61.6	2.36 $\pm$ .47

- AEM:
  - PBSMT outperforms both NMT models (BLEU, METEOR and chrF3)
  - NMT<sub>m</sub> performs as well as PBSMT (TER)
- Adequacy
  - NMT<sub>m</sub> performs as well as PBSMT
- Ranking
  - PBSMT: 56.3% preferred system
  - NMT<sub>m</sub>: 24.8%
  - NMT<sub>t</sub>: 18.8%



- Compare the performance between the mature patent MT engines used in production with novel NMT
- Systems
  - PBSMT (a combination of elements of phrase-based, syntactic, and rule-driven MT, along with automatic post-editing)
  - NMT (baseline)
- English into Chinese
- Data set: ~1M sentence pairs chemical abstracts, ~350K chemical titles, ~12M general patent, and ~2K glossaries.
- 2 reviewers
- Ranking
- Error analysis
  - Punctuation, part of speech, omission, addition, wrong terminology, literal translation, and word form.



System	Titles (BLEU)	Abstracts (BLEU)
Iconic MT	31.99	28.32
Neural MT	37.52	13.39



- AEM:
  - SMT outperforms NMT for abstracts, NMT outperforms SMT for titles
- Ranking
  - General: PBSMT 54% - NMT 39%
  - Long sentences: PBSMT 58% - NMT 33%
  - Short sentences: PBSMT 84% - NMT 8%
  - Medium-length sentences: PBSMT 36% - NMT 57%





- Error analysis
  - SMT: sentence structure 35% (10% NMT)
  - NMT: 37% omission (8% SMT)
  
- % segments with “no errors”
  - SMT 25%
  - NMT 2%



- Decide which system would provide better quality translations for the project domain
- Systems
  - PBMST (Moses)
  - NMT (baseline)
- English into German, Greek, Portuguese and Russian
- Data set:
  - OFD : ~24M (DE), ~31M (EL), ~32 (PT), ~22(RU)
  - In-domain : ~270K(DE), ~140K(EL), ~58K(PT), ~2M(RU)
- Ranking
- Post-editing
- Fluency and Adequacy (1-4 Likert scale)
- Error analysis: inflectional morphology, word order, omission, addition, and mistranslation



Lang.	System	BLEU	METEOR	HTER	Fluency	Adequacy
DE	SMT	41.5	33.6	49.0	2.60	2.85
	NMT	61.2 †	42.7 †	32.2	2.95	2.79
EL	SMT	47.0	35.8	45.1	2.86	3.44
	NMT	56.6 †	40.1 †	38.0	3.08	3.46
PT	SMT	57.0	41.6	33.4	3.15	3.73
	NMT	59.9	43.4	31.6	3.22	3.79
RU	SMT	41.9	33.7	44.6	2.70	2.98
	NMT	57.3 †	40.65 †	33.9	3.08	3.12



- AEM:
  - NMT outperforms SMT in terms of BLEU and METEOR
  - More PE for SMT
- Fluency and Adequacy
  - NMT is preferred across all languages for Fluency
  - Adequacy results a bit less consistent





Lang.	System	Technical Effort	Temporal Effort	WPS
DE	SMT	5.8	74.8	0.21
	NMT	3.9	72.8	0.22
EL	SMT	13.9	77.7	0.22
	NMT	12.5	70.4	0.24
PT	SMT	3.8	57.7	0.29
	NMT	3.6	55.19	0.30
RU	SMT	7.5	104.6	0.14
	NMT	7.2	105.6	0.14

- Post-editing
  - Technical effort improved for DE, but marginally for other languages
  - Temporal effort marginally improved
- Ranking
  - NMT is preferred across all languages (DE 80%, EL 56%, PT 61% and RU 63%)





NMT results are really promising!

But...

human evaluations show that results are not yet so clear-cut



- Translation industry is eager for improved MT quality in order to minimise costs
- The hype around NMT must be treated cautiously
- Overselling a technology that is still in need of more research may cause negativity about MT
  - “us vs them”
  - “MT is a threat to human translators”



# Thank you!

# Questions?

Sheila Castilho

[sheila.castilho@adaptcentre.ie](mailto:sheila.castilho@adaptcentre.ie)

