

# Mensch oder Maschine: Möglichkeiten und Grenzen des technologiegestützten Übersetzens

Vesna Lušicky, Zentrum für Translationswissenschaft, Universität Wien



# Mensch oder Maschine...?

„In fünf Jahren brauchen wir keine ÜbersetzerInnen mehr“

...und das seit 80 Jahren!

# Übersetzen mit Technologie?

## 1. Hilfsmittel für ÜbersetzerInnen

- Technologiestütztes Übersetzen ist eine Form der Übersetzung, bei der ein menschlicher Übersetzer Texte mit Hilfe von Computerprogrammen übersetzt.

## 2. Maschinelle Übersetzung (MÜ)

- *Human-Aided Machine Translation*
  - **Pre-Editing** des Ausgangstextes um die Qualität des MÜ-Ergebnis zu optimieren.
  - **Post-Editing** des MÜ-Ergebnisses um die Fehler zu beheben.
- Vollautomatisch, *Fully Automatic Translation*

# Was macht Übersetzen für Maschine schwierig?

- Beispiel:

„Die Volkswirtschaftslehre ist ein Teilgebiet der Wirtschaftswissenschaften.“

„The economics is a sub-area of economics.“

- Ambiguität, Kontext, Idiome, Terminologie, Domänenwissen...

# Maschinelle Übersetzung und künstliche Intelligenz

- Neuronale maschinelle Übersetzung:
  - Basiert auf sogenannten künstlichen neuronalen Netzen, die mit großen Mengen von Daten trainiert werden und so lernen sollen, eigenständig neue Übersetzungen anzufertigen.

## Künstliche Intelligenz für österreichisches Deutsch

**Digital.** Für den Ratsvorsitz Österreichs erlernte das maschinelle Übersetzungssystem der EU erstmals österreichische Varianten der Verwaltungssprache. Für wirklich gute Übersetzungen braucht es aber weiterhin den Menschen.

VON ERIKA PICHLER

Spezifika der österreichischen Sprache existieren nicht nur auf Speisekarten oder Straßenschildern, sondern auch in der Verwaltungssprache oder in Gesetzestexten. Wörterbücher und elektronische Übersetzungsprogramme, die in der Regel mit Daten aus Deutschland gefüllt werden, stoßen hier oft an Grenzen. Auch das maschinelle Übersetzungssystem der Europäischen Kommission, eTranslation, verfügte bisher kaum über Austria.

Um vor allem für die Zeit des österreichischen Ratsvorsitzes Abhilfe zu schaffen, hat das Zentrum für Translationswissenschaft (ZTW) der Uni Wien ein maschinelles Übersetzungssystem weiterentwickelt: den EU Council Presidency Translator. Ihm wurden gezielt österreichische Textbausteine, Dokumente und Begrifflichkeiten zugeführt. Das Übersetzungssystem vermag dadurch Spezifika der österreichischen Verwaltungssprache in allen EU-Sprachen

ständig zu machen, dies mit Fokus auf Schwerpunktthemen der Ratspräsidentschaft wie Digitalisierung oder Migration. Es ist frei zugänglich, profitieren sollen vor allem Delegierte, Organisatoren und Journalisten, die rasch an Informationen kommen müssen.

### System lernt selbst dazu

Der EU Council Presidency Translator wurde vom ZTW bereits für den estnischen Ratsvorsitz 2017 entwickelt, und zwar in einem EU-Projekt, das zusammen mit dem baltischen Sprachtechnologie-Unternehmen Tilde betrieben wurde. Damals basierte das System auf statistischen Verfahren, seit dem österreichischen EU-Vorsitz erstmals auf künstlicher Intelligenz: Es werden selbst aus künstlichen geschaffenen Sätzen erlernt, wie man sie

Verbindung setzen“, sagt Vesna Lušicky, Postdoc-Forscherin und Teamleiterin am ZTW.

Die Technologie des neuronalen maschinellen Übersetzens wird erst seit 2014 in einem nennenswerten Umfang erforscht. „Der Qualitätssprung ist im Vergleich zum statistischen Verfahren bemerkenswert. Vor allem im Bezug auf die Flüssigkeit und grammatische Richtigkeit wurden große Fortschritte gemacht“, sagt Lušicky. Die Qualität der Rohübersetzung könne jedoch in Abhängigkeit von der Art und Qualität der Ausgangstexte stark variieren.

In der Tat ist es in den Presstexten

Systeme immer bessere Ergebnisse in diesem Bereich liefern, werden wir vergänglich auf entsprechende Ergebnisse bei stilistisch ausgefeilten und kulturgebundenen Texten warten, wie etwa bei Werbe- oder juristischen Texten, für die Fachwissen und Rechtssicherheit notwendig sind“, sagt Lušicky. Benötige man qualitativ hochwertige oder druckreife Übersetzungen, würden daher auch weiterhin professionelle Übersetzer unerlässlich bleiben.

### LEXIKON

Neuronale maschinelle Übersetzung basiert auf künstlichen neuronalen Netzen, die mit Daten trainiert werden und so lernen, eigenständig neue Übersetzungen anzufertigen. Die Netze sind der neuronalen Vernetzung des menschlichen Gehirns nachempfunden. Ein Neuron nimmt von anderen Neuronen Informationen in Form von Eingangssignalen auf, verarbeitet diese und leitet sie an andere Neuronen weiter.

Quelle: www.translate2018.eu

**The Bible helps AI gain in translation**  
Researchers at the Dartmouth College have turned to the Bible to improve their artificial intelligence and Machine Learning

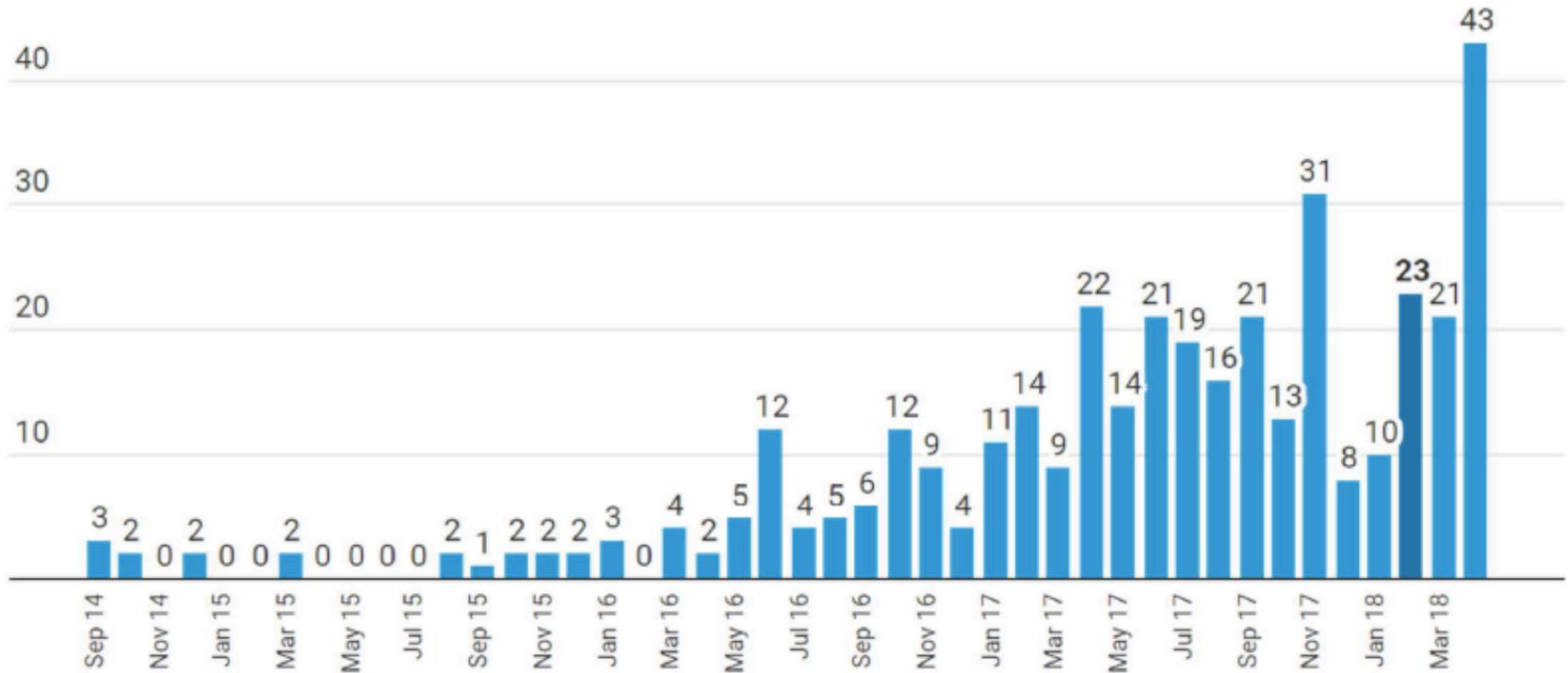
**China's Baidu challenges Google with A.I. that translates languages in real-time**  
Baidu unveiled an artificial intelligence-powered tool that can translate languages in real time on Wednesday.  
The simultaneous translation feature has been trained on two million pairs of English and Chinese sentences.  
The Chinese tech giant is posing a challenge to Google which last year unveiled its real-time translation tool.

Bildquellen: Die Presse, 1.9.2018; livemint.com, 24.10.2018; cnbc.com, 28.10.2018 (abgerufen am 10.11.2018)

# Maschinelle Übersetzung und künstliche Intelligenz

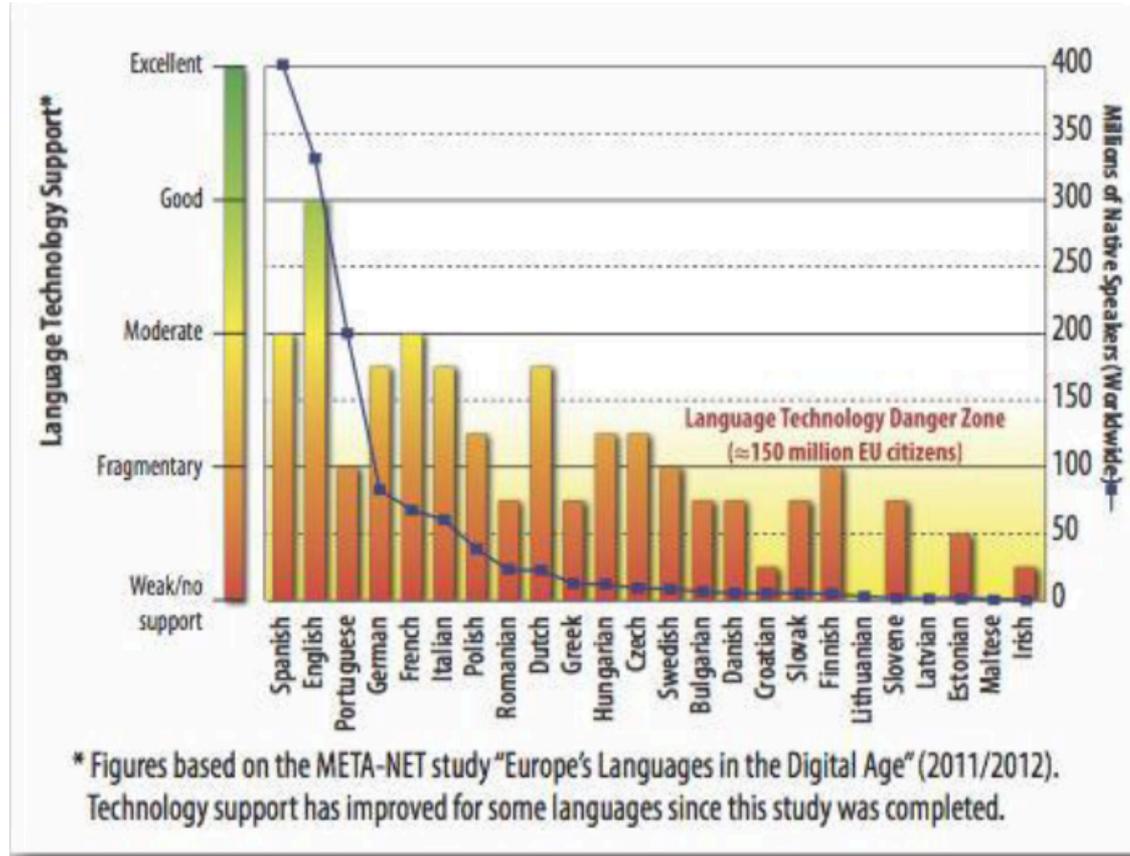
- Neuronale maschinelle Übersetzung:
  - Texte sind sehr flüssig
    - *Multidimensional Quality Matrix* (Lommel et al. 2015): *Fluency* (Bojar et al. 2016; Toral and Sánchez-Cartagena 2017)
  - Sehr gute Ergebnisse: Morphologie, Wortfolge (Toral und Sánchez-Cartagena, 2017)
  - Post-editing: Geringer Aufwand (Bentivogli et al. 2016)
  - Andere Ansätze schneiden bei langen Segmenten besser ab
  - Fehlerkategorien: Auslassungen, Ambiguitäten, Terminologie, Namen

# Neuronale Maschinelle Übersetzung: Forschung



Bildquelle: Slator, <https://slator.com/academia/neural-machine-translation-research-output-ballooned-in-the-first-half-of-2018/> (abgerufen am 10.11.2018)

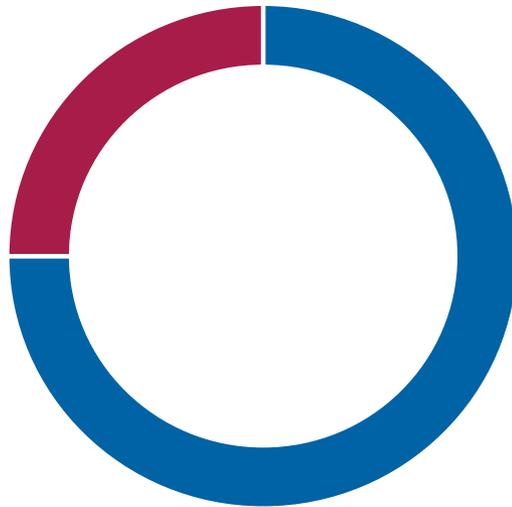
# Sprachtechnologie in Europa



Bildquelle: Rehm et al. (2017). Language Technologies for Multilingual Europe - Towards a Human Language Project. [cracker-project.eu/wp-content/uploads/SRIA-V1.0-final.pdf](http://cracker-project.eu/wp-content/uploads/SRIA-V1.0-final.pdf) (abgerufen am 26.9.2018).

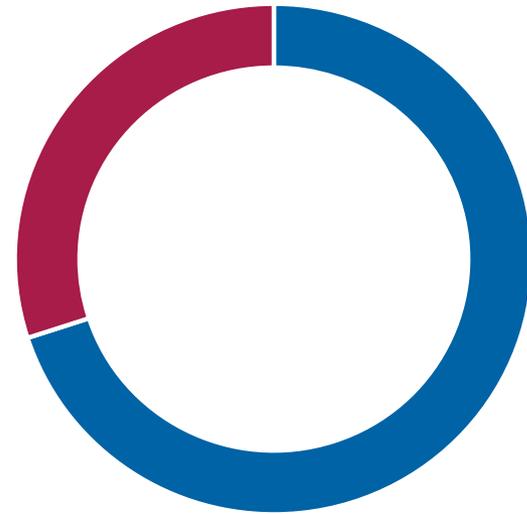
# Sprachenvielfalt im Internet

1998



■ Englisch ■ Andere

2017



■ Chinesisch, Arabisch, Spanisch, Russisch, Portugiesisch usw.  
■ Englisch

Quelle: Tilde, Präsentation „EU Council Presidency Translator – Essential Tools for a Multilingual World“, Workshop Neuronales maschinelles Übersetzen, 13.9.2018, Universität Wien.

# Anforderungen an die Qualität

- Langzeit- oder Kurzzeit-Inhalte  
(*perishability of content*)
- Angemessen (*fit for purpose*)

**J'ai un probleme avec les paiement par mois**

barbara gibertz de @recupdesign 6:55 am 03 oct. 2013 EDT

je trouver que c'est mal explique pour la vente, je ne pas envie de rester à vie sur un site ou je vende rien, il y a pas moyen de s'inscrire. je déjà arrêté mon produit sur le site et je voudrais bien me desabonner qui n'est pas possible, par contre je recou de mail pour une modique somme de 0.20 usd ,que si je paie sa me couste plus au frais par la viss, qui peut m'aide svp

**Geld zurück**

maitechr33 6:10 am 11. Nov 2015 EDT

Hallo ihr,  
ich habe vor kurzem etwas bestellt, was ich aber noch am selben tag habe stornieren lassen. Trotzdem wurde mir ein betrag in höhe von 20,57 euro abgebucht. Als ich den Verkäufer dann angeschrieben habe meinte der dass ich das über Etsy irgendwie klären muss und nicht über seinen shop. Wie mache ich das denn jetzt bloß, ich habe eine mail ans etsy support team geschickt aber die schreiben mir nicht zurück. / Kann mir einer von euch weiterhelfen und mir tipps geben wie ich jetzt vorgehen soll?  
danke schon einmal im voraus!

**problemi con le inserzioni**

chiara M. di @artgalleryChiaraM 2:04 pm 05 apr 2015 EDT

Salve, mi sono iscritta qui su Etsy qualche giorno fa, con l'intenzione di "aprire il mio negozio".  
Dopo avergli dato un nome, ho tentato svariate volte ormai di aggiungere un'inserzione.  
Questo e' l'avviso che puntualmente mi compare:  
"si è verificato un problema nel caricamento della pagina"  
Sapete aiutarmi?  
Grazie,  
Chiara.

**Help! Wel betaling gedaan, maar geen bestelling!**

Marlies Conves 2:31 pm 24 feb 2015 EDT

Gisteravond heb ik een bestelling gedaan via Etsy. Ik heb betaald via Ideal -> op het moment van terugverwijzen (nadat de betaling was afgerond) liep het proces vast (geen reactie van de server).  
De bestelling zit nu nog in mijn winkelmandje (dat is consistent met de terugkoppeling van de verkoper die aangeeft geen bestelling te hebben ontvangen) en ik kan bij Etsy geen zaak starten (want volgens mijn Etsy account heb ik geen bestelling geplaatst), maar het bedrag is al wel van mijn rekening afgeschreven. Ben al bijna een uur aan het zoeken op de Etsy site, maar het lijkt onmogelijk voor mij om contact op te nemen om het probleem te adresseren. Wie kan mij helpen???

**Problema al intentar pagar facturas con PayPal**

Raquel Serrano de @Raquel90Hendrade 1:24 pm 12 sep 2014 EDT

Hola, buenas tardes, llevo toda la tarde intentado abrir mi tienda, pero para ello necesito saber porque no me funciona. El caso es que todo el proceso de crear una tienda va bien hasta que llego a la ultima parte, cuando ya solo tengo que elegir como pagar las facturas. Puesto que no tengo tarjeta de credito decidí escoger "Autorizar cuenta de PayPal" parece que todo va bien, pero al final me dice que PayPal ha dado error.  
He preguntado a PayPal a que se puede deber y esto es lo que ellos me han dicho:  
- La conexión de Internet era lenta y la página de confirmación del pago expiró.  
- Pudo salirse por error de la página web mientras la página de confirmación del pago estaba aún en proceso. (Esto puede suceder si se pulsa el botón de volver a la página anterior en el navegador o se pulsa un enlace que lleva a otra página.)  
- La página de confirmación se cerró mientras la transacción estaba todavía en proceso.

Quelle: Tilde, Präsentation „EU Council Presidency Translator – Essential Tools for a Multilingual World“, Workshop Neuronales  
machinelles Übersetzen, 13.9.2018, Universität Wien.

# Faktoren

- Ziel und Zweck,
- Funktion, Anforderungen,
- Domäne, Textsorte, Sprachkombination,
- Qualität von Quelltexten,
- vorhandenen Sprachressourcen, personellen Ressourcen,
- Workflow,
- vorhandenen Systemen und deren Integrationsfähigkeit,
- Zeitfaktor, Kostenfaktor,
- Lernkurve,
- Know-How...

Vgl. Thicke 2013.

# Sicherheit

- Datensicherheit

## Warning about translation web site: Passwords and contracts accessible on the Internet

Sensitive information that has been translated at a popular digital translation service has been stored in the cloud and could be searchable via Google. Tekna warns Norwegian authorities, companies and private citizens against naïve use of the web site.



Bilduelle: <https://www.nrk.no/urix/warning-about-translation-web-site -passwords-and-contracts-accessible-on-the-internet-1.13670874> (abgerufen am 10.11.2018).

## eTranslation

- Maschinelle-Übersetzung-System für alle Amtssprachen
  - Domänen-spezifisch
  - Training: u.a. EURAMIS, + 1 Milliarde Segmente
  
- Juli 2017– März 2018: 3,7 Mio. Seiten übersetzt
  
- Weitere Informationen:  
<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>

Quelle: Markus Foti, Präsentation „CEF eTranslation: Multilingualism for Europe“, EuropeanaTech Conference, 15.5.2018.

---



## EU Council Presidency Translator

- Der EU Council Presidency Translator ist ein maschinelles Übersetzungssystem speziell für den Vorsitz im Rat der Europäischen Union.
  - Estland
  - Bulgarien
  - Österreich
- Der EU Council Presidency Translator ist ein sicheres maschinelles Übersetzungssystem für alle offiziellen EU-Sprachen.



## Maschinelle Übersetzung

- Neuronale maschinelle Übersetzung
- eTranslation, das maschinelle Übersetzungsservice der Europäischen Kommission

+

künstliche Intelligenz



## Training

- EN – DE – EN
  - Domänenadaptation: Nachrichten (Österreich), EU-Ratsvorsitz
    - 39.1 K parallele Segmentenpaare aus domänenspezifischen Translation Memories
    - 5.3 M parallele, domänenspezifische Segmente
    - 2.1 M einsprachiges, domänenspezifisches Korpus
    - 5 Terminologiebestände

# Translate2018.eu



The screenshot shows the 'EU Council Presidency Translator' interface. At the top left, the text 'e 2 0 u 1 8 a t' is displayed vertically, with 'Austrian Presidency of the Council of the European Union' to its right. The main title 'EU Council Presidency Translator' is centered. On the right, there is a 'Menu' icon and language options 'EN DE'. Below the title are three buttons: 'Translate text' (dark red), 'Translate document' (red), and 'Translate website' (blue). The interface is set against a background image of a mountain range. Below the buttons, the 'From' language is set to 'English' and the 'To' language is set to 'German'. A 'Translate' button is on the right. The main area contains a text input field with the placeholder 'Enter the text you want to translate.' and a disclaimer: 'Machine translation results help to understand the meaning of a source text, but do not equal translation by a human.' At the bottom left, it says 'Powered by Tilde MT and eTranslation, the EU's automated translation service'.

e 2 0  
u 1 8  
a t

Austrian  
Presidency  
of the  
Council of the  
European Union

## EU Council Presidency Translator

Menu EN DE

Translate text Translate document Translate website

From: **English** German Bulgarian ▼

To: **German** Bulgarian Czech ▼

Translate

Enter the text you want to translate.

Machine translation results help to understand the meaning of a source text, but do not equal translation by a human.

Powered by **Tilde MT** and **eTranslation**, the EU's automated translation service



## Translate2018.eu

- Web
- CAT Tool-Plugin: SDL Trados
- Mobil

## Mensch mit Maschine...?

- Kein einheitlicher Ansatz
- Generische Tools vs. maßgeschneiderte Tools
- Datensicherheit, Rechtssicherheit
- Angemessene Daten, Prozesse, ExpertInnen
- Berufsprofil
- Ausbildung von TranslatorInnen:
  - European Master in Translation (EMT)
- Bewusstsein bei Entscheidungsträgern

# Bibliographie

- Bentivogli, L. et al. (2016): "Neural versus phrase-based machine translation quality: a case study." arXiv preprint, <https://arxiv.org/abs/1608.04631>.
- Lommel, A. et al. (2015): „Multidimensional Quality Metrics (MQM) Definition.“ <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>
- Thicke, L. (2013). "The industrial process for quality machine translation." *JoSTrans*, 19/2013. [http://www.jostrans.org/issue19/art\\_thicke.php](http://www.jostrans.org/issue19/art_thicke.php)
- Toral, A., & Sánchez-Cartagena, V. M. (2017). "A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions." *arXiv preprint*, <https://arxiv.org/abs/1701.02901>