



Das eJournal der Europäischen Rechtslinguistik (ERL)
Universität zu Köln

Machine Translation and Neural Networks for a multilingual EU

Pascale Chartier-Brun
Katharina Mahler

14.11.2018

[urn:nbn:de:hbz:38-85979](https://nbn-resolving.org/urn:nbn:de:hbz:38-85979)

www.zerl.uni-koeln.de



ABSTRACT ENGLISH

This paper presents an overview of the current developments and use of Machine Translation (MT) and Neural Machine Translation (NMT), specifically eTranslation, in the European Institutions. An insight into the state-of-the-art of NMT as currently in development in the Directorate-General for Translation (DG TRAD) of the European Parliament is provided by Pascale Chartier-Brun. Problems in machine translation support requiring further research and development for processing languages with complex morphosyntax are discussed in the outlook. This paper was developed from the presentation "IT integrated environment for optimising the translation of legislative documents in the EP" by Pascale Chartier-Brun at the workshop "Europäische Rechtslinguistik und Digitale Möglichkeiten / EU Legal Linguistics and Digital Perspectives", held at the University of Cologne July 7th/8th, 2017.



ABSTRACT DEUTSCH

Dieser Beitrag bietet einen Überblick in die aktuellen Entwicklungen und den Einsatz von Machine Translation (MT) und Neural Machine Translation (NMT), insbesondere eTranslation, in den Europäischen Institutionen. Einen Einblick in den Stand der Technik der NMT, wie er derzeit in der Generaldirektion Übersetzung (DG TRAD) des Europäischen Parlaments entwickelt wird, gibt Pascale Chartier-Brun. Probleme bei der Unterstützung durch maschinelle Übersetzung, die weitere Forschung und Entwicklung für die Verarbeitung von Sprachen mit komplexer Morphosyntax erfordern, werden im Outlook diskutiert. Dieser Beitrag wurde aus dem Vortrag "IT-integrierte Umgebung zur Optimierung der Übersetzung von Gesetzesdokumenten im EP" von Pascale Chartier-Brun im Rahmen des Workshops "Europäische Rechtslinguistik und Digitale Möglichkeiten / EU Legal Linguistics and Digital Perspectives" an der Universität Köln am 7. und 8. Juli 2017 entwickelt.

Contents

1. Multilingual EU: Many languages need to be translated	1
2. Machine translation at the EU: MT@EC and eTranslation	2
3. Natural Language Processing and Neural Machine Translation	4
4. Translation and NMT at the EU Parliament.....	7
4.1 Phase 1 : Pretreatment and collect of translation memories.....	7
4.2 Phase 2: Translation using memories	8
5. Outlook: further development necessary for processing languages with complex morphosyntax	10
6. Conclusion.....	13
7. References	15

The authors would like to thank Dr. Jürgen Hermes, Institute for Digital Humanities at the University of Cologne for his specialist expertise in proofreading this paper. All shortcomings remain the responsibility of the authors.

1. Multilingual EU: Many languages need to be translated

< 1 >

As noted by the Commission of the European Communities, the "harmonious co-existence of many languages in Europe is a powerful symbol of the EU's aspiration to be united in diversity, one of the cornerstones of the European project."¹ A successful multilingualism policy is of utmost importance to bridge the linguistic diversity in the European Union (EU) with its 28 member states, 500 million citizens, 3 alphabets and 24 official languages.² The 24 official languages of the EU enjoy equal status and obligate the EU institutions to ensure the highest degree of multilingualism. There are 552 possible language combinations that must be managed in document translation, as each language can be translated into 23 other languages.³

< 2 >

To meet this challenge, the EU institutions, including the European Commission (EC) and the European Parliament (EP), have established "highly efficient interpreting, translation and legal text verification services".⁴ A wide range of tools and technologies speed up the translation process, reduce the risk of human error, and improve consistency through use of translation memories and reference to terminological and documentary databases (see, e.g., CHARTIER-BRUN 2018). Also, new tools are regularly developed to respond to the needs of modern translations services and to further improve the quality and efficiency of the translators' work.⁵ This includes the development and implementation of Machine Translation (MT).

< 3 >

Current state-of-the-art machine translation systems are trained on very large collections of parallel texts translated by humans. With the use of sophisticated algorithms, MT systems are able to mine this parallel data to produce automatic translations.

Although MT may not offer the same level of accuracy and quality as human translation, it does give insight into the general meaning of a text, "thus helping [...] to cross language

¹ See <https://web.archive.org/web/20181219115349/http://ec.europa.eu/transparency/regdoc/rep/1/2008/EN/1-2008-566-EN-F1-1.Pdf>, 4. All links were accessed 19.12.2018.

² Currently these are: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish and Swedish. See https://web.archive.org/web/20181219115513/https://europa.eu/european-union/topics/multilingualism_en.

³ Or 506 combinations (see video at <https://web.archive.org/web/20181219115552/http://www.qt21.eu/>), if Irish is excluded: "EU regulations and other legislative texts are published in all official languages except Irish (only regulations adopted by both the EU Council and the European Parliament are currently translated into Irish)", see https://web.archive.org/web/20181219115513/https://europa.eu/european-union/topics/multilingualism_en.

⁴ See <https://web.archive.org/web/20181219115731/http://www.europarl.europa.eu/about-parliament/en/organisation-and-rules/multilingualism>.

⁵ See https://web.archive.org/web/20181219115821/http://www.europarl.europa.eu/pdf/multilinguisme/EP_translators_en.pdf.

barriers between nations and facilitate multilingual communication".⁶ Further, MT technology is used by the translation departments of EU institutions, for example, the Directorate-General for Translation (DGT) of the European Commission and Directorate-General for Translation (DG TRAD) of the European Parliament, to provide translation segments for the human translators.

< 4 >

The following pages will present a look at the current state of the development and implementation of machine translation at the EU institutions, with a special focus on Neural Machine Translation (NMT), see section below, and the usage of MT and NMT at the European Parliament.

2. Machine translation at the EU: MT@EC and eTranslation

< 5 >

Machine translation has been available for EU translators since June 2013, when the MT@EC system (machine translation system) developed by the European Commission went online.⁷ The MT@EC system is based on the MOSES open-source translation toolkit, which is a Statistical Machine Translation (SMT) system.⁸ Currently MT@EC is being gradually replaced by a new system based on Neural Machine Translation. This program is called eTranslation⁹ and was launched in November 2017.

< 6 >

The eTranslation service¹⁰ can translate between any pair of the 24 official EU languages, as well as Icelandic and Norwegian (Bokmål): it can handle formatted documents and plain text; it translates multiple documents into multiple languages in "one go"; it accepts diverse input formats including XML and PDF; it retains formatting; and it provides specific output formats

⁶ See <https://web.archive.org/web/20181219115920/http://lr-coordination.eu/discover>.

⁷ See <https://web.archive.org/web/20181219120107/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/What+is+eTranslation+-+MT@EC+and+eTranslation>.

⁸ For more information on MOSES, see <https://web.archive.org/web/20181219120205/http://www.statmt.org/moses/>.

⁹ eTranslation is a building block of Connecting Europe Facility (CEF), and is also referred to as CEF eTranslation. See <https://web.archive.org/web/20181219120244/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation> and <https://web.archive.org/web/20181219120351/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/What+is+eTranslation+-+Overview>.

¹⁰ See <https://web.archive.org/web/20181219120544/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/How+does+it+work+-+eTranslation>.

for computer-aided translation, i.e., TMX¹¹ and XLIFF¹². MT@EC and eTranslation can be used by the European institutions and bodies, e.g., the Commission, Parliament, Council, Court of Justice, as well as by public administrations in the EU/EEA (European Economic Area) countries and online services funded or supported by the EU.¹³ Depending on the language pair chosen for translation, eTranslation draws upon either the statistical or neural machine translation system.

< 7 >

Both phases of the machine translation system (MT@EC and eTranslation) are trained using the translation memories contained in Euramis (European advanced multilingual information system), which currently encompasses around 1.2 billion translation segments in the 24 official EU languages produced by translators of EU institutions over the past decades.^{14, 15} Euramis serves as a "multilingual, multidirectional repository of clearly labelled equivalent phrases ("segments")".¹⁶ These segments are matched up, i.e., aligned, with the equivalent translation segments in other languages.¹⁷ By drawing upon these equivalently translated segments, the translation systems are thus capable of handling and processing specific EU policy and legal terminology with a high level of accuracy.¹⁸

< 8 >

The quality of the translation depends on three main factors: the complexity of the language pairs being translated; the language style (e.g., better output for EU official style, very long and very short sentences can be difficult, literary and conversational style are not supported

¹¹ TMX stands for Translation Memory eXchange and is an XML-compliant format, <https://web.archive.org/web/20181219120725/https://www.gala-global.org/tmx-14b>. This format is used in, e.g., Euramis, see below.

¹² XLIFF stands for XML Localization Interchange File Format, see <https://web.archive.org/web/20181219120830/http://docs.oasis-open.org/xliff/xliff-core/v2.1/xliff-core-v2.1.html>. This format is not currently used in DG TRAD but it is sometimes requested (correspondence with DGT, Information Technology Unit, EC, July 2018).

¹³ Public administrations in the EU and EEA countries and online services, see PILOS (2016), "How can public services benefit from the CEF.AT platform", https://web.archive.org/web/20181219120940/http://r-coordination.eu/sites/default/files/Denmark/20160307%20-%20ELRC_benefits_D%20SpyrosK.pdf, 12-15.

¹⁴ Correspondence with the Information Technology Unit, DGT, EC, July 2018. Euramis contains translation memories in all 24 languages, with varying sample volumes. The translation segments come from public legislation of the EU and also current translations of the EU Commission, Parliament, Council, Court of Justice, etc.

¹⁵ While Euramis itself is not directly accessible for the public, extractions of Euramis translation units can be accessed freely via DGT-TM (DGT-Translation Memory), see <https://web.archive.org/web/20181219121230/https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory>. As of 2018, DGT-TM contains ca. 120 million aligned translation units or 2 billion words, see https://web.archive.org/web/20181219121331/https://wt-public.emm4u.eu/Resources/DGT-TM_Statistics.pdf.

¹⁶ See https://web.archive.org/web/20181219121503/http://www.europarl.europa.eu/meetdocs/2009_2014/documents/budg/dv/2010_c4_implem_euramis_dgtrad_/2010_c4_implem_euramis_dgtrad_en.pdf.

¹⁷ The translation segments in Euramis are generally aligned with about 10 languages. Correspondence with the Information Technology Unit, DGT, EC, July 2018.

¹⁸ See <https://web.archive.org/web/20181219121555/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/How+does+it+work+-+eTranslation>.

so well); and the subject matter (e.g., for unknown domains and terminology, some terms might not be translated). Regarding complexity, for example, "pronoun systems do not map well across languages, e.g., due to differences in gender, number, case, formality, or humanness, and to differences in where pronouns may be used. Translation divergences typically lead to mistakes in SMT, as when translating the English "it" into French ("il", "elle", or "cela"?) or into German ("er", "sie", or "es"?).¹⁹

< 9 >

In answer to questions regarding the technical development of eTranslation, the Directorate-General for Translation (DGT) at the European Commission²⁰ notes that the development in NMT is in full flow, and that part of their task is to monitor and test different configurations. Open source toolkits are used to build translation engines for implementation. There is no commitment to a single approach: currently DGT is working heavily with convolutional models (see section 3), but this is likely to change in the future. The training is supervised and takes place in an "engine factory" of DGT's own design, allowing monitoring, debugging, restarting failed or unsatisfactory trainings from an earlier point, etc. Currently the favored toolkit is OpenNMT,²¹ but developments in other toolkits are being followed closely. "Basically at this point we are keeping our options open, and we believe that to be the appropriate approach given the current state of the technology."²²

3. Natural Language Processing and Neural Machine Translation

< 10 >

Neural machine translation is based on the model of neural networks (NN) which consist of many 'neurons' in different layers and imitate the functioning of biological neural networks. Artificial neural networks were originally developed for visual recognition, e.g., Optical Character Recognition (OCR), but can be used very efficiently for natural language processing (NLP).²³

¹⁹ See <https://web.archive.org/web/20181219121826/http://www.statmt.org/wmt16/>.

²⁰ E-Mail correspondence with the DGT, EC, April 2018.

²¹ See the OpenNMT homepage at <https://web.archive.org/web/20181219121851/http://opennmt.net/> for more information.

²² E-Mail correspondence with the DGT, EC, April 2018.

²³ See, e.g., the Neural Machine Translation ACL 2016 tutorial by LUONG, Thang / CHO, Kyunghyun / MANNING, Christopher (2016), <https://web.archive.org/web/20181219121924/https://sites.google.com/site/acl16nmt/home>, (slides under <https://web.archive.org/web/20181219122019/https://nlp.stanford.edu/projects/nmt/Luong-Cho-Manning-NMT-ACL2016-v4.pdf>).

< 11 >

If one wants to learn Finnish, for example, as Josef van Genabith, Scientific Director of the German Research Center for Artificial Intelligence (DFKI), explains, one must be "prepared to deal with a complex grammar that includes fifteen different cases. The grammatical cases are marked in part by appending syllables to nouns resulting in a dizzying array of word forms and expressive possibilities." ²⁴ It would be "exceptionally difficult" to teach a computer to process and translate all these "grammatical nuances" into a different language.²⁵

< 12 >

Instead of feeding translation systems with grammar rules and linguistic details, the systems implement neural networks that "are taught to recognize patterns in huge text repositories and to learn from them".²⁶ When the neural network is composed of many layers, this is called "Deep Learning".

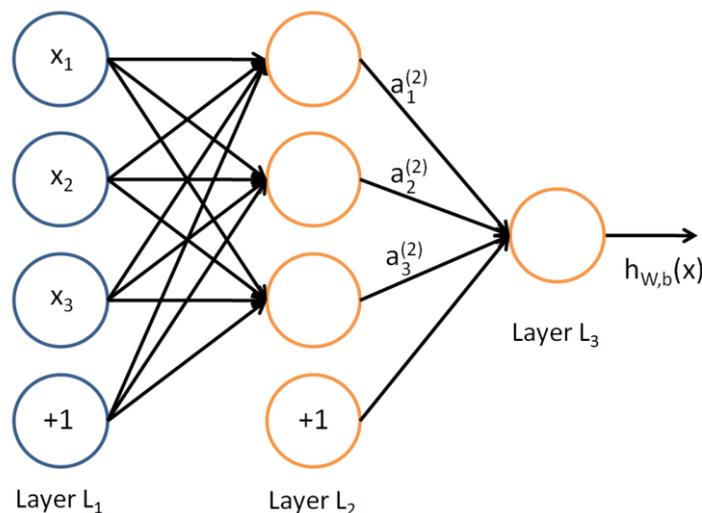


Figure 1: Neural Network with an input layer (L₁), hidden layer (L₂) and output unit (L₃) ²⁷

< 13 >

Neural networks can be understood as an encoder-decoder architecture.²⁸ The input layer L₁ (encoder) encodes the surface representation (a sequence of words, etc.) and maps it to a

²⁴ See MEYER ZU TITTINGDORF, Friederike (2016), „Researchers want to achieve machine translation of the 24 languages of the EU“, <https://web.archive.org/web/20181219122118/https://idw-online.de/de/news655897>.

²⁵ Ibid.

²⁶ Ibid.

²⁷ Image from <https://web.archive.org/web/20181219122308/http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/>.

²⁸ GOODFELLOW, Ian / BENGIO, Yoshua / COURVILLE, Aaron (2016). *Deep Learning*. Boston: MIT Press, 469. <https://web.archive.org/web/20181219122348/https://www.deeplearningbook.org/>.

semantic representation in the hidden layer (L₂). This hidden semantic representation is then decoded by the output unit (decoder), as illustrated below.

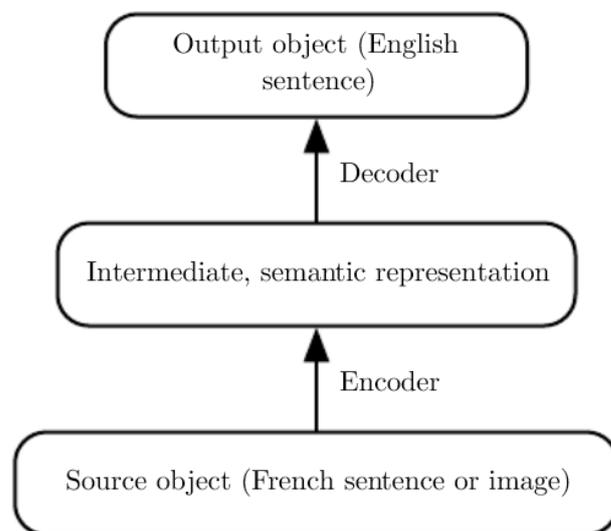


Figure 2: Encoder-decoder architecture with hidden semantic representation ²⁹

Thus, neural networks trained for pattern recognition can learn to process a certain input (e.g., French sentence) and return a specific output in another language (e.g., English sentence).

< 14 >

Different types of architectures for neural networks are possible.³⁰ For example, Convolutional Neural Networks (CNNs), as currently implemented by the EC, have a feed-forward structure composed of an input and output layer and multiple hidden layers.³¹ The research project QT21, investigating solutions for morphologically rich languages, uses Recurrent Neural Networks (RNNs), see section 5 below. This type of architecture is an extension of conventional feedforward networks and can handle sequential input of variable lengths. "The RNN handles the variable-length sequence by having a recurrent hidden state whose activation at each time is dependent on that of the previous time", see Chung et al. (2014: 1). An RNN thus re-uses the output of the hidden layer(s) to process the next sequences, i.e., the

²⁹ Image from GOODFELLOW / BENGIO / COURVILLE (2016: 469).

³⁰ See GOODFELLOW / BENGIO / COURVILLE (2016) or, for a colorful overview, e.g., Fjodor VAN VEEN (2016), "A mostly complete chart of Neural Networks", <https://web.archive.org/web/20181219122429/http://www.asimovinstitute.org/neural-network-zoo/>.

³¹ For convolutional neural networks research, see, e.g., KIM, Yoon (2014), "Convolutional Neural Networks for Sentence Classification", <https://web.archive.org/web/20181219122554/http://emnlp2014.org/papers/pdf/EMNLP2014181.pdf> or KALCHBRENNER, Nal / GREFFENSTETTE, Edward / BLUNSOM, Phil (2014), "A Convolutional Neural Network for Modelling Sentences", <https://web.archive.org/web/20181219122718/http://www.aclweb.org/anthology/P14-1062>.

output of the internal state is stored in a memory and can be used to predict following words and to capture long distance dependencies.

4. Translation and NMT at the EU Parliament

< 15 >

The DG TRAD at the European Parliament is in the process of establishing a proof-of-concept (PoC) in two steps: the purpose of the first step is to rank the candidate segments coming from Euramis and MT by pertinence, the second step is to post edit automatically the best segment (more pertinent) using extra resources such as terminology. The main objectives are to better classify the translation segments proposed to translators through their tools and to do automatic post editing, as will be explained in more detail below. Many different applications work together to support the translation workflow at the European Parliament, which has two phases: pretreatment and translation.

4.1 Phase 1 : Pretreatment and collect of translation memories

< 16 >

Original documents (documents to be translated) are automatically pretreated by the SPA (Safe working Protocol Automation) system. The main goal of the protocol is to retrieve translation memories in Euramis that correspond to the original document in the target languages and also to define the list of reference documents linked to the original in the legislative procedure and to retrieve their translation from Euramis. The original document is split into segments of text using mainly punctuation (segmentation rules) and sent to Euramis to retrieve the translation of these segments with a matching rate (retrieval process). For the reference documents (that were translated earlier), SPA downloads the available memories. After the pretreatment, SPA feeds the central memory server called FullCat using XML standard format (TMX Translation Memory eXchange). The segments are ranked dynamically according to predefined rules during the translation.³² The SPA system is currently being modernized with a flexible workflow engine. This will allow flexibility and easy adaptation of the pretreatment workflow for different document types and to include additional steps such as quality control of the original text, a different text extraction, a better treatment of

³² See CHARTIER-BRUN, Pascale (2018), "Translating legislative documents at the European Parliament: e-Parliament, XML, SPA and the Cat4Trad workflow", <https://web.archive.org/web/20181219122904/http://zerl.uni-koeln.de/chartier-brun-2018-translation-workflow-ep.html>, for details on the practical implementation of the translation workflow.

multilingual texts or any other new feature. The original documents are mainly in two formats, Word and XML4EP for the documents of the legislative work.³³

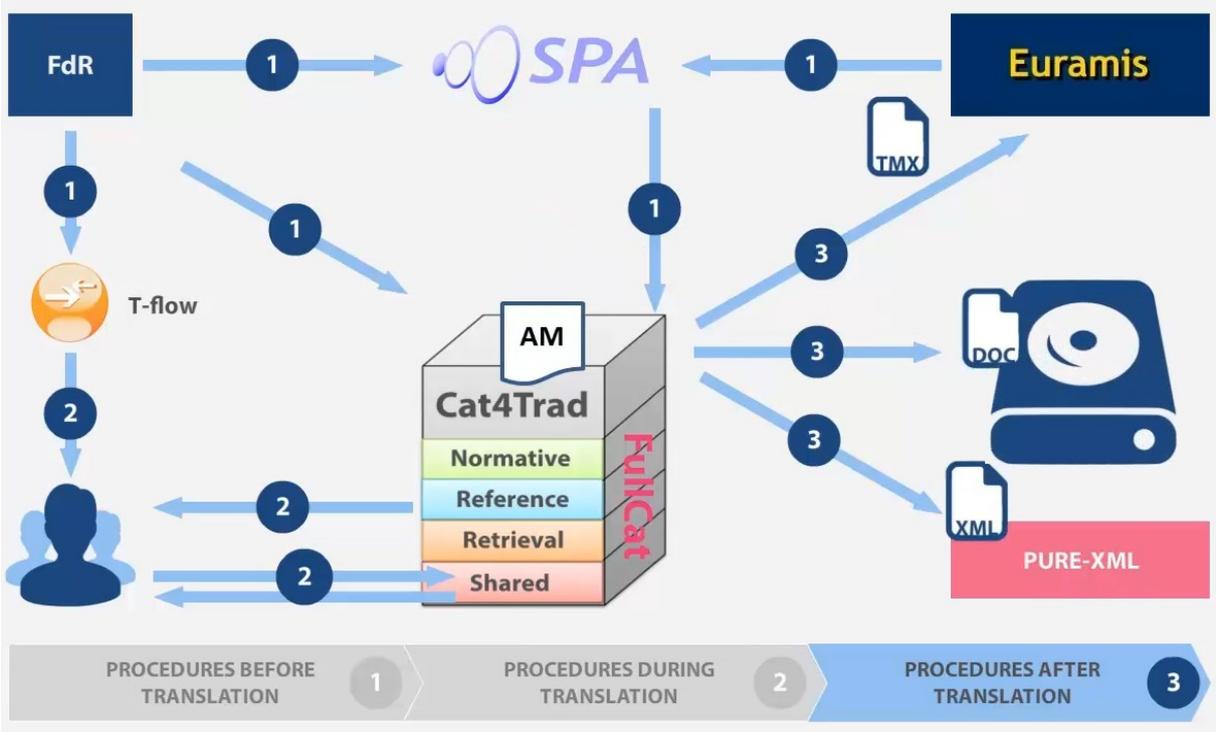


Figure 3: eParliament: translation workflow with SPA, Euramis, translation memories and Cat4Trad³⁴

Parallel to the collection of Euramis memories, SPA calls a service in the European Commission in order to translate the original using eTranslation. The result of the MT processing is a memory (TMX file) which is sent by SPA to a special index of FullCat.

4.2 Phase 2: Translation using memories

< 17 >

The identification of relevant translation memories for translations is based on a set of predefined rules. These rules are complex and difficult to maintain, and could probably benefit from the use of Machine Learning (ML) techniques. Thus, DG TRAD is establishing a proof-of-concept for the ranking of segments coming from Euramis (retrieved by SPA) and also

³³ XML4EP is an adaptation of Akoma Ntoso (XML format for parliamentary, legislative and judiciary documents) for the EP, see <https://web.archive.org/web/20181219123051/http://www.akomantoso.org>.

³⁴ Illustration used with permission, courtesy of EP (2014-2017). For details see CHARTIER-BRUN (2018).

coming from machine translation, i.e., NMT, using the eTranslation system of the European Commission.³⁵ The proof-of-concept has two main objectives.

< 18 >

The first objective is to better classify the segments that are proposed to translators through their CAT (computer-assisted translation) Tools (CAT4TRAD and SDL Trados Studio). Five segments from Euramis and one segment from MT are proposed to the translator. The results coming from MT receive a 30% penalty in the FullCAT Server, i.e., the match rate is reduced to 70%.³⁶ The other segments are ranked by their Euramis matching rate and then by their age with the youngest first. This situation penalizes the MT segment and does not take into account the context of the document, i.e., whether the segment originates from legislation, case law, terminology etc. The PoC based on machine learning will order these segments regardless of their origin. Four experiments are foreseen for scoring the segments. The simplest is based on Doc2vec (converts a document into vectors) algorithm and Euclidian distance, the more complex is using Word2vec (converts words into vectors) and neural network model. The calculation of the score is a product of the matching rate, the Bleu Score³⁷, the WAcc and METEOR³⁸.

< 19 >

The second objective is to do automatic post editing. The PoC will propose, based on the "best" segment of the six available, a new and improved segment that takes the context into account. The best segment selected in the first phase will be rewritten using a machine learning solution. Among the possible solutions, the simplest has been selected for experimentation (without using neuronal networks). This solution uses the Doc2vec algorithm, a corpus per language and a data model. Content from IATE (InterActive Terminology for Europe)³⁹ is integrated in the first experimentation, but other sources such as dictionaries, Trans fields or normative memories could be added in a second experiment phase.

³⁵ See <https://web.archive.org/web/20181219123217/http://www.europarl.europa.eu/the-secretary-general/en/directorates-general/trad>.

³⁶ To compensate for the lower quality of MT compared to a human translation, results from MT have a flat penalty of 30% applied. This does not capture the differences in quality between language combinations, or the evolution of quality in the MT results proposed by the new neural MT engines. Therefore, a more intelligent ranking system could identify the cases when MT is preferable to other matches.

³⁷ PAPANENI, Kishore / ROUKOS, Salim / WARD, Todd / ZHU, Wei-jing (2002), "BLEU: a method for automatic evaluation of machine translation", 311–318, <https://web.archive.org/web/20181219123352/https://www.aclweb.org/anthology/P02-1040.pdf>.

³⁸ See <https://web.archive.org/web/20181219123508/https://www.cs.cmu.edu/~alavie/METEOR/>.

³⁹ See <http://iate.europa.eu/home>. More on corpora: Steinberger et al. (2014), "An overview of the European Union's highly multilingual parallel corpora". *Language Resources & Evaluation* 48 (4): 679-707, https://web.archive.org/web/20181220100033/https://ec.europa.eu/jrc/sites/jrcsh/files/2014_08_LRE-Journal_JRC-Linguistic-Resources_Manuscript.pdf

< 20 >

The results of the experiment phase are promising, especially the rewriting step, which has provided very good results during the current tests. The automation of machine translation is in full production with great success. As of June 2018, the translators at DG TRAD are able to benefit from new engines based on Neural Machine Translation.

5. Outlook: further development necessary for processing languages with complex morphosyntax

< 21 >

All official EU languages are equal, but not all are equally supported. For some languages, the MT support is sub-optimal due to processing difficulties based on, e.g., free constituent order or rich morphology.⁴⁰ According to the findings of the META-NET (Multilingual Europe Technology Alliance Network) Language White Papers,⁴¹ in 2010-2011 only English, French and Spanish of the EU-27 languages⁴² enjoyed "moderate to good support by our machine translation technologies, with either weak (at best fragmentary) or no support for the vast majority of the EU-27 languages".⁴³ An update and extension of the META-NET study in 2014 confirmed the original results and painted an alarming picture, demonstrating even more dramatic differences in language technology support between the European languages. This technology gap can endanger languages to the point of digital extinction.⁴⁴

⁴⁰ See the FAQ of the ELRC, <https://web.archive.org/web/20181219123837/http://lr-coordination.eu/faq>, <http://lr-coordination.eu/faq#281>.

⁴¹ See the overview at META-NET (published in 2012), <https://web.archive.org/web/20181219124131/http://www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison>.

⁴² For EU-27, see, e.g., https://web.archive.org/web/20181219124250/https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_enlargements or <https://web.archive.org/web/20181219124331/https://datacollection.jrc.ec.europa.eu/eu-27>.

⁴³ "Introduction", QT21 Quality Translation, <https://web.archive.org/web/20181219115552/http://www.qt21.eu/>.

⁴⁴ This is especially true for non-official minority languages. As of 2012, at least 21 European languages were in danger of digital extinction, according to an assessment of 30 out of about 80 European languages that were examined regarding the level of language technology support, see <https://web.archive.org/web/20181219124842/http://www.meta-net.eu/whitepapers/press-release>. For a visual overview of the state of language technology support in 2014, see the poster at https://web.archive.org/web/20181219124939/http://www.ilc.cnr.it/ccurl2014/LREC2014-Workshops_CCURL2014-Poster-1_Presentation.pdf. For the update of the META-NET study, see REHM, Georg / USZKOREIT, Hans / DAGAN, Ido et al. (2014a), "An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age"", <https://web.archive.org/web/20181219125108/http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014-Workshop-CCURL2014-Proceedings.pdf>. See also, e.g., REHM, Georg (2013), "Europe's languages in the Digital Age: Multilingual Technologies for overcoming Language Barriers and Preventing Digital Language Extinction", <http://docplayer.gr/3524920-Europe-s-languages-in-the-digital-age.html> and REHM, Georg / USZKOREIT, Hans / ANANIADOU, Sophia et al. (2016), "The strategic impact of META-NET on the regional, national and international level", <https://web.archive.org/web/20181219125645/https://helda.helsinki.fi/bitstream/handle/10138/136267/METAImpact.pdf?sequence=1&isAllowed=y>.

< 22 >

Common traits of many of the languages that are not well-supported include morphological complexity and free and diverse word order. Another problem is that often not enough language resources for training and/or processing tools are available.⁴⁵ MT and NMT systems can only be as good as the data they are trained on, i.e., that they learn from. This means that the higher the quality of the language data on which learning is based, the better the translation results will be.

< 23 >

To bridge the technology gap and to improve the quality and coverage of CEF eTranslation, language resources and translation data are needed on a much larger scale. Thus, the European Commission has launched a comprehensive European Language Resource Coordination (ELRC)⁴⁶ endeavor to identify and gather relevant language and translation data across the 30 European countries that are participating in the CEF (Connecting Europe Facility) program.⁴⁷ The ELRC is currently "one of the most comprehensive collections of language data worldwide".⁴⁸ Also, further research in machine translation for developing modified neural networks that better suit these languages⁴⁹ has been funded by the European Union's Horizon 2020 research and innovation program.

< 24 >

QT21 (Quality Translation 21)⁵⁰ is one of the research groups funded by Horizon 2020, focusing on syntactically varied and morphologically complex languages. At the WMT 2016 conference⁵¹ on machine translation, which "brought a paradigm shift with the introduction of deep neural networks",⁵² QT21 showed excellent performance by further improving upon the use of neural networks as a base for MT tasks, outperforming Google Translate⁵³ (which was

⁴⁵ "Introduction", <https://web.archive.org/web/20181219115552/http://www.qt21.eu/>.

⁴⁶ See <https://web.archive.org/web/20181219125752/http://www.lr-coordination.eu/>.

⁴⁷ See <https://web.archive.org/web/20181219121555/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/How+does+it+work+-+eTranslation>.

⁴⁸ See MEYER ZU TITTINGDORF (2016), <https://web.archive.org/web/20181219122118/https://idw-online.de/de/news655897>.

⁴⁹ See <https://web.archive.org/web/20181219123837/http://lr-coordination.eu/faq>, <http://lr-coordination.eu/faq#281>.

⁵⁰ QT21 is a consortium of 14 leading machine translation research institutes in Europe and Hong Kong, including universities, research institutes and numerous companies, led by the DFKI, the German Research Center for Artificial Intelligence, see <https://web.archive.org/web/20181219115552/http://www.qt21.eu/>.

⁵¹ See <https://web.archive.org/web/20181219121826/http://www.statmt.org/wmt16/> and <https://web.archive.org/web/20181219130428/http://www.statmt.org/wmt16/pdf/W16-2301.pdf>.

⁵² See <https://web.archive.org/web/20181219130547/https://tilde.com/news/deep-learning-european-project-qt21-tops-international-machine-translation-competition-second>.

⁵³ Two months after this conference, Google Translate also moved to neural networks with noticeable improvements, citing a major QT21 paper as one of their most relevant papers, *ibid*.

using statistical MT at the time) and others. Of the five language pairs that QT21 focused on, four have English as a source language (English->German, English->Czech, English->Latvian, English->Romanian) and one has English as the target language (German->English).⁵⁴

< 25 >

The QT21 research brought forth extensive improvements in the translation of these morphologically rich languages. These achievements were "possible due to the training of deep neural networks in conjunction with a pre-processing step that creates artificial sub-word units or segments based on the Gage byte pair encoding compression algorithm (1994), in which, instead of merging frequent pairs of bytes, characters or character sequences are merged."⁵⁵ The project homepage provides technical details as well as language data, i.e., a corpus of Czech, English, German and Latvian, and the software that was developed (see "Resources").⁵⁶

< 26 >

Successful steps have thus been taken in the right direction to improve the coverage of at least some of the under-supported complex languages since the findings of the META-NET White papers in 2012 and 2014. At that time, German and Romanian were receiving "fragmentary support" at best, and Czech, Latvian, Bulgarian and Estonian were receiving "weak" to "no support". The situation for these languages has been greatly improved through concerted efforts, see the QT21 results and also the EU Council Presidency Translator. The EU Council Presidency Translator was initially implemented during the Estonian EU Council Presidency in 2017, where it "introduced the world's first AI-powered Neural MT engines for Estonian".⁵⁷

< 27 >

The eTranslation system, i.e., the CEF platform, is restricted to public administrations. Thus, it is not directly available to the general public and also not in competition with other openly available NMT systems such as DeepL or GoogleTranslate. The reason for this is that the Commission is not a private enterprise and it is not supposed to enter into competition with private enterprises. Thus, eTranslation is intended to support language technologies while avoiding distortion of the market.⁵⁸

⁵⁴ See "Achievements", <https://web.archive.org/web/20181219115552/http://www.qt21.eu/>.

⁵⁵ See <https://web.archive.org/web/20181219130547/https://tilde.com/news/deep-learning-european-project-qt21-tops-international-machine-translation-competition-second>.

⁵⁶ Further, "Core technical contributions include "back translation" that allows to synthetically augment the training data volume, Byte Pair Encoding (BPE) to compress vocabularies of Morphologically Rich Languages (MRL), layer normalisation and deeper gated recurrent neural networks (GRU)". See "Achievements", <https://web.archive.org/web/20181219115552/http://www.qt21.eu/>.

⁵⁷ Developed by Tilde (<https://web.archive.org/web/20181219130753/https://www.tilde.com/>), see <https://web.archive.org/web/20181219130913/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2017/10/19/CE+eTranslation+Used+During+Estonian+EU+Council+Presidency>.

⁵⁸ Correspondence with the Information Technology Unit, DGT, EC, July 2018.

< 28 >

At the same time, the CEF platform does offer grants to private enterprises to promote growth in the language sector. One of these grants is provided for the EU Council Presidency translator, which is available to the public. The partners in this project provide their own system and translation engines, and can use the eTranslation engines to cover languages that are not well supported yet.⁵⁹ As of the beginning of 2018, this service is presented on the website of the current Bulgarian Presidency of the Council of the European Union, with a custom engine for Bulgarian.⁶⁰ Thus, eTranslation is available to the general public through certain projects, i.e., their websites.⁶¹

6. Conclusion

< 29 >

Machine translation has been implemented with much success for EU translation, starting with the statistical MT@EC system and the further development to eTranslation, based on neural network technology. Differing demands on MT may be distinguished: within DG TRAD, MT is implemented as an extra method to select and propose translation equivalents to the professional translators, whereas in general cross-language communication, MT can be seen more as a very useful tool for supporting and improving multilingual communication in the EU and partner countries. No matter how good MT is or one day may be, proposed MT translations should always be checked by human translators before official use.

< 30 >

Further steps to improve the support for other languages must be taken, i.e., high-quality language data must be collected and implemented in customized resources. Such concerted efforts in research and development, as already being undertaken by the DGT and DG TRAD of the EC and the EP, the ELRC, and various research groups including QT21 and META-NET, show great promise to provide better MT support for languages that have not received optimal support in the past. The inclusion of specific language data and NMT tools specifically designed for more complex languages has already brought forth very good results.

⁵⁹ Correspondence with the Information Technology Unit, DGT, EC, July 2018.

⁶⁰ EU Council Presidency Translator, see <https://eu2018bg.bg/en/translation> and <https://web.archive.org/web/20181219132856/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2018/01/26/Official+Website+of+2018+Bulgarian+EU+Council+Presidency+Integrates+CEF+eTranslation>. The Bulgarian engine is also developed by Tilde.

⁶¹ Many other EU websites are connected to eTranslation in a similar manner, e.g., Eur-Lex (<https://web.archive.org/web/20181219131614/https://eur-lex.europa.eu/homepage.html>) or the Online Dispute Resolution of the EC (<https://web.archive.org/web/20181219131712/https://ec.europa.eu/consumers/odr/main/index.cfm?event=main.home2.show&lng=EN>). Correspondence with the Information Technology Unit, DGT, EC, July 2018.

Much has been achieved in MT and NMT for translating in the multilingual EU in the last years. These achievements inspire hope and raise expectations that the many still under-supported languages will find better digital support in the future.

To make the EU truly multilingual, and to protect under-supported languages from digital extinction, it is of utmost importance to further invest in machine and neural machine translation research and development. The official EU languages can only be truly equal when they all receive equal digital support.

7. References

- BOJAR, Ondřej / CHATTERJEE, Rajen / FEDERMANN, Christian et al. (2016). "Shared Task Papers". In: Association for Computational Linguistics (ed.). *Proceedings of the First Conference on Machine Translation (WMT16)*. Stroudsburg: The Association for Computational Linguistics. Volume 2: 131–198 <https://web.archive.org/web/20181219130428/http://www.statmt.org/wmt16/pdf/W16-2301.pdf> (13.12.2018)
- CHARTIER-BRUN, Pascale (2018). "Translating legislative documents at the European Parliament: e-Parliament, XML, SPA and the Cat4Trad workflow". *Zeitschrift für Europäische Rechtslinguistik (ZERL)*. Köln: Europäische Rechtslinguistik. <https://web.archive.org/web/20181219122904/http://zerl.uni-koeln.de/chartier-brun-2018-translation-workflow-ep.html> (13.12.2018). URN of PDF: <urn:nbn:de:hbz:38-82669>.
- CHUNG, Junyoung / GÜLÇEHRE, Çağlar / CHO, KyungHyun / BENGIO, Yoshua (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling". *arXiv e-prints* arXiv: 1412.3555v1 [cs.NE]. Ithaca: Cornell University Library. <https://arxiv.org/pdf/1412.3555.pdf> (13.12.2018)
- COMMISSION OF THE EUROPEAN COMMUNITIES (2008). "Multilingualism: an asset for Europe and a shared commitment". Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 4. Brussels. <https://web.archive.org/web/20181219115349/http://ec.europa.eu/transparency/regdoc/rep/1/2008/EN/1-2008-566-EN-F1-1.Pdf> (13.12.2018)
- DENKOWSKI, Michael / LAVIE, Alon (2010-2011). "METEOR (Automatic Machine Translation Evaluation System)". Pittsburgh: Carnegie Mellon University. <https://web.archive.org/web/20181219123508/https://www.cs.cmu.edu/~alavie/METEOR/> (13.12.2018)
- EUROPEAN COMMISSION (2014). "Glossary: EU enlargements". In: eurostat - Statistics Explained. Brussels. https://web.archive.org/web/20181219124250/https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_enlargements (13.12.2018)
- EUROPEAN COMMISSION, CONNECTING EUROPE FACILITY (CEF) (2016). "Machine Translation service, Version 1.2, Service Offering description". Brussels. [https://web.archive.org/web/20181219132128/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Machine+Translation?preview=/46992195/48766166/ServiceOfferingDescription\(SOD\)_eTranslation_v1.2.pdf](https://web.archive.org/web/20181219132128/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Machine+Translation?preview=/46992195/48766166/ServiceOfferingDescription(SOD)_eTranslation_v1.2.pdf) (13.12.2018)
- EUROPEAN COMMISSION, CONNECTING EUROPE FACILITY (2017c). "CEF eTranslation Used During Estonian EU Council Presidency". In: CEF (Connecting Europe Facility) Digital News. Brussels. <https://web.archive.org/web/20181219130913/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2017/10/19/CEF+eTranslation+Used+During+Estonian+EU+Council+Presidency> (13.12.2018)
- EUROPEAN COMMISSION (2017a). "DGT-Translation memory". In: EU Science Hub. Luxembourg: Directorate-General for Translation (DGT). <https://web.archive.org/web/20181219121230/https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory> (13.12.2018)
- EUROPEAN COMMISSION (2017b). "DGT-Translation Memory". In: EU Science Hub. Luxembourg: Directorate-General for Translation (DGT). <https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory#dgt-memory> (13.12.2018)
- EUROPEAN COMMISSION, JOINT RESEARCH CENTRE (2018). "List of words – EU-27 and EU-28". In: DataCollection. Luxembourg. <https://web.archive.org/web/20181219124331/https://datacollection.jrc.ec.europa.eu/eu-27> (13.12.2018)

- EUROPEAN COMMISSION (2018). "Official Website of 2018 Bulgarian EU Council Presidency Integrates CEF eTranslation". In: CEF (Connecting Europe Facility) Digital News. Brussels. <https://web.archive.org/web/20181219132856/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2018/01/26/Official+Website+of+2018+Bulgarian+EU+Council+Presidency+Integrates+CEF+eTranslation> (13.12.2018)
- EUROPEAN COMMISSION (n.d. a). "Linguistic Diversity". In: Education and Training - Policies. Brussels. https://web.archive.org/web/20181219133234/https://ec.europa.eu/education/policies/multilingualism/linguistic-diversity_en (13.12.2018)
- EUROPEAN COMMISSION (n.d. b). "eTranslation". In: CEF Digital Home. Brussels. <https://web.archive.org/web/20181219120244/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation> (13.12.2018)
- EUROPEAN COMMISSION (n.d. c). "What is eTranslation? 1. Overview". In: CEF Digital Home eTranslation. Brussels. <https://web.archive.org/web/20181219133555/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/What+is+eTranslation+-+Overview> (13.12.2018)
- EUROPEAN COMMISSION (n.d. d). "What is eTranslation? 2. eTranslation and MT@EC". In: CEF Digital Home eTranslation. Brussels. <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/What+is+eTranslation+-+MT@EC+and+eTranslation> (13.12.2018)
- EUROPEAN COMMISSION (n.d. e). "How does it work?". In: CEF Digital Home eTranslation. Brussels. <https://web.archive.org/web/20181219120544/https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/How+does+it+work+-+eTranslation> (13.12.2018)
- EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) CONSORTIUM (2015-2016a). "ELRC-SHARE repository". n.p. <https://web.archive.org/web/20181219134145/https://elrc-share.eu> (13.12.2018)
- EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) CONSORTIUM (2015-2016b). "How can we access eTranslation?" In: FAQ of the ELRC. n.p. <https://web.archive.org/web/20181219123837/http://lr-coordination.eu/faq> <http://lr-coordination.eu/faq#281> (13.12.2018)
- EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) CONSORTIUM (2015-2016c). "European Language Resource Coordination – supporting Multilingual Europe". n.p. <https://web.archive.org/web/20181219125752/http://www.lr-coordination.eu/> (13.12.2018)
- EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) CONSORTIUM (2015-2016d). "Discover Automated Translation". n.p. <http://lr-coordination.eu/discover#f1%20%20automated%20translation> <https://web.archive.org/web/20181219115920/http://lr-coordination.eu/discover> (13.12.2018)
- EUROPEAN PARLIAMENT (2010). "Implementation of Euramis in DG TRAD". Strasbourg. https://web.archive.org/web/20181219121503/http://www.europarl.europa.eu/meetdocs/2009_2014/documents/budg/dv/2010_c4_implem_euramis_dgtrad_/2010_c4_implem_euramis_dgtrad_en.pdf (13.12.2018)
- EUROPEAN PARLIAMENT (n.d. a). "Multilingualism in the European Parliament". In: About Parliament. Organisation and rules. Strasbourg. <https://web.archive.org/web/20181219134805/http://www.europarl.europa.eu/about-parliament/en/organisation-and-rules/multilingualism> (13.12.2018)
- EUROPEAN PARLIAMENT (n.d. b). "EP Translators". Strasbourg. https://web.archive.org/web/20181219135027/http://www.europarl.europa.eu/about-parliament/files/organisation-and-rules/multilingualism/en-ep_translators.pdf (13.12.2018)

- EUROPEAN PARLIAMENT (n.d. c). "Directorate-General for Translation". In: The Secretary-General. Strasbourg. <https://web.archive.org/web/20181219123217/http://www.euoparl.europa.eu/the-secretary-general/en/directorates-general/trad> (13.12.2018)
- EUROPEAN UNION (2018a). "Multilingualism". In: EU by topic. Brussels. https://web.archive.org/web/20181219115513/https://europa.eu/european-union/topics/multilingualism_en (13.12.2018)
- EUROPEAN UNION, Translation Centre for the Bodies of the European Union (2018b). "Inter Active Terminology for Europe (IATE)". Luxembourg. <https://iate.europa.eu/home> (13.12.2018)
- GOODFELLOW, Ian / BENGIO, Yoshua / COURVILLE, Aaron (2016). *Deep Learning*. Boston: MIT Press. <https://web.archive.org/web/20181219122348/https://www.deeplearningbook.org/> (13.12.2018)
- KALCHBRENNER, Nal / GREFFENSTETTE, Edward / BLUNSON, Phil (2014). "A Convolutional Neural Network for Modelling Sentences". In: TOUTANOVA, Kristina / WU, Hua (eds.). *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: The Association for Computational Linguistics. Vol. I: 655–665. <https://web.archive.org/web/20181219122718/http://www.aclweb.org/anthology/P14-1062> (13.12.2018)
- KIM, Yoon (2014). "Convolutional Neural Networks for Sentence Classification". In: Association for Computational Linguistics (ed.). *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg: The Association for Computational Linguistics. 1746-1751. <https://web.archive.org/web/20181219122554/http://emnlp2014.org/papers/pdf/EMNLP2014181.pdf> (13.12.2018)
- KLEIN, Guillaume / KIM, Yoon / DENG, Yuntian / SENELLART, Jean / RUSH, Alexander M. (2017). "OpenNMT: Open-Source Toolkit for Neural Machine Translation". *arXiv e-prints* [arXiv: 1701.02810v2](https://arxiv.org/abs/1701.02810) [cs.CL]. Ithaca: Cornell University Library. <https://arxiv.org/abs/1701.02810> (13.12.2018)
- KOEHN, Philipp / HOANG, Hieu (2018). "Moses - Statistical Machine Translation System". Edinburgh. <https://web.archive.org/web/20181219120205/http://www.statmt.org/moses/> (13.12.2018)
- LUONG, Than / CHO, Kyunghyun / MANNING, Christopher (2016). "Neural Machine Translation". Tutorial presented at the ACL 2016. Stanford. <https://web.archive.org/web/20181219122019/https://nlp.stanford.edu/projects/nmt/Luong-Cho-Manning-NMT-ACL2016-v4.pdf> (13.12.2018)
- MEYER ZU TITTINGDORF, Friederike (2016). "Researchers want to achieve machine translation of the 24 languages of the EU". *Informationsdienst Wissenschaft (idw)*. Bayreuth & Universität des Saarlandes. <https://web.archive.org/web/20181219122118/https://idw-online.de/de/news655897> (13.12.2018)
- MINISTRY OF THE BULGARIAN PRESIDENCY OF THE EU COUNCIL (2018). "Bulgarian Presidency of the Council of the European Union". Sofia. <https://eu2018bg.bg/en/translation> (13.12.2018)
- NG, Andrew / NGIAM, Jiquan / FOO, Chuan Yu et al. (n.d.). "Multi-Layer Neural Network. Unsupervised Feature Learning and Deep Learning (UFLDL) Tutorial". Stanford: Stanford University. <https://web.archive.org/web/20181219122308/http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/> (13.12.2018)
- OASIS & UNITED NATIONS, Africa i-Parliament Action Plan (2018). "Akoma Ntoso, XML for parliamentary, legislative & judiciary documents". <https://web.archive.org/web/20181219123051/http://www.akomantoso.org> (13.12.2018)

- PAPINENI, Kishore / ROUKOS, Salim / WARD, Todd / ZHU, Wei-jing (2002). "BLEU: a method for automatic evaluation of machine translation". In: Association for Computational Linguistics (ed.). *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: The Association for Computational Linguistics. 311–318. <https://web.archive.org/web/20181219123352/https://www.aclweb.org/anthology/P02-1040.pdf> & <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.9416> (13.12.2018)
- PILOS, Spyridon (2016). "How can public services benefit from the CEF.AT platform". *ELRC Workshop Report*, Copenhagen 2016. <https://web.archive.org/web/20181219120940/http://lr-coordination.eu/sites/default/files/Denmark/20160307%20-%20ELRC%20benefits%20SpyrosK.pdf> (13.12.2018)
- QT21 CONSORTIUM (2017). "QT21 - Quality Translation 21. Introduction. Achievements. Video". Kaiserslautern: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI): <https://web.archive.org/web/20181219115552/http://www.qt21.eu/> (13.12.2018)
- REHM, Georg (2013). "Europe's Languages in the Digital Age: Multilingual Technologies for overcoming Language Barriers and Preventing Digital Language Extinction". *Workshop "State of the Art of Machine Translation – Current Challenges and Future Opportunities"*, (STOA). Brussels: European Parliament. <http://docplayer.gr/3524920-Europe-s-languages-in-the-digital-age.html> (13.12.2018)
- REHM, Georg / USZKOREIT, Hans / DAGAN, Ido et al. (2014a). "An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age"". In: PRETORIUS, Laurette / SORIA, Claudia / BARONI, Paola (eds.). *Proceedings of the Workshop on "Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era" (CCURL 2014)*. European Language Resources Association (ELRA). Pisa: Istituto di Linguistica Computazionale. 30-37. <https://web.archive.org/web/20181219125108/http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014-Workshop-CCURL2014-Proceedings.pdf> (13.12.2018)
- REHM, Georg / USKOREIT, Hans / DAGAN, Ido et al., (2014b). "An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age"". Poster presented at the *LREC 2014 Workshop on "Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era" (CCURL 2014)*. [https://web.archive.org/web/20181219124939/http://www.ilc.cnr.it/ccurl2014/LREC2014-Workshops CCURL2014-Poster-1 Presentation.pdf](https://web.archive.org/web/20181219124939/http://www.ilc.cnr.it/ccurl2014/LREC2014-Workshops%20CCURL2014-Poster-1%20Presentation.pdf) (13.12.2018)
- REHM, Georg / USZKOREIT, Hans / ANANIADOU, Sophia et al. (2016). "The strategic impact of META-NET on the regional, national and international level". *Language Resources & Evaluation* 50 (2): 351-374. <https://web.archive.org/web/20181219125645/https://helda.helsinki.fi/bitstream/handle/10138/136267/METAImpact.pdf?sequence=1&isAllowed=y>. <https://doi.org/10.1007/s10579-015-9333-4> (13.12.2018).
- SAVOUREL, Yves (2018). "TMX 1.4b Specification" (TMX - Translation Memory eXchange). Seattle: Globalization and Localization Association GALA. <https://web.archive.org/web/20181219120725/https://www.gala-global.org/tmx-14b> (13.12.2018)
- SAVOUREL, Yves / REID, John / JEWUSHENKO, Tony / RAYA, Rodolfo M. (2018). "OASIS - XLIFF Version 1.2" (XLIFF - XML Localization Interchange File Format). n.p. OASIS. <https://web.archive.org/web/20181219120830/http://docs.oasis-open.org/xliff/xliff-core/v2.1/xliff-core-v2.1.html> (13.12.2018)
- SCHLÜTER, Patrick (2018). "Statistics on the DGT-Translation Memory (DGT-TM)". Luxembourg: Directorate-General for Translation (DGT). https://web.archive.org/web/20181219121331/https://wt-public.emm4u.eu/Resources/DGT-TM_Statistics.pdf (13.12.2018)

- SOCHER, Richard (2016). "CS224d. Deep NLP. Lecture 8: Recap, Projects and Fancy Recurrent Neural Networks for Machine Translation". Stanford: Stanford University.
<https://web.archive.org/web/20181220095702/https://cs224d.stanford.edu/lectures/CS224d-Lecture9.pdf> (13.12.2018)
- STEINBERGER, Ralf / EBRAHIM, Mohamed / POULIS, Alexandros et al. (2014). "An overview of the European Union's highly multilingual parallel corpora". *Language Resources & Evaluation* 48 (4): 679-707. https://web.archive.org/web/20181220100033/https://ec.europa.eu/jrc/sites/jrcsh/files/2014_08_LRE-Journal_JRC-Linguistic-Resources_Manuscript.pdf. <https://doi.org/10.1007/s10579-014-9277-0> (13.12.2018)
- USZKOREIT, Hans (ed.) (2012a). "At Least 21 European Languages in Danger of Digital Extinction". In: *META-NET White Paper Series: Press Release*. Kaiserslautern: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI). <https://web.archive.org/web/20181219124842/http://www.meta-net.eu/whitepapers/press-release> (13.12.2018)
- USZKOREIT, Hans (ed.) (2012b). "META-NET White Paper Series: Key Results and Cross-Language Comparison". In: *META-NET White Paper Series*. Kaiserslautern: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI). <https://web.archive.org/web/20181219124131/http://www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison> (13.12.2018)
- TILDE COMPANY (2018a). "Deep Learning - European project QT21 tops international Machine Translation competition for the second year running". Riga, Vilnius, Tallinn. <https://web.archive.org/web/20181219130547/https://tilde.com/news/deep-learning-european-project-qt21-tops-international-machine-translation-competition-second> (13.12.2018)
- TILDE COMPANY (2018b). "Official website of the 2018 EU Council Presidency integrates Tilde's Neural MT service". Riga, Vilnius, Tallinn. <https://web.archive.org/web/20181220100202/https://www.tilde.com/news/official-website-2018-eu-council-presidency-integrates-tildes-neural-mt-service> (13.12.2018)
- TILDE COMPANY (2018c). "Language technologies for a connected world". Riga, Vilnius, Tallinn. <https://web.archive.org/web/20181219130753/https://www.tilde.com/> (13.12.2018)
- VEEN, Fjodor van (2016). "A mostly complete chart of Neural Networks". In: *The Neural Network Zoo*. Utrecht: The Asimov Institute. <https://web.archive.org/web/20181219122429/http://www.asimovinstitute.org/neural-network-zoo/> (13.12.2018)