Comparative Legilinguistics
vol. 2022/52
DOI: http://dx.doi.org/10.14746/cl.52.2022.14

Attention mechanism and skip-gram embedded phrases: short and long-distance dependency n-grams for legal corpora

PANAGIOTIS G. KRIMPAS, Tenured Associate Professor
Faculty of Classics and Humanities
Democritus University of Thrace, Greece
P. Tsaldari 1 Str., Komotini 69100
pkrimpas@bscc.duth.gr

ORCID: https://orcid.org/0000-0001-7271-9653

CHRISTINA VALAVANI, Phd
Journalism Computational Linguistics Lab (JCL Lab)
National and Kapodistrian University of Athens, Greece
Panepistimiou 30 Str., Athens 10679
cvalavani@hotmail.com

ORCID: https://orcid.org/0000-0002-2944-0734

Abstract: This article examines common translation errors that occur in the translation of legal texts. In particular, it focuses on how German texts containing legal terminology are rendered into Modern Greek by the Google translation machine. Our case study is the Google-assisted translation of the original (German) version of the Constitution of the Federal Republic of Germany into Modern Greek. A training method is proposed for phrase extraction based on the occurrence frequency, which goes through the Skip-gram algorithm to be then integrated into the Self Attention Mechanism proposed by Vaswani et al. (2017) in order to minimise human effort and contribute to the development of a robust machine translation system for multi-word legal terms and special phrases. This Neural Machine Translation approach aims at developing vectorised phrases from large corpora and process them for translation. The research direction is to increase the in-domain training data set and enrich the vector dimension with more information for legal concepts (domain specific features).

Keywords: computational linguistics; legal terminology; legal translation; Neural Machine Translation; Self Attention Mechanism; short and long-distance dependency n-grams; skip-gram algorithm.

SELF ATTENTION ΚΑΙ ΦΡΑΣΕΙΣ ΠΟΥ ΕΝΣΩΜΑΤΩΝΟΝΤΑΙ ΣΤΟ SKIP-GRAM: Ν-ΓΡΑΜΜΑΤΑ ΣΕ ΚΟΝΤΙΝΗ ΚΑΙ ΜΑΚΡΙΝΗ ΕΞΑΡΤΗΣΗ ΓΙΑ ΣΩΜΑΤΑ ΝΟΜΙΚΩΝ ΚΕΙΜΕΝΩΝ

Περίληψη: Αυτό το άρθρο εξετάζει συνήθη μεταφραστικά σφάλματα που σημειώνονται κατά τη μετάφραση νομικών κειμένων. Ειδικότερα, εστιάζει στον τρόπο με τον οποίο αποδίδει στα νέα ελληνικά η μηχανή μετάφρασης της Google γερμανικά κείμενα που περιέχουν νομική ορολογία. Η μελέτη περίπτωσης που χρησιμοποιούμαι αφορά τη μετάφραση της πρωτότυπης (γερμανικής) εκδοχής του Συντάγματος της Ομοσπονδιακής Δημοκρατίας της Γερμανίας στα νέα ελληνικά, μέσω της μηχανής μετάφρασης της Google. Προτείνεται μια μέθοδος εκπαίδευσης για την εξαγωγή φράσεων βάσει της συχνότητας εμφάνισης τους, η οποία διέρχεται από τον αλγόριθμο Skip-gram για να ενσωματωθεί κατόπιν στον Μηχανισμό Αυτοπροσοχής (Self Attention Mechanism) των Vaswani et al. (2017), προκειμένου να ελαχιστοποιήσει την ανθρώπινη προσπάθεια και να συμβάλει στην ανάπτυξη ενός ισχυρού συστήματος μηχανικής μετάφρασης για πολυλεκτικούς νομικούς όρους και ειδικές φράσεις. Αυτή η προσέγγιση, στο πλαίσιο της Νευρωνικής Μηχανικής Μετάφρασης, αποσκοπεί να αναπτύξει διανυσματοποιημένες φράσεις από μεγάλα σώματα και να τις επεξεργαστεί με στόχο τη μετάφραση. Η έρευνά μας κατευθύνεται προς την αύξηση των συνολικών δεδομένων εκπαίδευσης εντός δεδομένου θεματικού πεδίου και να

εμπλουτίσει τη διανυσματοποιημένη διάσταση με περισσότερες πληροφορίες για νομικές έννοιες (ιδιαίτερα χαρακτηριστικά του θεματικού πεδίου).

Λέξεις-κλειδιά: αλγόριθμος skip-gram; ν-γράμματα σε κοντινή και μακρινή εξάρτηση; νευρωνική μηχανική μετάφραση; νομική μετάφραση; νομική ορολογία; self attention mechanism; υπολογιστική γλωσσολογία.

SELF-ATTENTION-MECHANISMUS UND SKIP-GRAM EINGEBETTETE PHRASEN: N-GRAMME IN NAH- UND FERNABHÄNGIGKEIT FÜR RECHTSKORPORA

Zusammenfassung: Dieser Artikel untersucht häufige Übersetzungsfehler, die bei der Übersetzung von Rechtstexten auftreten. Insbesondere geht es darum, wie deutsche Texte mit juristischer Terminologie von der Google-Übersetzungsmaschine ins Neugriechische übertragen werden. Unsere Fallstudie ist die Google-gestützte Übersetzung der deutschen (originalen) Fassung der Verfassung der Bundesrepublik Deutschland ins Neugriechische. Für die Extraktion von häufigen Phrasen wird eine Trainingsmethode vorgeschlagen, die den Skip-Gram-Algorithmus durchläuft und wird dann in den von Vaswani et al. (2017) vorgestellten Selbstaufmerksamkeitsmechanismus integriert, um den menschlichen Aufwand zu minimieren und zur Entwicklung eines robusten maschinellen Übersetzungssystems für Mehrwortrechtstermini und -phrasen beizutragen. Dieser Ansatz der neuronalen maschinellen Übersetzung zielt darauf ab, vektorisierte Phrasen aus großen Korpora zu entwickeln und sie zur Übersetzung zu verarbeiten. Unsere Forschungsrichtung besteht darin, den domäneninternen Trainingsdatensatz zu erweitern und die Vektordimension mit mehr Informationen um Rechtskonzepte (domänenspezifische Merkmale) anzureichern.

Schlüsselwörter: Computerlinguistik; juristische Terminologie; juristische Übersetzung; n-Gramme von Nah- und Fernabhängigkeiten; neuronale maschinelle Übersetzung; Self-Attention-Mechanismus; Skip-Gramm-Algorithmus.

1. Introduction

This article examines common errors that occur when machine-translating legal texts. In particular, we discuss legalese translated from German into Modern Greek with focus on how the Google

Translation™ machine renders German legal terms, term elements, appellations and special phrases (hereinafter: legal language units) into Modern Greek at sentence level. Our case study is a Google-translated Modern Greek version of the German (original) text of the Constitution of the Federal Republic of Germany, the so-called *Grundgesetz für die Bundesrepublik Deutschland* (Bundesministerium der Justiz, Bundesamts für Justiz, http://www.gesetze-im-internet.de/gg/GG.pdf), hereinafter referred to as the GBD Parallel Text Corpus. We test the accuracy of sentence-level Google translation by comparing, in two adjacent columns, the target text sentence with the source text sentence generated by Google Translate in order to detect specific error types whose study may lead to steps and suggestions for a more sensitive Machine Translation. The errors of the automatic tool were identified within the corpus and were then analysed and classified according to specific criteria (see below in this unit). As explained by Stanisław Goźdź-Roszkowski (2021: 1524):

> "The influence of corpus linguistics methodology on how legal phraseology has been investigated extends beyond technological advances in text processing. Rather, corpus linguistics phraseology has paved the way for new and innovative studies which have begun to reveal the potential for investigating various roles and functions performed by different multi-word units in legal discourse."

Our study falls under what is called corpus-driven approaches (Tognini Bonelli 2001: 84–100), given that we make no prior assumptions and our source of information is the corpus itself (Goźdź-Roszkowski 2021: 1517). The utility of web legal resources as legal corpora has been discussed and supported by various scholars (e.g. Giampieri 2018), while the same is true of the relevance of machine-translation research for legal translation, especially for paedagogical purposes (e.g. Wiesmann 2019).

Since the 1990s there has been a shift from the dominant rule-based methods to statistical approaches. Following this background, deep learning goes further down, and gradually becomes the de facto technique of the mainstream statistical landscape (Liu et al. 2017). Neural Machine Translation (Kalchbrenner and Blunsom 2013) has demonstrated impressive performance in recent years. In this article we propose a training method for multi-word legal language unit extraction that goes through the Skip-gram algorithm (Mikolov et al. 2013) to be then integrated into the Self Attention Mechanism

(Vaswani et al. 2017); this process can minimise human effort and contribute to the development of a robust machine translation system for multi-word units. The aim of this Neural Machine Translation approach is to develop vectorised phrases from large corpora and process them in a way novel for translation.

After a thorough review of the machine translation output, some error types have been recorded, largely based on existing categorisations (Tezcan, Hoste and Mackel 2017; van Brussel, Tezcan, and Mackel 2018: 3800–3803; Krimpas 2017b: 79–96) by distinguishing between fluency and accuracy errors (Tezcan, Hoste and Mackel 2017). The error types recorded in our sample, adapted from Tezcan, Hoste and Mackel (2017), are as follows (our adaptations/additions appear in square brackets; the abbreviation TRM stands for 'term'):

Accuracy errors

- Mistranslation
  - Multi Word Expressions (MIS-MWE) [MIS-MWE-TRM]
  - POS
  - Sense (MIS-SE) [(MIS-SE-TRM)]
  - Mistranslation of verb tense and voice, number (nouns) (MIS-TVN)
  - Partial (MIS-PA)
  - Semantically unrelated (MIS-SU) [MIS-SU-TRM]
- Do not translate (DNT) (words have been translated unnecessarily e.g. for proper names)
- Untranslated (UT) [German Word/-s (GW)]
- Addition (AD) [AD-TRM]
  - Content Word
  - Function Word
- Omission (OM) [OM-TRM]
  - Content Word
  - Function Word

- Mechanical (non-meaning errors e.g. punctuation) (MECH)

Fluency errors

- Grammar (GR)
  - Word Form (WF)
  - Word Order (WO)
  - Extra Word(s) (EW)
  - Missing Word(s) (MW)
  - Multi Word Syntax (MWS)
- Lexicon (LEX)
  - Nonexistent (LEX-NE)
  - Lexical choice (LEX-CH) [LEX-CH-TRM] [Phrase Lexical Choice Term (PH-LEX-CH-TRM) and Partial Phrase Lexical Choice Term (PPH-LEX-CH-TRM)]
- Orthography (ORTH)
- Multiple errors (MULER)

To alleviate the translation errors documented in this research we propose a better, enriched version of the Skip-Gram and Self Attention mechanism proposed by Vaswani et al. (2018), where we modify the System so as to process legal and, in general, special multi-word units. In general, we make use of the Pointwise Mutual Information (PMI) (Bouma 2009) method (bigram extraction) (see Figures 1, 2, 3) and of the Short and Long-Distance Dependencies Extraction Algorithm (SLDDExAl) before inserting our words into the Skip-gram algorithm (Mikolov et al. 2013); then we insert the output vectors into the Self Attention Mechanism for more meaning, in which case the words become re-embedded. Attention is a concept that has helped improve the performance of NLP applications (Jay Alammar, jalammar.github.io, Visualizing machine learning one concept at a time, article posted June 27, 2018), including Machine Translation.

We also extend the Skip-gram model (Mikolov et al. 2013) by customising it to our needs. This paper largely incorporates a Self

Attention model that takes into account n-grams in relation to the predicted word. We begin with searching for bigrams, trigrams and tetragrams (hereinafter n-grams) to then embed them with Skip gram for better results (Mikolov et al. 2013). Then we train the System in the Self Attention Mechanism in order to develop vectorised n-grams (bigrams, trigrams, tetragrams in short and long-distance dependency).

In short, unit two presents all categories of errors in legal language units with examples extracted from the corpus. Their classification helps identify weak points of the translation tool with special focus on legal language units. Unit three discusses ways of automatic extraction of both bilects and multi-word units in short and long distance dependency. In unit four Skip-gram training takes place in order to obtain embeddings with more meaning. Unit five describes the attention mechanism with the new elements and how exactly the n-grams are integrated for correct translation purposes. Unit six records the proposed mechanism's steps as well as new proposals. Unit seven summarises by presenting advantages and disadvantages of our proposal.

2. Error documentation

The difficult-to-process character of legal language units often results in pronounced discrepancies in both human and machine translation, the DE > EL language pair being no exception.

To reflect the actual will of the legislator it is vital for the legal terminology used in the target language (TL) to cover the same conceptual areas as the source text (ST). In practice, however, the attempt to find legally equivalent terms is not always straightforward due to the asymmetry of legal systems (Dubě̌da 2021: 61, 68, 69; Prieto Ramos 2021: 175–176), even if they belong to the same family of law, as is the case with the Greek and the German ones. At times the asymmetry can be purely terminological-semantic rather than conceptual, but this can be equally problematic for the legal translator (Krimpas 2017a).

Tables 1–27 below show examples of n-grams, some of which are interdependent with other, correctly translated units in the sentence, while others are semantically mistranslated independently of

context. Context is included whenever appropriate. All examples given below are taken from the aforementioned GDB Parallel Text Corpus (see unit 1). This corpus comprises approximately 5,000 pairs of sentences, whose translation into Greek was carried out at the sentence level by Google Translate; approximately 300 sentence pairs out of them were translated by both Google Translate and a human translator; the examples below come from this particular subset. The tables show sentence parts that are essential to illustrate machine translation errors; whole sentences are given only when necessary. Units involved in one or more machine translation errors were manually extracted. Underlined text in the first row of each table (source text) shows translational correspondence with underlined text in the third row of each table (target text), as a way to highlight text involved in the machine translation error. The second row of each table shows the machine-translated target text, accompanied by the error code (see unit 1). In cases of clear correspondence between the first and third row there is no underlining.

Tables 1–11: Examples of unigrams involved in context-dependent errors

1.

Source text:	Grundgesetz für die Bundesrepublik Deutschland
Google translation:	Βασικός νόμος (MIS-SE-TRM)
Correct rendering:	**Σύνταγμα**

2.

Source text:	(weggefallen)
Google translation:	(εγκαταλείφθηκε) (MIS-SE-TRM)
Correct rendering:	**(καταργήθηκε)**

3.	Source text:	Jeder Deutsche hat in jedem Lande die gleichen staatsbürgerlichen Rechte und Pflichten.
	Google translation:	χώρα (MIS-SE-TRM)
	Correct rendering:	**ομόσπονδο κρατίδιο**

4.	Source text:	….soweit der Bundesrat ihm zustimmt.
	Google translation:	Bundesrat […]. (GW(s)+ MIS-SE-TRM)
	Correct rendering:	**Ομοσπονδιακό Συμβούλιο**

5.	Source text:	den Wasserhaushalt
	Google translation:	ισοζύγιο νερού (MIS-SE-TRM)
	Correct rendering:	**διαχείριση (των) υδάτινων πόρων**

6.	Source text:	[…] zwei Jahren nach der Durchführung der Volksbefragung ein […]
	Google translation:	πραγματοποίηση του δημοψηφίσματος (LEX-CH-TRM)
	Correct rendering:	**διεξαγωγή του δημοψηφίσματος**

7.	Source text:	Oberster Gerichtshof für die in Absatz 1 und 2 genannten Gerichte ist der Bundesgerichtshof.
	Google translation:	Ομοσπονδιακό Δικαστήριο (OM-TRM)
	Correct rendering:	**Ομοσπονδιακό Ακυρωτικό Δικαστήριο**

8.	Source text:	Kunst und Wissenschaft, Forschung und Lehre sind frei.
	Google translation:	δωρεάν (LEX-CH-TRM)
	Correct rendering:	**ελεύθερες**

9.	Source text:	Ihre Gründung ist frei.
	Google translation:	Η εγκατάσταση σας (LEX-CH-TRM)
	Correct rendering:	**Η ίδρυσή της**

10.	Source text:	(Vollzitat:)
	Google translation:	(Πλήρες απόσπασμα:) (PPH-LEX-CH-TRM)
	Correct rendering:	(**πλήρες παράθεμα:**)

11.	Source text:	Bundesrecht bricht Landesrecht. *(a context-independent unigrams)*
	Google translation:	Ο ομοσπονδιακός νόμος παραβιάζει τον κρατικό νόμο. (MIS-MWE-TRM)
	Correct rendering:	**Το ομοσπονδιακό δίκαιο παραβιάζει το πολιτειακό δίκαιο**

Tables 12–19: Examples of bigrams involved in short-distance dependency errors

12.	Source text:	daß ein billiger Ausgleich erzielt,
	Google translation:	φθηνή αποζημίωση (MIS-MWE-TRM)
	Correct rendering:	**εύλογο συμψηφισμό**

13.	Source text:	Die konkurrierende Gesetzgebung erstreckt sich auf folgende Gebiete
	Google translation:	ανταγωνιστική νομοθεσία (MIS-MWE-TRM)
	Correct rendering:	**συντρέχουσα νομοθετική αρμοδιότητα**

14.	Source text:	Den unehelichen Kindern sind durch die Gesetzgebung die gleichen Bedingungen für ihre leibliche und seelische Entwicklung und ihre Stellung in der Gesellschaft zu schaffen wie den ehelichen Kindern.
	Google translation:	[…] παράνομα παιδιά […] νόμιμα παιδιά. (MIS-MWE-TRM)
	Correct rendering:	**[…] τέκνα εκτός γάμου […] τέκνα γεννημένα σε γάμο**

15.	Source text:	(+++ Textnachweis Geltung ab: 14.12.1976 +++)
	Google translation:	(+++ Η απόδειξη ισχύει από: 14.12.1976 +++) (MIS-MWE-TRM)
	Correct rendering:	**Η παρούσα εκδοχή τέθηκε σε ισχύ από: 14.12.1976 +++**

16.	Source text:	Zur Wahrung der Einheitlichkeit der Rechtsprechung ist ein Gemeinsamer Senat der in Absatz 1 genannten Gerichte zu bilden.
	Google translation:	[…] ομοιομορφία της νομολογίας, […] μεικτή σύγκρουση […] (MIS-MWE-TRM + LEX-CH-TRM)
	Correct rendering:	**[…] ενιαίου χαρακτήρα της νομολογίας […] Μείζων Ολομέλεια**

17.	Source text:	Sie soll hierbei ihre Auffassung darlegen.
	Google translation:	[…] εξηγήσει την άποψή της. (MIS-MWE-TRM and LEX-CH-TRM).
	Correct rendering:	**[…] καταθέτει τη γνώμη της.**

18.	Source text:	…..hat sich das Deutsche Volk kraft seiner verfassungsgebenden Gewalt dieses Grundgesetz gegeben.
	Google translation:	λόγω των συστατικών του δυνάμεων (MIS-MWE-TRM)
	Correct rendering:	**δυνάμει της συντακτικής του εξουσίας**

19.	Source text:	Zwischen dem Antrage und der Wahl müssen achtundvierzig Stunden liegen.
	Google translation:	να υπάρχουν (MIS-MWE-TRM)
	Correct rendering:	**να μεσολαβούν**

Tables 20–27: Examples of phrases or sentences with words involved in short- and long-distance dependency errors

20.	Source text:	Der Verlust der Staatsangehörigkeit darf nur auf Grund eines Gesetzes und gegen den Willen des Betroffenen nur dann eintreten […] (LEX-CH-TRM)
	Google translation:	Η απώλεια της ιθαγένειας μπορεί να συμβεί […] (NRTDNL)
	Correct rendering:	**Η απώλεια της ιθαγένειας επέρχεται […]**

21.	Source text:	Im Falle eines vorsätzlichen Verstoßes kann auf Entlassung erkannt werden.
	Google translation:	η απόλυση μπορεί να αναγνωριστεί. (MIS-MWE-TRM)
	Correct rendering:	**μπορεί να τεθεί σε διαθεσιμότητα.**

22.	Source text:	Frauen vom vollendeten achtzehnten bis zum vollendeten fünfundfünfzigsten Lebensjahr
	Google translation:	από την ηλικία των δεκαοκτώ έως την ηλικία των πενήντα – Πέμπτο. (MIS-MWE-TRM and EW(s))
	Correct rendering:	**με συμπληρωμένη ηλικία από δεκαοκτώ έως πενήντα ετών**

23.	Source text:	Ihre hauptamtlichen Richter müssen die Befähigung zum Richteramt haben.
	Google translation:	Οι κριτές πλήρους απασχόλησης πρέπει να είναι κατάλληλοι για να υπηρετούν ως κριτές. (MIS-MWE-TRM)
	Correct rendering:	**Οι δικαστές πλήρους απασχόλησης πρέπει να έχουν την ικανότητα ανάληψης του δικαστικού λειτουργήματος.**

24.	Source text:	Wehrpflichtige, die nicht zu einem Dienst nach Absatz 1 oder 2 herangezogen sind
	Google translation:	Στρατιωτικοί που δεν εμπλέκονται σε υπηρεσία (MIS-MWE-TRM)
	Correct rendering:	**Επίστρατοι/Κληρωτοί που δεν καλούνται να υπηρετήσουν τη θητεία τους**

25.	Source text:	Die Verwirkung und ihr Ausmaß werden durch das Bundesverfassungsgericht ausgesprochen.
	Google translation:	Η κατάπτωση και η έκτασή της θα εκφραστεί από το Ομοσπονδιακό Συνταγματικό Δικαστήριο. (MIS-MWE-TRM)
	Correct rendering:	**Η έκπτωση δικαιώματος και ο βαθμός της αποφασίζονται από το Ομοσπονδιακό Συνταγματικό Δικαστήριο.**

26.	Source text:	Gesetzesvorlagen werden beim Bundestage durch die Bundesregierung, aus der Mitte des Bundestages oder durch den Bundesrat eingebracht
	Google translation:	Οι λογαριασμοί εισάγονται στο Bundestag από την Ομοσπονδιακή Κυβέρνηση, από το κέντρο του Bundestag ή από το Bundesrat. (LEX-CH-TRM and UT-GW(s))
	Correct rendering:	**Τα νομοσχέδια εισάγονται στο Ομοσπονδιακό Κοινοβούλιο από την Ομοσπονδιακή Κυβέρνηση, από το βήμα του Ομοσπονδιακού Κοινοβουλίου ή από το Ομοσπονδιακό Συμβούλιο.**

27.	Source text:	Gewählt ist, wer die Stimmen der Mehrheit der Mitglieder des Bundestages auf sich vereinigt.
	Google translation:	ποιος έχει την πλειοψηφία των μελών του Bundestag (UT-GW(s) and LEX-CH-TRM)
	Correct rendering:	**ποιος συγκεντρώνει την πλειοψηφία των μελών του Ομοσπονδιακού Κοινοβουλίου.**

It should be borne in mind that in such cases translation quality depends on the available pool of German legal texts having been translated so far into Modern Greek (in this case not as many) by the machine translator (in this case Google Translate), as well as on the model used by the latter. An additional factor is whether the text to be translated will be machine-translated at the word, sentence (as is here the case) or whole text level.

A closer look at the above data suggests that areas where syntax interference is detected show a high correlation with wrong syntactic interpretation errors, occurring at the stage of decoding or recoding of the syntactic structure of the text (we code them,

respectively, as MWS, WO), while the areas where lexical interference is detected are mainly associated with errors in decoding or recoding of the semantics of the text (LEX-CH-TRM, MIS-SE, MIS-SU).

There are also numerous language register-related errors in the test text (we code them as LEX-CH-TRM); in such cases the texts have been semantically, grammatically and syntactically correctly rendered in Greek, but there is a register discrepancy with respect to Greek legalese. Moreover, Modern Greek LSP is notorious for its heavy learned element, which poses lexical, phonotactic, morphological and syntactic difficulties even for native speakers (Valeontīs and Krimpas 2014: 49–54; cf. Krimpas 2019), a fact reflected also in the translation of LSP texts, which becomes particularly hard especially in institutional thematic areas (Valeontīs and Krimpas 2014: 21) such as law, economics, religion etc. This linguistic landscape favours translation errors due to 'non-recognition of text-specific deviations from normal language usage' (Wiesmann 2019: 137) (an error type that we code as MIS-SE, MIS-SU).

Finally, at various places in the test text there is some indication of probable syntax interference of the source language. In such cases, the meaning from the source text is not transferred to the target text, a translation error that we code as MWS, WO.

In the above cases the difficulty of transferring the exact meaning of the source language (German) into the target language (Modern Greek) is obvious, either because the relative concept does not exist in the target language, or because of intersystemic differences, which often lead to wrong term choice or even non-translation; such cases are coded as LEX-CH-TRM.

3. Pre-training of the corpus and n-grams mining

Mining n-grams is the automatic extraction of frequent phrases (Del, Tättar, and Fishel 2018), such as multi-word terms and special phrases, from a corpus. First, we POS tag, parse (syntactic dependencies) (Klemen, Krsnik, and Robnik-Šikonja 2022), lemmatise and tokenise the whole corpus and then extract bigrams, trigrams and tetragrams, hereinafter referred to as n-grams (verbs,

nouns, adverbs, adjectives, participles and prepositions) to subsequently take them as input into Word2Vec, in particular into the Skip-gram algorithm, which generates vectorised words of high dimensionality (Camacho-Collados and Pilehvar 2018) with more meaning (see Figure 9). The threshold for the n-grams will be high, so that high quality legal LSP words (especially with short- and long-distance dependencies), phrases are extracted.

The mechanism for extracting frequent n-grams is as follows: If x and y represent bigrams in the legal corpus, y follows x. Whenever x and y appear together many times, the Pointwise Mutual Information (PMI) (Bouma 2009) will have a high value (see Figures 1, 2, 3), while it will have a value of 0 if x and y are completely independent, i.e. if they appear in different sentences (Moshe Hazoom, Towards Data Science, article posted December 22, 2018). This can be extended to three or four words e.g. a tetragram [a, b, x, y] could collocate in a document by using the Short and Long-Distance Dependency Extraction Algorithm (since the PMI formula is tailored for pairs and combinations of two items).

$$E \in R^{|V| x d}$$

$$|V| = vocabulary\ size$$

Figure 1: Embedding matrix after Word2Vec training (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

$$PMI(x;y) = \log \frac{p(x,y)}{p(x)p(y)}$$

Figure 2: Pointwise Mutual Information (PMI). PMI of concrete occurrences of x and y (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

PMI helps us find bigrams in order to build phrase vectors and embed them (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

$$NPMI(x;y) = \frac{\log \frac{p(x,y)}{p(x)p(y)}}{-\log p(x,y)}$$

Figure 3: Normalised Pointwise Mutual Information of x and y (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

Additionally, an n-gram could often co-occur in a sentence but in a long-distance dependency (see Figure 8). With the aim of being able to extract words that are also in a long- distance dependency but syntactically related, we built a simple algorithm, the Short and Long Distance-Dependency Extraction Algorithm (SLDDExAl). First, the corpus is processed and more specifically parsed, lemmatised, all stop words are removed and all tokens per sentence are collected (from within each sentence). Lemmata remain within the sentences they belong to. Co-occurrence counting is only done at parser-defined sentence boundaries. For example, if the word x co-occurs with the word y in sentence s1, then this is registered by the algorithm. Every sentence of the corpus where the word x appears is checked. If, at a later point, the word x appears again in another sentence along with the word y, then the algorithm adds this information to the frequency count list.

In case the word y appears in sentences without the word x, then the word y is not counted since we are only interested in its co-occurrence with the word x. This process is repeated/iterated for all lemmata in the corpus. Thus it is established which words frequently co-appear at any distance within a sentence, while at the same time their syntactical relation is detected by the parser. Some issues may arise with respect to the automatic translation of eventual out-of-domain parts of the corpus.

Figure 4: Examples of frequent n-grams for embedding them with one vector.

In particular, our system builds one vector for n-grams that collocate (with high occurrence frequency) (see Figure 4). In a parsed text, the system is able to know, inter alia, which words are related to each other and focuses on extracting them as they often co-occur in a sentence. We are not looking in the text for words with a specific syntactical relationship between them, e.g. verb-subject, but for frequently co-occurrence words that may have any syntactical relationship with each other. For the aforementioned reasons we don't use the two generalizations for multivariate distributions of Pointwise Mutual Information, presented by Tim Van de Cruys (2011).

4. N-gram embedding with Skip-gram

Word2vec can be applied to a big amount of data and Skip-gram (Mikolov et al. 2013) is one of the unsupervised learning techniques (it can work on any raw text) used to find the most relative words for a given word (Mikolov et al. 2013), especially with infrequent words.

Skip-gram predicts the context words from the target word and -in our proposed approach- it can learn legal concept embeddings from different data sources, including journals and legal narratives. Creating representations for legal concepts by training the System with legal corpora is highly recommended (domain adaption) (Diniz da Costa et al. 2022).

In the output vector of Skip-gram there is semantic information and representation of the relation between words, which is not the case for one-hot representations. Then those n-grams are inserted into Word2Vec (Mikolov et al. 2013) to be trained in the Skip-gram algorithm (Mikolov et al. 2013). The algorithm will take as input one-hot encodings which represent n-grams but will process the collocating n-grams as one vector and the co-occurring ones simultaneously. The architecture of Skip-gram is presented in Figure 5 below:

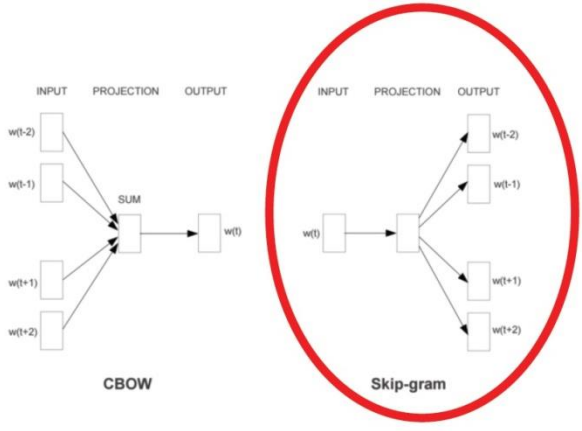

Figure 5: The Skip-gram model architecture (Source: https://arxiv.org/pdf/1301.3781.pdf Mikolov el al., 2013)

Skip-gram is a simple Neural Network with only one hidden layer (Mikolov et al. 2013). The input to the network is a one-hot encoded vector representation of a target-word; all of its dimensions are set to zero, apart from the dimension corresponding to the target-word (one-hot representation). The output is the probability distribution over all

words in the vocabulary, which defines the likelihood of a word being selected as the input word's context (Paula, Cambridge Spark, article posted November 9, 2018). Figure 6 below illustrates the Skip-gram model in more detail.

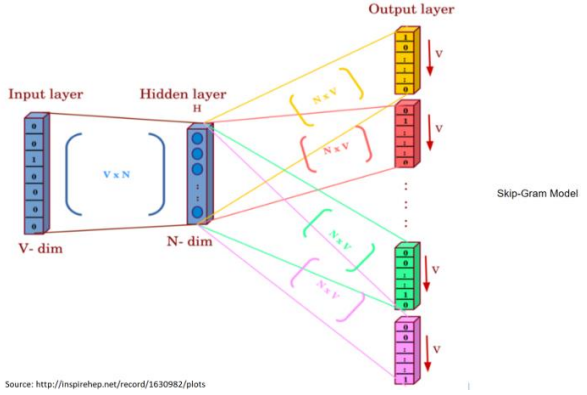

Figure 6: Skip-gram model architecture in detail (Doshi 2019).

With the SLDDExAl we can extract n-grams and use them as one-word (see Figure 4) vector for collocations and as multiple-word vectors for short and long-distance dependencies (depending on the number of words) (see Figure 8). This happens additionally to the one-word vectors vocabulary. When two words appear at a distance then they will be extracted as such and inserted together into the Skip-Gram with a gap, e.g. "Verlust […] eintreten" will be inserted as a unit into the Skip-Gram. Skip-Gram will train the vectors by simultaneously setting a window of two words left and right of each word, the two words will be trained in the shared context of the words. If four words do not appear between the two, then the system automatically shrinks the window and adapts to those that exist. Moreover, when a gap stands for more words, the system is still trained on the basis of a two-word window, as is usual with Skip-Gram.

Window1	Window 2	Verlust	GAP (of unknown size)	eintreten	Window 1	Window 2

Figure 7: Long-distance dependency legal term inserted in Skip-Gram.

After training in Skip-gram we have the embeddings with more targeted meaning.

5. Self-attention mechanism

Most successful approaches to machine translation (e.g. Wu, Zhao, and Li 2020; Bahdanau, Cho, and Bengio 2016; Vaswani et al. 2017; Gehring et al. 2017) rely on the availability of parallel corpora. Supervised Neural Machine Translation (NMT) (Kalchbrenner and Blunsom 2013) works with the encoder- decoder, where the encoder reads the source sentence and produces its representation, which is then fed to the decoder in order to generate the target sentence word-by-word (Del, Tättar, and Fishel 2018) (see Figure 9). Cross entropy loss is usually used as a training objective and beam search algorithm is used for inference (Del, Tättar, and Fishel 2018). Such neural models rely on vast amounts of parallel data. We employ the Self Attention Mechanism as presented in Vaswani et al. (2017). The closer the vectors are, the bigger the dot product is. By computing the cosine similarity we find the similarity between vectors, and we can also measure the Euclidean distance d for it.

5.1. Why self-attention in legal language

With respect to legal language units long memory might be regarded as not required by a system of Neural Machine Translation, since they can be one-word (simple or complex) terms, multi-word terms or other multi-word (context-conditioned or fixed) special phrases. Legal language, however, is pretty demanding as such and if ones wishes to structure a mechanism that translates correctly while maintaining the

language register and rendering it accordingly to the target language, then the Self Attention Mechanism is quite appropriate. For example, it is worth considering the German sentence: 'Der Verlust der Staatsangehörigkeit darf [...] eintreten', which was Google-translated as 'Η απώλεια της ιθαγένειας μπορεί να συμβεί'. Although the translation is perfectly understandable even with the general-language verbal phrase μπορεί να συμβεί 'can/may happen', the expected wording in Modern Greek legal language would require the LSP verb επέρχεται 'takes place'. In this case the participle is semantically bound to the legal language. High attention is necessary, especially due to the long-distance dependency between the two words.

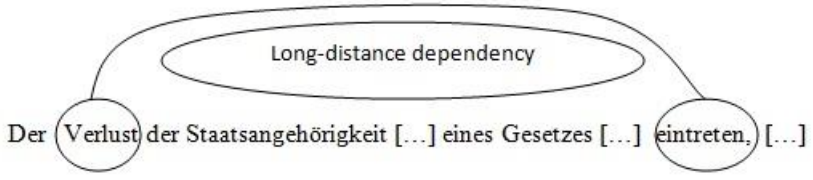

Figure 8: Long-distance dependency unigrams.

Το δικαίωμα […] εγγυημένο.

Example 1: A legal phrase showing that high attention is required due to the long dependency between the words.

In this case the two unigrams will be simultaneously processed by Skip-Gram (see Figure 7), while the created vectors of the interdependent n-grams will result only from a context where such n-grams co-occur. The numerical representation of such co-occurring units will contain information of their shared context; the same will be done in the target language during training.

5.2. Self-attention with n-grams

Self-attention (Vaswani et al. 2017) is a sequence-to-sequence operation: a sequence of vectors goes in and a sequence of vectors

comes out. The input vectors are e.g. x1,2,…,xn and the corresponding output vectors are y1,y2,…,yn with a d dimension (Peter Bloem, peterbloem.nl, article posted August 18, 2019). N-gram embeddings are the mathematical expression of phrases and single units (unigrams) (Jurafsky and Martin 2022).

The Self Attention mechanism (see Figure 9) is applied unaltered, the only difference being that it accepts as input n-grams embedded with the Skip-gram algorithm. The processing of words so as to become vectors before being inserted into the Self Attention mechanism is performed for both the source and the target language. The mechanism enriched with the novel interventions and the exact points of the latter are illustrated below.

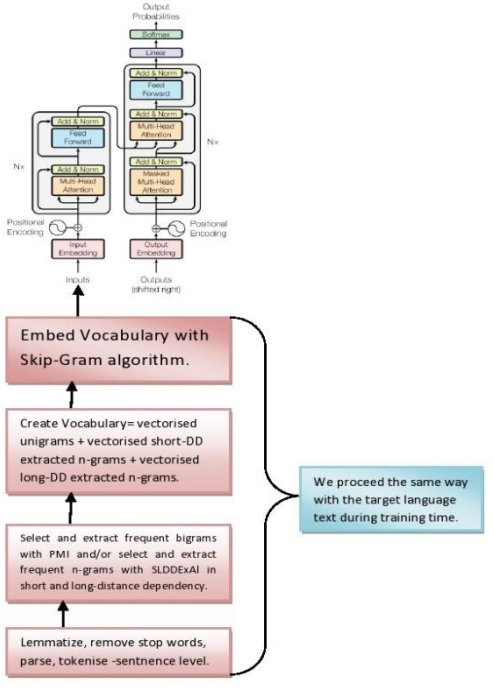

Figure 9: The proposed n-gram-extended Self Attention Mechanism (Vaswani et al. 2017) with Skip-gram, PMI (Bouma 2009) and SLDDExAl.

6. General implementation steps

After stating in detail the main points of the proposed mechanism and the new suggestions, we list below the implementation steps in their actual order, focusing on the proposed steps of the novel system. The main process of the system is the following:

(1) POS corpus annotation, lemmatisation, tokenisation syntactic parsing and stop word list.
(2) Creation of an algorithm for extracting short and long-distance dependencies (Short and Long-Distance Dependency Extraction Algorithm).
(3) Selection of frequent n-grams (bigrams, trigrams and tetragrams) from the corpus by using the Pointwise Mutual Information for bigrams (PMI) (Bouma 2009) method (Moshe Hazoom, Towards Data Science, article posted December 22, 2018) and/or SLDDExAl with a high threshold. The system selects only nouns, verbs, adverbs, adjectives, participles and prepositions.
(4) Limitation of selection to n-grams with sufficient frequency (Shang et al. 2018).
(5) Embedment of all the unique words as well as the n-grams (consisting of more words in short and long-distance dependencies) from the vocabulary $|V|$ with one-hot representation, with a dimension of $[1, |V|]$.
(6) Attribution of a single numerical representation to frequent lexical collocating units.
(7) Simultaneous insertion of frequent lexical co-occurring n-grams (in short and long-distance dependency) into Skip-gram and training thereof in their shared context.
(8) Creation of an n-gram vectorised vocabulary $|V|$.
(9) Integration of the one-hot representations of the n-grams into the Skip-gram algorithm (Mikolov et al. 2013).
(10) Use of the Skip-gram implementation from the Word2Vec model and of the Gensim library (Mikolov et al. 2013) to train embeddings.

The output of the Skip-gram consists of meaningful vectorised n-grams with only high frequency (Shang et al. 2018). The objective of the Skip-gram is to maximise P(Vtarget|Vsource), the probability of Vtarget being predicted as Vsource context for all training pairs in the corpus. The n-gram vectors are the input vectors for the Self Attention

Mechanism (Vaswani at al. 2017). This preparation is performed for both languages and the words of the source and target language (see Figure 9).

The main features of and requirements for our proposed model to be functional can be summarised as follows:

(1) Training with legal domain parallel corpora for performance improvement.
(2) Extension of the embedded words with features of legal content. This minimises ambiguity to the extent possible.
(3) Reliance on large corpora of legal domain containing hundreds of thousands of documents to help deliver superior performance (Shang et al. 2018).
(4) Phrase learning from an unsupervised text (Del, Tättar, and Fishel 2018).
(5) Domain-independence (it can support multiple domains) (Shang et al. 2018).
(6) Development of a single numeric representation for combining words (e.g. bigrams) (see Figure 4).
(7) Data-driven approach.
(8) Reward of frequent phrases, as frequency of the phrase occurrence is important; e.g., if "A B" is frequent, then "A B" is a phrase candidate.
(9) Choice of high frequency since PMI (Bouma 2009) and SLDDExAl can reflect the frequency counts rather than the quality of the phrases. The assumption is that if it appears in the corpus frequently, then it is a quality multi-word term/appellation/phrase.
(10) Careful choice of minimum threshold for the selected phrases (in a high rank) in order not to vectorise infrequent n-grams. The set limit for common phrases mainly depends on the size of the parallel Text Corpus and whether it is a domain corpus.
(11) Support of any language (language agnostic system).

(12) Improvement of the neural-based translation system capability by modelling both word and phrase (n-grams) (Del, Tättar, and Fishel 2018).
(13) Learning of phrase embedding by minimisation of the semantic distance between translation equivalents and maximisation of the semantic distance between non-translation pairs (Zhang et al. 2014).
(14) Introduction of the PMI (Bouma 2009) and SLDDExAl methods to generate phrase level memory in vector form.
(15) Embedment of both the source and the target phrase with the same vectors, having the same dimension. The idea for phrase embedding has been picked up from LASER (GitHub, Language-Agnostic Sentence Representations, updated July 6, 2022), where sentence embedding takes place.
(16) Re-embedment of the vectors for more meaning.
(17) Use of the Self Attention Mechanism as presented in Vaswani et al. (2017).

7. Pros and cons of the proposal

In this article, which attempts to serve as a proposal for improving legal translation at both the lexical and the structural level, we pull existing methods and techniques together in a new way. Admittedly, the main obstacle to the implementation of this novel proposal is the lack of large Modern Greek special text corpora, let alone parallel ones. Anyway, some advantages of our proposal are the following:

(1) It introduces LSDDExAl.
(2) Skip-gram processes frequent co-occurring vectorised words simultaneously and trains them in their shared context.
(3) It combines already existed methods and techniques in a novel way.
(4) It can be trained for any domain.
(5) It deals with long-distance dependencies.
(6) It is language agnostic.
(7) It focuses in particular on the correct translation of multi-word special phrases.
(8) It contributes to overcoming the previously known errors in Neural Machine Translation.
(9) It limits the post-editing errors

Beyond doubt, any model has also disadvantages, and the ones of our approach are probably the following:

(1) There is no benchmark available.
(2) The model requires more time and space for training in comparison with the original Self Attention Mechanism.
(3) In order for the performance to be enhanced, a vast amount of data of legal content is needed, especially bilingual parallel legal corpora for the German-Greek language pair.

8. Future work

In our proposed approach the existing phrase-mining potential is complemented (since units connected meaning-wise but distant syntax-wise are extracted as well) and the Pointwise Mutual Information (PMI) method (Bouma 2009) is presented; this method ranks the extracted phrases by their term frequency. We also extend the Self attention Mechanism (Vaswani et al. 2017) with the Skip-gram algorithm (Mikolov et al. 2013) and the embedded vectors by adding domain specific (legal) features in order to eliminate semantic ambiguities. Our method requires availability of domain parallel corpora. Our baseline system follows principles of the Self Attention Mechanism (Vaswani et al. 2017) where we integrate n-gram vectors. The vectors for n-grams are learned as individual vocabulary entries. Yet, the effect of legal phrase embeddings is still to be investigated. The research direction is to increase the in-domain training data set and enrich the vector dimension with more information for legal concepts (domain specific features).

CONTRIBUTION OF AUTHORS: Both authors have contributed equally to the article.

CONFLICT OF INTEREST: The authors declare that there is no conflict of interest.

Bibliography

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2016. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*. arXiv:1409.0473v7 [cs.CL]. DOI: https://doi.org/10.48550/arXiv.1409.0473.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5: 135–46. https://aclanthology.org/Q17-1010.pdf (accessed December 28, 2022).

Bouma, Gerlof. 2009. Normalized (Pointwise) Mutual information in collocation extraction. In *From Form to Meaning: Processing Texts Automatically: Proceedings of the Biennial GSCL Conference 2009*, eds. Christian Chiarcos, Richard Eckart de Castilho and Manfred Stede, 31–40. Tübingen: Gunter Narr.

Camacho-Collados, José, and Mohammad Taher Pilehvar. 2018. From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research* 63: 743–88. DOI: https://doi.org/10.1613/jair.1.11259.

Diniz da Costa, Alexandre, Mateus Coutinho Marim, Ely Edison da Silva Matos, and Tiago Timponi Torrent. 2022. Domain Adaptation in Neural Machine Translation using a Qualia-Enriched FrameNet. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, eds. Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk and Stelios Piperidis, 1–12. Paris: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2022/LREC-2022.pdf (accessed December 28, 2022).

Duběda, Tomáš. 2021. Direction-asymmetric equivalence in legal translation. *Comparative Legilinguistics* 47: 57–72. DOI: http://dx.doi.org/10.2478/cl-2021-0012.

Giampieri, Patrizia. 2018. The web as corpus and online corpora for legal translations. *Comparative Legilinguistics* 33: 35–55. DOI: http://dx.doi.org/10.14746/cl.2018.33.2.

Goźdź-Roszkowski, Stanisław. 2021. Corpus linguistics in legal discourse. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique* 34: 1515–1540. DOI: https://doi.org/10.1007/s11196-021-09860-8.

Jurafsky, Daniel, and James H. Martin. 2022. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd edition draft). https://web.stanford.edu/~jurafsky/slp3/ed3book_jan122022.pdf (accessed December 28, 2022).

Kalchbrenner, Nal, and Phil Blunsom. 2013. Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, 18-21 October 2013 (EMNLP 2013)*, 1700–09. Stroudsburg: Association for Computational Linguistics. https://aclanthology.org/D13-1176.pdf (accessed December 28, 2022).

Klemen, Matej, Luka Krsnik, and Marko Robnik-Šikonja. 2022. Enhancing deep neural networks with morphological information. *Natural Language Engineering* 28(3): 1–26. DOI: https://doi.org/10.1017/S1351324922000080.

Krimpas, Panagiotis G. 2017a. Terminological preciseness or translational and legal effectiveness? Terminology of commodatum in no > el language pair. In *Konteksty súdneho prekladu a tlmočenia VI*, ed. Zuzana Guldanová, 66–84. Bratislava: Univerzita Komenského v Bratislave. https://fphil.uniba.sk/fileadmin/fif/katedry_pracoviska/kgn/transius/na_stiahnutie/Kontexty_sudneho_prekladu_a_tlmocenia_VI_2017.pdf (accessed December 28, 2022).

Krimpas, Panagiōtīs G. 2017b. *Eisagōgī stī theōria tīs metafrasīs* [Introduction to Translation Theory]. Athīna: Grīgorī.

Krimpas, Panagiōtīs G. 2019. Pseudologioi typoi kai yperdiorthōsī stī Neoellīnikī Koinī me vasī ta epipeda glōssikīs analysīs [Pseudo-learned forms and hypercorrection in Standard Modern Greek on the basis of linguistic analysis levels]. In *Apo ton oiko sto spiti kai tanapalin… To logio epipedo stī sygchronī nea ellīnikī: Theōria, Istoria, Efarmogī* [From *oikos*

to *spiti* and vice versa: The learned register in Standard Modern Greek: Theory, History, Practice], eds. Asimakīs Fliatouras and Anna Anastasiadī-Symeōnidī, 57–126. Athīna: Patakī.

Maksym Del, Andre Tättar, and Mark Fishel. 2018. Phrase-based unsupervised machine translation with compositional phrase embeddings. In *Proceedings of the Third Conference on Machine Translation (WMT), Volume 2: Shared Task Papers, Belgium, Brussels, October 31 - Novermber 1, 2018, 361–67*. Stroudsburg: Association for Computational Linguistics. DOI: https://doi.org/10.18653/v1/W18-64034.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2014. Distributed representations of words and phrases and their compositionality. In *27th Annual Conference on Neural Information Processing Systems 2013, December 5-10, 2013 Lake Tahoe, Nevada, USA, Volume 1 of 4*, 3128–36. New York: Curran.

Prieto Ramos, Fernando. 2014. Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In *The Ashgate Handbook of Legal Translation*, eds. Le Cheng, King Kui Sin, and Anne Wagner, 121–134. Abingdon: Routledge.

Shang, Jingbo, Jialu Liu, Meng Jiang, Xiang Ren, Clare R. Voss, Jiawei Han. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering* 30(10): 1825–1837. DOI: https://doi.org/10.1109/TKDE.2018.2812203.

Tezcan, Arda, Véronique Hoste, and Lieve Macken. 2017. SCATE Taxonomy and Corpus of Machine Translation Errors. In *Trends in e-Tools and Resources for Translators and Interpreters*, eds. Gloria Corpas Pastor and Isabel Durán Muñoz, 219–248. Leiden: Brill. DOI: https://doi.org/10.1163/9789004351790_012.

Tognini Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins.

Valeontīs, Kōnstantinos E., and Panagiōtīs G. Krimpas. 2014. *Nomikī Glōssa, Nomikī Orologia: Theōria kai Praxī* [Legal Language, Legal Terminology: Theory and Practice]. Athīna: Nomikī Vivliothīkī/Ellīnikī Etaireia Orologias.

van Brussel, Laura, Arda Tezcan, and Lieve Macken. 2018. A Fine-grained Error Analysis of NMT, PBMT and RBMT Output for English-to-Dutch. In *Proceedings of the 11th Conference on Language Resources and Evaluation (LREC 2018)*, eds. Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk, Stelios Piperidis and Takenobu Tokunaga, 3799–3804. Paris: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2018/index.html (accessed December 28, 2022).

Van de Cruys, Tim. 2011. Two Multivariate Generalizations of Pointwise Mutual Information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality (DiSCo'2011)*, 16–20. Stroudsburg: Association for Computational Linguistics.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention Is All You Need*. DOI: https://doi.org/10.48550/arXiv.1706.03762.

Vaswani, Ashish, Samy Bengio, Eugene Brevdo, Francois Chollet, Aidan N. Gomez, Stephan Gouws, Llion Jones, Łukasz Kaiser, Nal Kalchbrenner, Niki Parmar, Ryan Sepassi, Noam Shazeer, and Jakob Uszkoreit. 2018. *Tensor2Tensor for neural machine translation. In Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: MT Researchers' Track)*, eds. Colin Cherry and Graham Neubig, 193–99. Association for Machine Translation in the Americas. https://aclanthology.org/W18-1819.pdf (accessed December 28, 2022).

Wiesmann, Ewa. 2019. Machine translation in the field of law: A study of the translation of Italian legal texts into German. *Comparative Legilinguistics* 37: 117–153. DOI: http://dx.doi.org/10.14746/cl.2019.37.4.

Zhang, Jiajun, Shujie Liu, Mu Li, Ming Zhou, and Chengqing Zong. 2014. Bilingually-constrained Phrase Embeddings for Machine Translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Volume 1: Long Papers, ACL 2014, June 22–27, Baltimore,*

111–21. Stroudsburg: Association for Computational Linguistics. DOI: https://doi.org/10.3115/v1/P14-1011.

Websites and blogs

Bundesministerium der Justiz, Bundesamts für Justiz – www.gesetze-im-internet.de (accessed October 12, 2021).

Cambridge Spark. Tutorial: Build your own skip-gram embeddings and use them in a neural network. https://blog.cambridgespark.com/tutorial-build-your-own-embedding-and-use-it-in-a-neural-network-e9cde4a81296 (accessed March 25, 2022).

Jay Alammar, jalammar.github.io. Visualizing machine learning one concept at a time. The Illustrated Transformer.https://jalammar.github.io/illustrated-transformer/ (accessed March 25, 2022).

peterbloem.nl. Transformers from scratch.https://peterbloem.nl/blog/transformers (accessed March 25, 2022) (accessed March 25, 2022).

LASER Language-Agnostic SEntence Representations. GitHub - facebookresearch/LASER: Language-Agnostic SEntence Representations (accessed July 22, 2022).

Towards Data Science. Word2Vec for phrases learning embeddings for more than one word. https://towardsdatascience.com/word2vec-for-phrases-learning-embeddings-for-more-than-one-word-727b6cf723cf (accessed March 25, 2022).