Analyzing variation in translation through neural semantic spaces

Yuri Bizzoni
University of Saarland, Germany
yuri.bizzoni@uni-saarland.de

Elke Teich
University of Saarland, Germany
e.teich@mx.uni-saarland.de

Abstract

We present an approach for exploring the lexical choice patterns in translation on the basis of word embeddings. Specifically, we are interested in variation in translation according to translation mode, i.e. (written) translation vs. (simultaneous) interpreting. While it might seem obvious that the outputs of the two translation modes differ, there are hardly any accounts of the summative linguistic effects of one vs. the other. To explore such effects at the lexical level, we propose a data-driven approach: using neural word embeddings (Word2Vec), we compare the bilingual semantic spaces emanating from source-to-translation and source-to-interpreting.

1 Introduction and Related Work

Our research question stems from the field of translation studies. Revisiting the notion of ’translationese’ (Gellerstam, 1986), i.e. the specific linguistic traces left in the translation product by the process of translation, we are interested in patterns of lexical choice in translation versus interpreting. To explore this, we need (a) summaries of the dominant lexical choices made in translation and interpreting and (b) a method of comparing them.

Existing research on translationese (Baker, 1996) has mainly focused on (sets of) predefined features (e.g. type-token ratio, sentence length, part-of-speech distributions), applied in classification tasks comparing translations and original texts (Baroni and Bernardini, 2006; Volansky et al., 2015; Rubino et al., 2016). While this work has brought genuine insights regarding the language of translation, we still have a fairly fragmented picture of translation behavior and its many facets (Lapshinova-Koltunski, 2013, 2015). For instance, it has been shown that translations exhibit source language interference or shining-through (Toury, 1995; Teich, 2003), against the assumption of translation universals; or that certain groups of translators show higher convergence in translation choice than others (see e.g. (Martínez Martínez and Teich, 2017) who study the outputs of translation learners and professionals by entropy).

Here, we are interested in written translation vs. simultaneous interpreting. Among the known differences are more frequent and different kinds of omissions in interpreting (He et al., 2016) and, depending on the source - target language pair, unusual word orders (Collard et al., 2018). However, there is no comprehensive, systematic picture yet, also due to the fact that specific and systematic studies of interpretation are a relatively recent phenomenon (Pöchhacker, 2016).

Nonetheless, we can formulate some hypotheses. Beyond the notorious difficulties of bridging a “message” between two languages - difficulties that are constantly analyzed in translation studies (Eades, 2011; Li, 2019) - the process of interpreting is complicated by the dire time constrains of the process and by the absence of an editing phase, essential in many translation processes (Schaeffer et al., 2019). We might thus assume that due to high cognitive pressure, interpreters may not be able to adapt their output to the target language norms as well as translators do, which might be reflected in lower lexical richness and lexically less coherent interpreting output compared to translation output.

On the computational side, the approach proposed here is related to attempts at making word embeddings fruitful for linguistic analysis, notably modeling diachronic language change (Dubossarsky et al., 2017; Fankhauser and Kupietz, 2017; Bizzoni et al., 2019). Also, there is some

Proceedings of the 12th Workshop on Building and Using Comparable Corpora at RANLP 2019, Varna, Bulgaria.
Serge Sharoff, Pierre Zweigenbaum, Reinhard Rapp (eds.)

resemblance to the problem of creating domain-specific word embeddings (Zhang et al., 2019; Wang et al., 2018).

Concretely, the method we propose here aims at building bilingual word embeddings from aligned corpora. In the last years, a significant amount of research has gone into the construction of more effective multilingual word embeddings (Zhang et al., 2017; Artetxe et al., 2018) from smaller datasets (Artetxe et al., 2017) or with the help of multimodal data (Singhal et al., 2019).

But while most works on multilingual distributional semantics focus on creating consistent spaces (Huang et al., 2018) showing robust properties across languages (Brychcín et al., 2019), our aim is creating semantic spaces that model the lexical choices of a *specific* kind of linguistic behavior, i.e. translation, which we call here *translation spaces*. Specifically, we train two bilingual distributional models on two monolingually comparable corpora, a larger one of translation, and a smaller one of interpreting, and we compare them to detect differential patterns of translation mode-induced lexical choice.

It is important to underline that in this first stage, the gist of our analysis comparing semantic spaces will be qualitative (Sections 4.1-4.3). Qualitative analyses are somewhat easier on bilingual than on monolingual spaces, for the reason that in bilingual spaces we often know the "ground truth" (e.g. we know that the Spanish translation of *Germany* is *Alemania*), while the similarities displayed by monolingual word embeddings are harder to judge case by case. Therefore, our bilingual word embeddings are directly comparable and we are able to present a conclusive quantitative perspective on the spaces' overall topology as well (Section 4.4). In that case, we will just consider the mean distances, without looking at the actual words in a cluster.

For all our experiments we used gensim's implementation of Word2Vec (Mikolov et al., 2013).

2 Corpora

Our data set is composed of Spanish and German translations of the same English source (speeches from the European Parliament) (Karakanta et al., 2018) as well as interpreted speeches in the same two target languages. For written translation, each language is represented by circa 20 million characters and 130.000 sentences. For German, we have created a corpus of interpreted speech of English into German from the European Parliament including materials from existing corpora (Sandrelli and Bendazzoli, 2005; Bernardini and Milievi, 2016). The resulting interpreting corpus is strictly comparable to the translation corpus in terms of register and domain but contains much less material (568.230 characters and 3.397 sentences per language).

3 Methodology

3.1 Creating translation spaces

The input for a translation space is constituted by the tokenized, concatenated aligned sentences of a source-translation corpus. In other words, each sentence from a source text X is concatenated with its translation in a target text Y, creating a bilingual pseudo-sentence. If we were dealing with an idealized word-by-word translation, this pseudo-sentence would be simply composed by lexical source-target pairs; in the case of a more realistic translation, we can still confidently expect that a percentage of the words in the source language will find a direct target correspondence within the same pseudo-sentence. After creating the pseudo-sentences, we train a standard skip-gram Word2Vec model on them, using as context window the mean + standard deviation length of the sentences (in our case, we set each word's context at 160 words, which is the double of the mean sentence length plus standard deviation). Before training, the words in each aligned sentence were shuffled: this proved to yield slightly better results.

The logic of this approach is that words having a consistent translation in an aligned corpus will share very similar contexts, ending up in close proximity in the resulting distributional space.

An important problem to address is the variability of Word2Vec's results at different run times. While the specific cosine similarity is bound to undergo oscillations between different runs, all the rankings we present in the following tables have been verified through multiple runs: in other words, if the cosine similarities slightly changed, the ordering and the magnitude of the results remained the same. In future we intend to verify the stability of our spaces more consistently (see Section 5).

3.2 Probing the translation spaces

Words that translate each other in a very consistent way throughout the corpus appear to be very close in the resulting semantic space, and are often each other's nearest neighbours. For example, in the English-Spanish translation space, the nearest neighbour of *Germany* is *Alemania*, of *Italy Italia*, and so forth. Also, country names in both languages create a tight semantic cluster, and happen to be in the same lexical neighborhood (see Table 1).

This example shows the qualities of a space deriving from a translation where each term has one and one only direct equivalent in the other language: the group of country names forms a cluster which is both bilingually sound (*Alemania* is the closest word to *Germany*) and semantically coherent (*Italy*, *Italia* and *France* are the three following neighbours of *Alemania*). We can consider this cluster as representative of extremely faithful translations: situations where each word in X has its undiscussed equivalent in Y. Such peculiar cases of "extreme" source text fidelity guarantee both semantically and translationally sound distances. On the opposite side of the spectrum, we can find elements that rarely have a single, obvious equivalent in another language: function words. Words like Spanish *el* or English *to* and *if*, do not have a meaningful closest neighbour in the other language and are on average further apart from other words than words with one obvious translation: they form looser clusters.

Translation spaces are of particular interest in the cases between these two extremes. Both the identity and the distance of neighbours become indicative of a translation "style". For example, in the same space, we find that *war* is closest to *guerra* (cosine similarity .91) but also relatively close to *terror* (0.71), *fria* (0.69), *cold* (0.69): *war* seems to be consistently translated, and in a semantically quite coherent cluster. The nearest neighbour of *voz* is *voice*, with a cosine similarity of 0.91, but its second nearest neighbour, *solidaridad*, has a similarity of only 0.57, followed by words mainly in Spanish, such as *sola* and *expresarse*; *voz* has a consistent translation, but belongs to a less obvious paradigm.

The comparison with the country names, where each word is nearest to its translation and very near to other country names in both languages, is helpful to see how we are moving towards more semantically complex cases: *war* and *voz* are words that have a preferential translation, but do not belong to conventionalized paradigms as predictably translated as country names.

Germany	**war**	**voz**
Alemania.95 Italia.86 Italy.86	guerra.91 terror.71 fria.69	voice.91 solidaridad.57 sola.56

Table 1: Three words and their three nearest neighbours with cosine similarities in the Spanish-English translation space.

If instead a word is *not* consistently translated, there are two possible configurations in the translation space:

1. The word is close to its various translations in the space, but the similarity is relatively low. This seems to represent the case of well defined polysemy, where one word is consistently translated with one among N choices in the target language: for example *fear* is close to *temo*, *miedo* and *temor*, and their cosine similarities are between 0.62 and 0.7.

2. The word isn't close to any translating term in the other language, and does not present a high similarity to its neighbours. This seems to represent the case of words that are particularly hard or impossible to translate with one term in the target language. Such words produce many contextual translations, rephrasing, or omissions, and this "'productivity" in turn makes their distributional profile relatively idiosyncratic, distancing them from all other points in the space. For example, *somehow* has no close neighbours in Spanish, and its nearest term in the space, *foolish*, has a cosine similarity of only 0.49; the nearest neighbour of *weekend* is *week* (0.57) and the nearest neighbour of *insight* is *spirit* (0.41).

This simple mirroring between the source fidelity of a translation and the tightness of a distributional cluster can be a special way to detect several translation behaviours (see Table 2).

4 Translation spaces for comparable corpora

As a use case, we want to adopt this system of building distributional spaces to compare lexical

gentes	**the**	**palestinian**
oppressed.54	mandato.23	palestino.9
gente.52	de.23 ca-	palestinos.88
pueblo.52	chemir.22	israeli.87
population	**quiero**	**sucesor**
inhabitants.61	quisiera.88	successor.84
viven.57	deseo.75	gallant.6
living.57	desearia.7	franco.56

Table 2: Words with no direct translation in loose clusters (*gentes*, *the*), words with direct translation in tight semantic clusters (*palestinian*), words with some semantic tightness but no direct translation (*population*, *quiero*), words with direct translation in a loose semantic cluster (*sucesor*). In the majority of cases, no direct translation means lower semantic similarity with the nearest word, ergo looser clusters. An "untranslatable", be it real or perceived, doesn't have single words that share its context with the same regularity of a translating term, and thus tendentially creates looser groups.

fidelity between the translation corpus and a comparable interpreting corpus.

We conduct first a qualitative analysis of the differences, and then a quantitative analysis of the topological differences between the two spaces.

4.1 English-German translation space

Following the procedure described in the previous section, we train a translation space on the tokenized and aligned sentences of the English-to-German written translation corpus, with a context window of 160 words and a dimensionality of 300.

This space seems to behave coherently with what we would expect:

1. Words with single, highly preferred translations form translation and semantic tight groups: *Germany* is close to *Deutschland* (0.94) and *Belgien* (0.84); *Mord* is close to *murder* (0.95) and *brutale* (0.86). Technical terms too tend to display high nearest neighbour similarities: *unemployment* - *Arbeitslosigkeit* (0.89), *decriminalisation* - *Entkriminalisierung* (0.96).

2. Words belonging to semantically complex paradigms fall relatively close to their preferential translation when they have one, but their clusters are looser: *force* is the nearest neighbour of *Kraft* (0.67) and *Friedenstruppe* (0.64).

3. Words with various translations fall close to their equivalents, but their similarities are low: *happy* is close to *glücklich* (0.62), *erfreut* (0.52), *zufrieden* (0.51).

4. Words without a single term translation are at the center of very loose clusters, with nearest neighbours' similarities ranging between 0.6 and 0.4.

Both in this space and the English-Spanish one, geometric analogies of the sort of "'man : woman = king : x" (Mikolov et al., 2013) are possible with various terms: "man : woman = Mann : x" returns *Frau*; "glücklich : sad = happy : " returns *traurige*; "Freiheit : Presse = freedom : x" returns *press* and *newspapers*. In other words, the sum vector of *Freiheit* + *freedom* minus *Presse* returns a point that is closest to *press*.

While such results are the effects of consistent translation embeddings, this particular space also shows peculiarities that are due to the specifics of German compounding: the sum vector of *freedom* + *press* is close to *pressefreiheit* (0.70); the sum vector of *freedom* + *expression* is closest to *meinungsfreiheit* (0.88), and so forth.

Interestingly, the closest neighbours of *meinungsfreiheit* are *expression* and *freedom* with relatively high degrees of similarity (0.88 and 0.79): terms that have a multi-word consistent translation can still exhibit tight clustering properties.

4.2 English-German interpreting space

The interpreting space shows properties in common with the translation space, but with relevant differences.

1. Words with a highly preferred single translation fall closest to such translation, but do not seem to form semantically cohesive clusters: *Germany* is closest to *Deutschland* (0.98), but is not in a cluster of country names.

2. Words that showed a variety of translation neighbours in the translation space either present a single, very close meaningful neighbour (*zufrieden* has a 0.86 similarity to *satisfied*, but no other English words appear in its immediate vicinity), or tend to show no meaningful clustering at all.

Full Translation Space	Small Translation Space	Interpreting Space
Germany: Deutschland.94, Belgien.84, Frankreich.84	**Germany:** Deutschland.99, vivendi.84, France.84	**Germany:** Deutschland.98, politically.8, tragbar.78
somehow: irgendwie.7, Wahrheit.6, erwecken .59,	**somehow:** irgendwie.84, anhängen.83, pollute.8	**somehow:** enjoy.64, speaks.63, volumes.63
happy: glücklich.63, erfreut.53, zufrieden.51	**happy:** glücklich.88, verspatung.74, soweit.72	**happy:** glad.67, glücklich.65, m.65
Vertrag: treaty.79, Nizza.77, Nice.72	**Vertrag:** treaty.89, idea.66, settle.65	**Vertrag:** treaty.93, Lissabon.84, Lisbon.8

Table 3: Three nearest neighbours of *Germany*, *somehow*, *happy* and *Vertrag* for the full scale Translation Space, the down-sampled Translation Space, and the Interpreting Space.

3. Finally, words with no direct translation remain inside loose, sparse clusters.

4.3 Subsampling and comparison

The main problem with comparing the two spaces is difference in corpus size, the translation corpus being significantly larger than the interpreting corpus.

To take this aspect into account, we randomly sampled the translation corpus in order for it to have the same number of aligned sentences as the interpreting corpus, and trained a new distributional space on it.

A qualitative presentation of the difference between the three spaces is in Table 3.

Four main observation can be made:

1. The down-sampled translation model keeps some looser semantic cohesion with country names, while the interpreting model seems able to "only" retrieve the direct term translation;

2. Adverbs such as *almost*, *probably*, *irgendwie* etc. retrieve an equivalent in the translation spaces, but not in the interpreting space.

3. In some cases, such as in the case of *happy*, the effect of data scarcity is that of strengthening the relation between a term and one of its possible translations, probably due to the absence of alternatives in the down-sampled corpus; this makes the loose similarity of *happy* with *glücklich* in the interpreting corpus more relevant.

4. The relation between translatability and cosine similarity seems to hold through the spaces: if two neighbours translate each other, their similarity tends to be higher than if they are simply semantically related.

4.4 Topological comparison of the spaces

Given these observations, we can proceed to a comparison of some topological properties of the translation sub-sampled space and the interpreting space (see Table 4 for a summary).

We note that despite being of equal length, the translation model has more words than the interpreting model: 18 592 versus 10 524. For this comparison, we will focus on the 6 753 words that they have in common.

The average word similarity within the translation model is 0.26, six points higher than the average similarity within the interpreting model. But if we limit our computation to every word's nearest neighbour in each model, we see a different picture emerging. The distance between the models shrinks to no true significance, with the interpreting space showing even a slightly higher similarity than the translation: nearest neighbours in the translation space have an average cosine similarity of 0.85, those in the interpreting space of 0.86.

The average word similarity is different between the two spaces, but the nearest neighbour similarity is approximately the same. In other words, interpreting spaces present a less homogeneous distribution than translation spaces: they display words with nearer neighbours in looser clusters. This distribution seems to go along with our observations of a two-folded fidelity to the source: interpreting seems to show a high level of fidelity with respect to unambiguous, domain-specific words (*treaty* - *Vertrag*, *president* - *Präsident*) where the translation space presents a lower degree of similarity (more diversity in the translation). The result of these cases is close nearest neighbours followed by lower similarity words in interpreting spaces; and more distant nearest neighbour followed by closer alternatives in translation spaces.

At the same time, other categories of words, such as adverbs (*irgendwie*, *really*, *next*), and some non-domain specific words (*tag*, *muss*) seem to have no systematic equivalent in the interpreting space, while they do retrieve a translating nearest neighbour in the translation spaces, both full and "down-sampled". This may suggest a preference on the part of the interpreter for a precise translation of domain-specific content at the expense of interpersonal or textual expressions. These opposed tendencies could be the cause of the special topology we seem to observe in the spaces.

	Translation	**Interpreting**
vocab size	18592	10524
avg simil.	0.268	0.213
1st neigh.	0.851	0.860
10th neigh.	0.723	0.685

Table 4: Vocabulary size, average overall similarity, first and tenth nearest neighbour average similarity for the down-sampled translation space and the interpreting space. The mean difference between the first and tenth neighbour also shows the "loosening" of the similarity queues in the interpreting space, symptom of generally looser word clusters.

5 Conclusions and Future Work

We have presented a method of exploring variation in translation, here focusing on translation vs. interpreting, using neural word embeddings. Creating two models, a source-translation model and a source-interpreting model, from the same domain (European Parliament speeches) we explored similarities and differences between the two lexical-semantic spaces. To obtain better comparability, we down-sampled the dimensions of the translation corpus in order to avoid mistaking frequency effects for true translation behaviours.

Our comparison has revealed both differences in the overall topology of two semantic spaces (looser word clusters in interpreting compared to translation) as well as differences in how translators vs. interpreters handle certain types of vocabulary (e.g. domain-specific vs. general words). We can speculate on some possible reasons of the differences between the spaces: for example, the two-folded fidelity of interpretation could be due to the fact that while interpreting forces a more deliberate rephrasing of the source (which also comes with the apparent sacrifice of interpersonal or textual expressions), formulaic or highly predictable words are easier to translate always with the same equivalent. Nonetheless, we find that more research has to be done in order to make such claims substantial.

In our ongoing work, we use the same method for looking at other variables, e.g. the influence of source language on the translation output and the level of translation *expertise* (learner vs. professional), and analyze translation spaces further in terms of entropy, as an index of lexical variation. Another matter we want to address more consistently is that of Word2Vec's possible sensitivity to words' frequency. We think that our spaces are more resistant than monolingual spaces to random initialization simply because they are modelling a more clear-cut phenomenon: if a low frequency word has a consistent translation, its distributional profile will still be uniquely similar to that of the translation. Nonetheless, we intend to evaluate this method more substantially, comparing the spaces' results to bilingual dictionaries and synthetic data, which could also help us assess the impact of frequency effects. Also, we intend to compare this method's results with the results of a post-training aligned bilingual space, and to use the proposed method for translation evaluation, complementing it with other means of comparative textual analysis, such as relative entropy.

References

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pages 451–462.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. *arXiv preprint arXiv:1805.06297* .

Mona Baker. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers, editor, *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, John Benjamins, Amsterdam and Philadelphia.

Marco Baroni and Silvia Bernardini. 2006. A new approach to the study of translationese: Machinelearning the difference between original and translated text. *Literary and Linguistic Computing* 21(3):259–274.

S. A. Ferraresi Bernardini and M. Milievi. 2016. From epic to eptic exploring simplification in interpreting and translation from an intermodal perspective. *Target* 28:61–86.

Yuri Bizzoni, Stefania Degaetano-Ortlieb, Katrin Menzel, Pauline Krielke, and Elke Teich. 2019. Grammar and meaning: Analysing the topology of diachronic word embeddings. In *Proceedings of First Workshop on Computational Approaches to Historical Language Change, ACL.*

Tomáš Brychcín, Stephen Taylor, and Lukáš Svoboda. 2019. Cross-lingual word analogies using linear transformations between semantic spaces. *Expert Systems with Applications* .

Camille Collard, Heike Przybyl, and Bart Defrancq. 2018. Interpreting into an s.o.v. language: memory and the position of the verb, a corpus-based comparative study of interpreted and non-mediated speech. *Meta* 63(3):695–716.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 conference on empirical methods in natural language processing*. pages 1136–1145.

Domenyk Eades. 2011. Translating english modal expressions: an arab translator trainees perspective. *Babel* 57(3):283–304.

Peter Fankhauser and Marc Kupietz. 2017. Visualizing language change in a corpus of contemporary german .

Martin Gellerstam. 1986. *Translationese in Swedish novels translated from English*. CWK Gleerup.

He He, Jordan Boyd-Graber, Jordan Boyd Graber, and Hal Daumé III. 2016. Interpretese vs. translationese: The uniqueness of human strategies in simultaneous interpretation. In *Proceedings of NAACL-HLT*. San Diego, CA, pages 971–976.

Lifu Huang, Kyunghyun Cho, Boliang Zhang, Heng Ji, and Kevin Knight. 2018. Multi-lingual common semantic space construction via cluster-consistent word embedding. *arXiv preprint arXiv:1804.07875* .

Alina Karakanta, Mihaela Vela, and Elke Teich. 2018. Preserving metadata from parliamentary debates. In Darja Fier, Maria Eskevich, and Franciska de Jong, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).

Ekaterina Lapshinova-Koltunski. 2013. VARTRA: a comparable corpus for analysis of translation variation. In *Proceedings of the 6th Workshop on Building and Using Comparable Corpora*. Sofia, Bulgaria, pages 77–86.

Ekaterina Lapshinova-Koltunski. 2015. Variation in translation: Evidence from corpora. *New directions in corpus-based translation studies* pages 93–113.

Saihong Li. 2019. A corpus-based multimodal approach to the translation of restaurant menus. *Perspectives* 27(1):1–19.

José Manuel Martínez Martínez and Elke Teich. 2017. Modeling routine in translation with entropy and surprisal: A comparison of learner and professional translations. In Larisa Cercel, Marco Agnetta, and María Tereza Amido Lozano, editors, *Kreativitt und Hermeneutik in der Translation*, Narr, Tbingen, pages 103–126.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111–3119.

Franz Pöchhacker. 2016. *Introducing interpreting studies*. Routledge.

Raphael Rubino, Ekaterina Lapshinova-Koltunski, and Josef van Genabith. 2016. Information density and quality estimation features as translationese indicators for human translation classification. In *Proceedings of NAACL*. San Diego, CA, pages 960–970.

Annalisa Sandrelli and Claudio Bendazzoli. 2005. Lexical patterns in simultaneous interpreting: A preliminary investigation of epic (european parliament interpreting corpus). In *Proceedings of the Corpus Linguistics Conference Series 1*.

Moritz Schaeffer, Anke Tardel, Sascha Hofmann, and Silvia Hansen-Schirra. 2019. Cognitive effort and efficiency in translation revision. In *Quality Assurance and Assessment Practices in Translation and Interpreting*, IGI Global, pages 226–243.

Karan Singhal, Karthik Raman, and Balder ten Cate. 2019. Learning multilingual word embeddings using image-text data. *arXiv preprint arXiv:1905.12260* .

Elke Teich. 2003. *Cross-Linguistic Variation in System und Text. A Methodology for the Investigation of Translations and Comparable Texts*. Mouton de Gruyter, Berlin.

Gideon Toury. 1995. *Descriptive Translation Studies and beyond*. John Benjamins, Amsterdam/Philadelphia. https://doi.org/10.1075/btl.4.

Vered Volansky, Noam Ordan, and Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities* 30(1):98–118. https://doi.org/10.1093/llc/fqt031.

Yanshan Wang, Sijia Liu, Naveed Afzal, Majid Rastegar-Mojarad, Liwei Wang, Feichen Shen, Paul Kingsbury, and Hongfang Liu. 2018. A comparison of word embeddings for the biomedical natural language processing. *Journal of biomedical informatics* 87:12–20.

Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017. Adversarial training for unsupervised bilingual lexicon induction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pages 1959–1970.

Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, and Zhiyong Lu. 2019. Biowordvec, improving biomedical word embeddings with subword information and mesh. *Scientific data* 6(1):52.