PBML

The Prague Bulletin of Mathematical Linguistics

NUMBER 119 OCTOBER 2022 23–36

Translating Argumentation: Distributional Semantic Analysis of Argumentation Resources in Parallel Corpora

Vahram Atayan, Bogdan Babych

Institute for Translation and Interpreting, Heidelberg University

Abstract

In the paper, we report results of our experiments on identifying distributional semantic characteristics of different types of lexical items used in argumentation: connectors, meta-argumentative words, key notions of a given discourse and the evaluative/connotative lexicon. These characteristics are contrasted within monolingual English corpora of different genres (Europarl-EN and Cord COVID-19) and in translation context for German-into-English direction (Europarl-DE and Europarl-EN-from-DE). For the analysis, we propose a number of new methods that better characterize distributional semantic differences between the argumentatively relevant lexical items, such as measuring the knee in the mutual information-ranked list curve, testing categories for span variation and different selection procedures. In our experiments, meta-argumentative lexical items show the biggest differences in their distribution with other word types on several of such measures. The analysis based on word vector allows us to create a selection heuristic for candidate lists for different categories of argumentative lexicon.

1. Introduction

The characterization of linguistic resources used in argumentation has been an important challenge for contrastive linguistics, qualitative and corpus-based translation studies (e.g. Atayan, 2007). In particular, different languages build argumentation structures with different inventories of lexical, morphosyntactic, and discursive means, as well as usage patterns, which involve interaction across various linguistic levels and paradigmatic sub-systems of a language (argumentative connectors, evaluative/connotated lexicon, meta-argumentative constructions, etc.). Also syntagmatically the argumentation patterns may span non-local context, extending through

 Corresponding author: atayan@uni-heidelberg.de
Cite as: Vahram Atayan, Bogdan Babych. Translating Argumentation: Distributional Semantic Analysis of Argumentation Resources in Parallel Corpora. The Prague Bulletin of Mathematical Linguistics No. 119, 2022, pp. 23–36. doi: 10.14712/00326585.021.

several sentences and larger discourse units. However, it is difficult to automatically identify and align multiword argumentation patterns in multilingual corpora with sufficient accuracy. This seriously limits the applicability of standard corpus-based methods, tools and annotation resources for their study, since most traditional approaches primarily target phenomena in the local context of corpus searches (e.g., morphosyntactic or lexical patterns within a window of a few words). Specifically, while there have been several corpus-based studies of systematic differences between original texts and translations (so called 'translationese') on the levels of the general lexicon, modal markers, morphosyntactic patterns, indirect equivalents (e.g., Babych et al., 2007; House, 2011; Hoey, 2011; Kranich and Gast, 2015; Gast, 2022), the range of such studies for argumentation patterns across languages has been limited.

Our paper addresses this methodological gap in contrastive corpus-based analysis of argumentation, namely we suggest a number of new distributional semantic properties of argumentation resources, which allow us to quantify differences in their usage across languages and genres or in original and translation corpora within the same language. In the paper we report the results of our experiments on evaluating relevant distributional parameters of three types of argumentation patterns in multilingual and translation context: (a) meta-argumentative words; (b) key notions of a given discourse; (c) evaluatives/connotated lexicon. Lexical items of these three categories have been manually annotated in two different selections of ca. 1000 word types from each of our corpora: (1) The Europarl corpus of parliamentary proceedings (Koehn, 2005), where we selected original texts that have been authored in (1a) German and (1b) English, as well as (1c) English texts translated from German. (2) The CORD-19 English monolingual corpus of medical research articles about Covid-19 (Wang et al., 2020). Both corpora are POS-tagged and lemmatized (Schmid, 2013).

Therefore, this paper seeks to make two kinds of contribution. Firstly, we propose new methods for distributional analysis of argumentatively relevant lexical items, that extend standard collocation-based approaches used nowadays in corpus linguistics to characterize the general lexicon. Secondly, we identify distributional characteristics of the argumentative lexicon that best distinguish different types of argumentation resources and allow for linguistic interpretations that verify or extend existing theoretical models of argumentation. Finally, we use manually annotated lists of argumentatively relevant lexical items of different types to generate automatically further candidates for the three categories using word vector models.

The methods of distributional semantics are typically based on identification of certain *observable* features and annotations in text corpora, their statistical analysis and linguistic interpretation, which often creates a possibility to test quantitative predictions made by linguistic models and provides new linguistic insights or improved understanding of the phenomena, modelling their structure and interaction, possibly leading to new testable predictions. Because such explicit features are more easily found on the lexical and morphosyntactic levels, i.e., in the local context of linguistic constructions, there is a certain 'street-light effect' within the current research paradigm, which has mostly focussed on phenomena below the sentence level. This also applies to the study of cross lingual phenomena in corpus-based contrastive linguistics and translation studies (e.g., Kruger et al., 2011).

Still, corpus-based studies of certain discourse-level phenomena (such as discourse particles) have indicated that specific linguistic properties of these resources result in different sets of observable features, which distinguish them from the general lexicon. The research suggests that traditional corpus-based methods are much less effective for the study of phenomena on the discourse level, especially when such resources do not have direct translation equivalents (Gast, 2022, 323-324).

Argumentation is a phenomenon of communicative discourse (van Eemeren, 2018; van Eemeren et al., 2019, 5), so applicability of lexically-oriented corpus methodologies would be limited for modelling its distributional semantic properties. Therefore, the distributional analysis of argumentation resources and their translation equivalents across languages would involve the tasks of (1) identifying their characteristic observable features in corpora, also beyond the local context and (2) developing statistical measures that quantify them and could capture distributional properties of different types of argumentation resources, differences in their usage across genres, across different languages, as well as in original vs. translated discourse.

2. Methodology: distributional measures for argumentation lexicon

To analyze the impact of the distributional semantics on the argumentation, we defined four categories of argumentatively relevant lexical items: connectors (like *since*), evaluative or connotated words (like *dangerous* or *progress*), content-related non-connotated key notions of a given discourse (*Commission* for Europarl or medical terms like *venous* for COVID corpus) and meta-argumentative words (like *disagree* or *reason*). For the present study we are concentrating on the last three lexical groups. Our goal is to identify potential differences between these groups concerning their distributional properties, measured via different parameters related to the mutual information (MI) of a given word (with a cut-off at the frequency of 50 items/corpus) and its collocates (with a cut-off at least 10 tokens/corpus) defined as

$$\mathrm{MI} = \log_2(\mathrm{ObservedFreq}^2/\mathrm{ExpectedFreq}), \tag{1}$$

where ObservedFreq and ExpectedFreq are observed and expected frequencies of a collocate in the context of the word calculated in a given span (Evert (2008, 19) considers this relation as most popular in the MI-calculation). $\mathrm{MI} > 1$ is typically considered as indicator for a collocation relation between two words. In the previous research, collocation analyses have been conducted for different spans, ranging from 1 to dozens or hundreds of words; the choice of the specific span is of essential importance in the research (Evert, 2008, 12).

In our analysis, we are taking into consideration collocation spans from 1 to 9 in order to understand the potential differences in the behaviour of different types of

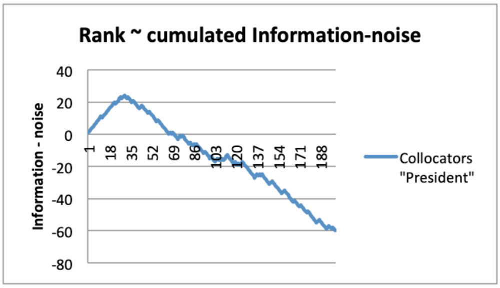


Figure 1: Knee: signal vs. noise in MI-ranked collocates of *president*

argumentative lexemes in local vs. rather clause-level context. Since the number of collocates is normally growing monotonically with the span, we suggest taking into consideration rather the number of most informative collocates measured by MI. Typically, the MI of the collocates of a given word seems to follow a Zipfian distribution with a small number of strong collocates and a wide range of collocates with low MIs, potentially generating more noise than useful information. To improve the information/noise ratio for a data set, it's possible, of course, to simply cut off the list by a fixed number or a selected threshold of MI. Yet both choices are rather arbitrary, so that we suggest instead taking into consideration only the collocates with an above average contribution to the overall MI of the whole collocate set. To do this, we calculate the knee of the discrete curve of decreasing MI values.

As Satopaa et al. (2011) point out, knee points of a curve represent in general the best balance for inherent trade-offs, in our case, between stronger collocates and rather accidental correlations. To illustrate informally the potential usefulness of this concept for our goals, we calculated the first 200 collocates of the word *president* in the English Europarl corpus (using

$$\mathrm{MI}_{\mathrm{lex}} = \mathrm{ObservedFreq}/\mathrm{ExpectedFreq} \times \log_2(\mathrm{CandidateFreq}) \tag{2}$$

as MI measure to select rather fully lexical items). In the next step, we annotated the items clearly related semantically or pragmatically to the lexeme *president*, such as proper names (*Obama*), organizations (*PPE*), countries (*Venezuela*), related political functions (*Chancellor*) and genre related collocates (*madame*) and weighted those with 1. We considered formally as noise non-evident collocates like *too, ask, incoming* etc. and weighted those with -1. Finally, we calculated the rank-dependent overall information as the sum of weights up to the given rank.

In Figure 1 we can easily see that the "noise" (here interpreted as statistically relevant but intuitively non-evident collocates) becomes dominant after first ca. 72 items. The calculation (`kneed` Python package) of the knee of the MI curve for *president* (with 3548 collocates with MI over 1) gives us 74 as knee value, which agrees rather well with our informal estimation (further studies with more lexemes will be conducted in the future to verify the usefulness of knee as information measure). We consider thus the knee value as a general informative characterization of the collocates set of a given lexeme. To obtain an overall result for our argumentatively relevant items, we calculate the average value of the knees of all elements for every given subset. Linguistically speaking, we are asking for the number of particularly informative collocates for the lexical items and try to understand the behaviour of this parameter, which reflects, for different spans, the interaction of a given word with local and extended context.

In our analysis, the COVID-19 corpus represents a culturally-neutral genre, which has distinct argumentation patterns typical for scientific discourse. The Europarl corpus is more culture-specific and uses argumentation patterns typical for political debates. For the analysis, we focus on genre comparison (COVID vs. Europarl) and on parallel English and German texts, both originally authored and translated. These corpora are used as a benchmark for the proposed methodology. They also allow us to develop linguistic interpretation of the results and to make further testable predictions about the behaviour of argumentation structures in translation. The sizes of the corpora: sub-corpus of COVID: 3.360.000 tokens; EP_EN: 7.281.000, sub-corpus of EP_DE: 2.581.000, EP_ENfDE: 3.854.751.

3. Experiment on knee comparison across different collocation spans

We extracted two different subsets of argumentative lexemes by annotating the categories of evaluative/connotated, key notion and meta-argumentative among ca. 1000 words with the highest average MI of all collocates (linguistically speaking, words with smaller sets of strong collocates) and among ca. 1000 randomly selected items for our four corpora. In general, we would assume a monotonic increase of the knees over spans, because broader spans typically introduce more weak collocates reinforcing indirectly the information contribution of stronger ones, which previously were slightly under the knee point, and then eventually coming to saturation at larger spans. For MI-based selections we expect generally lower knee values, due to the dominance of a small number of strong collocates typical in such cases. This hypothesis is confirmed by our data, cf. Figure 2. As far as genre and translation effects are concerned, we can report the following findings.

1. In the genre comparison, presented in Figure 2, we see the particularity of key notions in the COVID MI-based selection with a steady increase of the knee value over spans 1-9 as compared to evaluative/connotated and meta-argumentative words. This effect is possibly related to the internal inhomogeneity of the corpus texts (scientific articles) with well-defined functional parts (ABSTRACT, METHODS, DISCUSSION etc.). Due to this repartition, some parts of the vocabulary are not uniformly distributed in the texts but concentrated in one or other type of subtext. This could be the case for certain specialized key notions with high

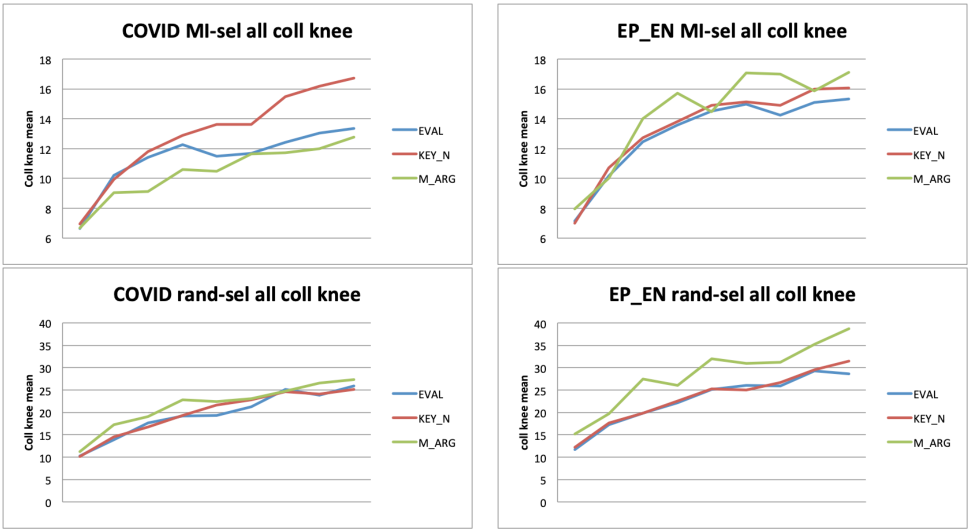


Figure 2: Size of knee in Covid vs. EP corpora: MI-ranked and random selection

average MIs, present in rather technical parts of an article. Now, an extension of the collocation span would in such cases integrate potential new collocates from the same subtext, where a part of the vocabulary would be generally over-represented in comparison with the other parts of the article, even without a specific relation to the original, i.e., collocation-based key notion. So we would have something like keywords of the given subtext type (as compared with the rest of corpus as reference corpus) instead of word collocates. This hypothesis will be tested in our future work by a detailed distributional analysis of the specific subtext types.

2. In the EP corpus, where we don't have a particular internal division of the single typically rather short interventions, such effects are not possible. Instead, it is the category of meta-argumentatives that shows a particular pattern (higher knee values than the other categories and non-monotonic curve with a decrease or plateau around the span of 4). If we analyze specifically the difference in the knee value between the spans 3 and 4 we see that frequent meta-argumentative verbs like *concede, reject, endorse* or *conclude* have higher collocation knees for span 4, i.e. maintain the monotonic increase. This seems to be related to frequent local sentence initial patterns like "Connector – subject – (negation) – argumentative verb": *Therefore the Commission does not support*, containing 3 or more words with limited variability (personal pronouns, typical political actors like European Commission, Parliament, parliamentary groups, negations etc.).

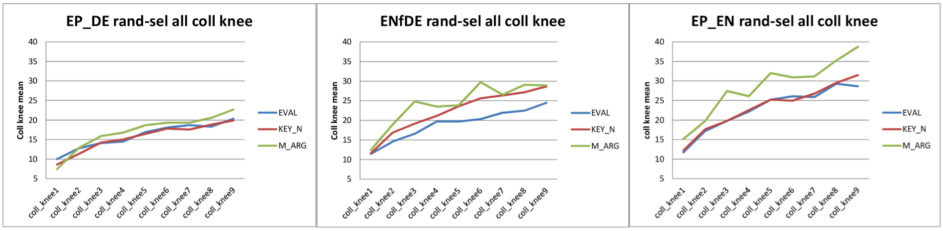


Figure 3: Knee size in Europarl: original DE, EN translated from DE and original EN

On the contrary, adverbs (*clearly*) and meta-argumentative nouns (*stance, contradiction*) with their limited scope and larger diversity of collocates, generate a cumulatively more important decrease of knee value, resulting in the overall non-monotonic pattern.

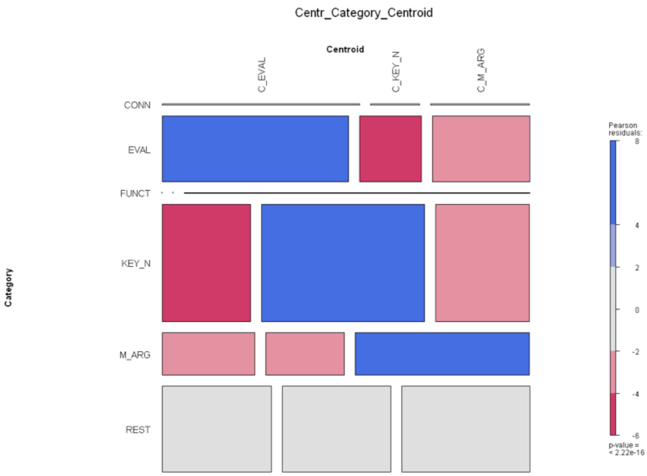
3. The effect concerning the difference between meta-argumentatives and other categories can be observed generally in the English corpora. Figure 3 presents the collocation knee value over spans from 1 to 9 for the original German EP corpus, English translations from German and the English originals for argumentative lexemes annotated in a random selection of ca. 1000 words. Here we see on the one side very similar patterns for the original and translated English corpora, there the meta-argumentatives show the same non-monotonic dynamics. Yet, the knee values are generally lower for the translated corpus and thus more similar to the German originals, possibly due to the general tendency of translation towards reduced variability of linguistic means, which could also reinforce this effect. The more regular pattern of meta-argumentatives in German could be related to the differences in the syntax, in particular the sentence final position of non-finite verb parts (participles and infinitives), as well as finite verbs in subordinate clauses, which reduces the probability of sentence initial pattern building.

4. Experiment on centroid prediction

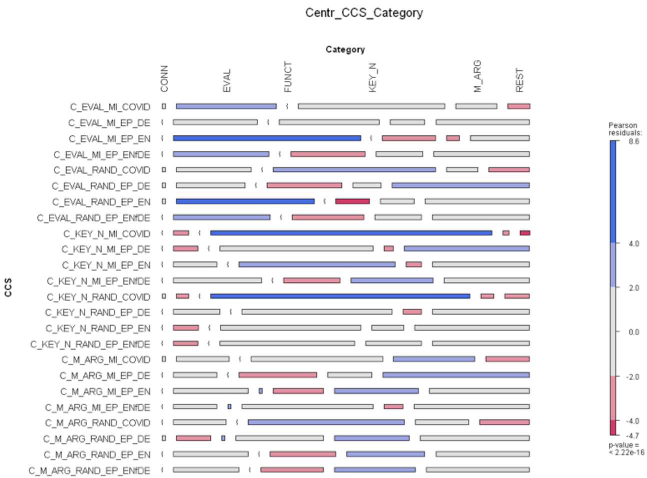
In our further experiment, we are using word embedding models generated with the Python `gensim` package to explore the possibility of semi-automatic identification of argumentatively-relevant lexical items. We use the annotation of the three fully lexical categories (evaluatives/connotated lexemes, key notions and meta-argumentatives) to automatically generate candidate lists and to evaluate the potential improvement in precision as compared to the initial frequencies of lexemes of a given type in our annotations.

centroid	% initial annot.	% correct predict.	improvement abs.	improvement rel.	arg. lex. overall initial annot.	arg. lex. centroid-based predict.
EVAL_MI_COVID	11.4	32	20.6	2.8	44.4	92
EVAL_MI_EP_DE	11.1	27	15.9	2.4	51.9	70
EVAL_MI_EP_EN	19.5	60	40.5	3.1	54.6	81
EVAL_MI_EP_ENfDE	17.2	31	13.8	1.8	49.5	70
EVAL_RAND_EP_DE	9.4	22	12.6	2.3	48.1	55
EVAL_RAND_EP_EN	12.4	45	32.6	3.6	39.5	67
EVAL_RAND_COVID	11.2	24	12.8	2.1	38.5	86
EVAL_RAND_EP_ENfDE	11.6	31	19.4	2.7	39.4	69
KEY_N_MI_COVID	26.2	90	63.8	3.4	44.4	97
KEY_N_MI_EP_DE	35.5	49	13.5	1.4	51.9	60
KEY_N_MI_EP_EN	31.0	50	19.0	1.6	54.6	69
KEY_N_MI_EP_ENfDE	26.3	18	-8.3	0.7	49.5	72
KEY_N_RAND_EP_DE	29.7	48	18.3	1.6	48.1	69
KEY_N_RAND_EP_EN	18.2	44	25.8	2.4	39.5	62
KEY_N_RAND_COVID	18.9	83	64.1	4.4	38.5	91
KEY_N_RAND_EP_ENfDE	20.1	43	22.9	2.1	39.4	69
M_ARG_MI_COVID	6.7	26	19.3	3.9	44.4	85
M_ARG_MI_EP_DE	5.4	14	8.6	2.6	51.9	53
M_ARG_MI_EP_EN	4.1	27	22.9	6.6	54.6	67
M_ARG_MI_EP_ENfDE	6.0	6	0.0	1.0	49.5	62
M_ARG_RAND_EP_DE	9.0	24	15.0	2.7	48.1	63
M_ARG_RAND_EP_EN	8.9	25	16.1	2.8	39.5	70
M_ARG_RAND_COVID	8.4	17	8.6	2.0	38.5	84
M_ARG_RAND_EP_ENfDE	7.7	26	18.3	3.4	39.4	67

Table 1: Improvements with centroid-based prediction for argumentative lexicon.



(a) Main centroid types



(b) Sub-categories

Figure 4: Pearson's residuals of distributions across categories of centroids

Our starting point is built by the annotations of eight selections of ca. 1000 items each from our corpora ({COVID, EP_EN, EP_DE, EP_ENfDE}x{MI-based selection, random selection}), annotated with 6 categories {CONN, EVAL, FUNCT, KEY_N, M_ARG, REST}. Our three fully lexical argumentative categories cover in different corpora between 9.4% and 19.5% (evauatives/connotated lexicon), 18.2% and 35.5% (key notions) and 4.15 and 9% (meta-argumentatives) of the initially annotated items with an overall percentage of argumentative elements ranging from 38.5% (COVID, random selection) to 54.6% (EP_EN, MI-based selection). We built then word embedding models for four corpora with a window of 5 items to the right and to the left, with vector size 100, taking into consideration only items with a frequency over 10. In the next step we calculated the word vectors for the items in our annotated lists of fully lexical argumentative elements, 24 in total (8 selections x 3 categories) and the centroid for each list (defined as the simple mean of word vectors for all lexemes for the list, cf. e.g. Brokos et al., 2016, 114). Then we extracted from the model the first 300 of most similar word to the centroid for every list and conducted an annotation of 100 words for each centroid (2400 items in total) using the categories defined above. Thus, we are trying to use the centroid for, e.g., evaluatives in EP_EN corpus to generate a list of potential candidates of the same category.

As could be easily seen from Table 1, we find an improvement (i.e., higher percentage of items of given category with respect to the initial annotation) for 22 of 24 centroids (besides KEY_N_MI_EP_ENfDE and M_ARG_MI_EP_ENfDE), with some major improvement like 6.6 times for M_ARG_MI_EP_EN, (though with a rather low frequency in the initial annotation), or 4.4 times for KEY_N_RAND_COVID. It is important to notice, that centroid-based identification of argumentative lexical items generates an overall higher percentage of argumentative lexicon (last two columns), that is, the centroid of evaluatives in a given corpus, for example, generates more evaluatives, but also more key notions and meta-argumentatives than functional or non-argumentative fully lexical items, etc. Still we can argue that the method we developed generates category-specific improvements of the candidate list, distinguishing between these three classes, too. We calculated the distribution of the elements annotated in centroid-based lists with respect to the centroid types as well as single centroids, using mosaic plots (Friendly, 1994) (generated using `vcd` package in `R`, Meyer et al., 2022) to represent Pearson residuals of the distribution. Figure 4a shows a very strong statistically significant correlation of generated evaluative lexical items: evaluatives, key notions and meta-argumentatives with their respective centroids, as well as clear or even strong statistically significant negative correlation between any of these categories and the centroids based on the other categories, e.g. between the key notion candidates and the centroid of evaluatives or meta-argumentatives (dark/light blue rectangles mark observed values exceeding the expected value by more than the four-/twofold of the square root of the expected value, red values show the under-representation of the same magnitude). In Figure 4b we can see a more detailed picture for every single configuration of corpus, category and selection procedure.

Here we observe partially mixed results, in particular the generally lower effectiveness of the procedure in ENfDE corpus. This lower performance of the vector model for the translated corpus could be a corollary of the lower number of informative words among the collocates used in words vectors, cf. the discussion in Section 3, point 3. On the other side we observe relevant improvement in the original EP_EN and COVID corpus. Putting it all together, we obtain an increased precision in the identification of candidates for the three classes of argumentatively relevant lexical items, defined above. Thus, our procedure gives us a heuristic to identify items of the given category with higher precision.

5. Discussion

Our results indicate that distributional characteristics of argumentation resources can be modelled with the proposed discourse-level metrics, such as the average size of knee of ranked collocation lists and vector centroids for word embeddings of argumentation lexicon. Specifically, these methods reduce noise in larger collocation spans, which are needed for capturing distributional properties beyond local context.

The dynamics of knee change across different spans indicates that meta-argumentative lexicon has distinct distributional characteristics in comparison with the general lexicon and other types of argumentation resources, (evaluatives, key notions, etc.): for most corpora their curve is flattening or going down in the middle-size spans. A possible linguistic interpretation of this result could be that meta-argumentatives find several 'islands of consistency' both in the local and more distant context within the discourse. This could indicate that meta-argumentatives form part of 'coordinated constructions' (Fillmore et al., 1988), e.g., on the one hand, they are integrated within the local syntactic structure, on the other hand, they have discourse-level valencies that are filled with more distant coordinated argumentation resources. This hypothesis could be experimentally tested, contributing to current research on discourse-level phenomena in construction grammars (e.g., Enghels and Sansiñena, 2021), extending them with construction coordination models for argumentation.

In the translation context, meta-argumentatives are also much stronger influenced by the target language in comparison to other types of argumentation lexicon, possibly because of greater asymmetry in the syntactic structure of English and German clauses, which could have a particular impact on meta-argumentative verbs.

More generally, the proposed methodology highlights interesting distributional characteristics of the argumentation lexicon used in translated corpora. Translations are often found to be influenced by the linguistic structures and usage patterns of the source language on the lexical, syntactic and textual levels, which can be confirmed with the corpus-based statistical analysis (e.g., Baroni and Bernardini, 2006). Such effects are often referred to as 'translationese', and they have been a serious limiting factor for the use of parallel corpora in contrastive linguistics research, as well as for the development and evaluation of modern Machine Translation (MT) systems that are trained on parallel corpora (e.g., Zhang and Toral, 2019; Graham et al., 2020; Vanmassenhove et al., 2021). The comparison of knee changes across collocation spans provides a better understanding of how different types of argumentation resources vary in this respect. Practical applications of this line of research could lead to improvements in MT training and evaluation procedures that will minimize the influence of the source language patterns found in parallel corpora.

Finally, we would like to emphasize that this work is a pilot study with corresponding important limitations. The annotation of categories of argumentative lexical items should in the future be realized by multiple annotators with control of inter-annotator agreement. The possible explanations of the distributional properties of our categories should be tested for other corpora and compared systematically with qualitative evaluation of the uses of argumentative lexical items in different types of corpora.

Bibliography

Atayan, Vahram. Argumentationsstrukturen–ein Äquivalenzparameter bei der Übersetzung. *Multiperspektivische Fragestellungen der Translation in der Romania. Hommage an Wolfram Wilss zu seinem*, 80:61–93, 2007.

Babych, Bogdan, Anthony Hartley, Serge Sharoff, and Olga Mudraya. Assisting translators in indirect lexical transfer. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 136–143, 2007.

Baroni, Marco and Silvia Bernardini. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3):259–274, 2006. doi: 10.1093/llc/fqi039.

Brokos, Georgios-Ioannis, Prodromos Malakasiotis, and Ion Androutsopoulos. Using Centroids of Word Embeddings and Word Mover's Distance for Biomedical Document Retrieval in Question Answering. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, pages 114–118, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/W16-2915. URL `https://aclanthology.org/W16-2915`.

Enghels, Renata and María Sol Sansiñena. Discourse-level phenomena in construction grammars. *Constructions and Frames*, 13(1):3–20, 2021. doi: 10.1075/cf.00045.int.

Evert, Stefan. Corpora and collocations. In Lüdeling, Anke and Merja Kytö, editors, *Corpus Linguistics. An International Handbook*, chapter 58, pages 1212–1248. Mouton de Gruyter, Berlin, 2008. URL `http://purl.org/stefan.evert/PUB/Evert2007HSK_extended_manuscript.pdf`.

Fillmore, Charles J, Paul Kay, and Mary Catherine O'connor. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, pages 501–538, 1988. doi: 10.2307/414531.

Friendly, Michael. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89(425):190–200, 1994. doi: 10.1080/01621459.1994.10476460.

Gast, Volker. Comparing Annotation Types and n-Gram Sizes. In Schützler, Ole and Julia Schlüter, editors, *Data and methods in corpus linguistics: Comparative approaches*, chapter 11, pages 323–352. Cambridge University Press, 2022. doi: 10.1017/9781108589314.012.

Graham, Yvette, Barry Haddow, and Philipp Koehn. Statistical power and translationese in machine translation evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 72–81, 2020. doi: 10.18653/v1/2020.emnlp-main.6.

Hoey, Michael. Lexical priming and translation. *Kruger, Alet, Kim Wallmach & Jeremy Munday (eds.)*, pages 153–168, 2011.

House, Juliane. Using translation and parallel text corpora to investigate the influence of global English on textual norms in other languages. *Corpus-based translation studies: Research and applications*, pages 187–208, 2011.

Koehn, Philipp. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x: papers*, pages 79–86, 2005.

Kranich, Svenja and Volker Gast. Explicitness of epistemic modal marking: Recent changes in British and American English. *Thinking Modality: English and Contrastive Studies of Modality. Newcastle upon Tyne: Cambridge Scholars Publishing*, pages 3–22, 2015.

Kruger, Alet, Kim Wallmach, and Jeremy Munday. *Corpus-based translation studies: Research and applications*. Bloomsbury Publishing, 2011.

Meyer, David, Achim Zeileis, Kurt Hornik, Florian Gerber, and Michael Friendly. *vcd: Visualizing Categorical Data*, 2022. URL `https://CRAN.R-project.org/package=vcd`. R package version 1.4-10.

Satopaa, Ville, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops*, pages 166–171. IEEE, 2011. doi: 10.1109/ICDCSW.2011.20.

Schmid, Helmut. Probabilistic part-ofispeech tagging using decision trees. In *New methods in language processing*, page 154, 2013.

van Eemeren, Frans H. *Argumentation theory: A pragma-dialectical perspective*. Springer, 2018. doi: 10.1007/978-3-319-95381-6.

van Eemeren, Frans H, Rob Grootendorst, and Tjark Kruiger. Handbook of argumentation theory. In *Handbook of Argumentation Theory*. De Gruyter Mouton, 2019. doi: 10.1515/9783110846096.

Vanmassenhove, Eva, Dimitar Shterionov, and Matthew Gwilliam. Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, 2021. doi: 10.18653/v1/2021.eacl-main.188.

Wang, Lucy Lu, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, et al. Cord-19: The covid-19 open research dataset. *ArXiv*, 2020.

Zhang, Mike and Antonio Toral. The Effect of Translationese in Machine Translation Test Sets. In *Proceedings of the Fourth Conference on Machine Translation* (*Volume 1: Research Papers*), pages 73–81, 2019. doi: 10.18653/v1/W19-5208.

Address for correspondence:
Vahram Atayan
`atayan@uni-heidelberg.de`
Institute for Translation and Interpreting
Heidelberg University, Plöck 57a, Heidelberg 69117, Germany