Exploration of Inter- and Intralingual Variation of Discourse Phenomena

Ekaterina Lapshinova-Koltunski
Saarland University
FR4.6, University Campus, D-66123 Saarbrücken
e.lapshinova@mx.uni-saarland.de

Abstract

In this paper, we analyse cross-linguistic variation of discourse phenomena, i.e. coreference, discourse relations and modality. We will show that contrasts in the distribution of these phenomena can be observed across languages, genres, and text production types, i.e. translated and non-translated ones. Translations, regardless of the method they were produced with, are different from their source texts and from the comparable originals in the target language, as it was stated in studies on *translationese*. These differences can be automatically detected and analysed with exploratory and automatic clustering techniques. The extracted frequency-based profiles of variables under analysis (languages, genres, text production types) can be used in further studies, e.g. in the development and enhancement of MT systems, or in further NLP applications.

1 Introduction

Although considerable research aiming at enhancing machine-translated texts with discourse properties achieved positive results in recent years, see e.g. (Webber et al., 2013; Hardmeier, 2014) or (Meyer et al., 2015), some document-wide properties of automatically translated texts still require improvement, as translation models are induced from stand-alone pairs of sentences. Moreover, target language models approximate the target language on the string level only, whereas target texts have properties that go beyond those of their individual sentences and that reveal themselves in the frequency and distribution of certain structures. These frequency- and distribution-based properties of translated and non-translated texts are in focus of corpus-based translation studies. However, these properties (in form of higher-level language models) may also be useful for natural language processing (NLP), including machine translation (MT).

In this paper, we show an example of a corpus-based analysis of interlingual (between English and German) and intralingual (across different genres) variation of discourse properties in translated and non-translated texts. In particular, this paper will focus on various types of discourse relational devices, pronominal referring expressions, as well as modal meanings expressed with particular modal verbs. The frequencies of these discourse features will be automatically extracted from English-German comparable corpora which also contain multiple translations produced with several methods, including manual and automatic ones. We will compare the distributions of these features in both languages, as well as in translations from English to German, paying attention to their variation across genres available in the dataset. We will also consider differences in their distributions in human and machine translation. For our analysis, we apply exploratory and unsupervised classification techniques. The obtained information on the frequency-based interlingual and intralingual differences may be valuable for linguistic studies on language contrasts, human translation, and may find application in NLP and especially MT.

2 Related Work

2.1 Discourse properties in English and German

Various discourse phenomena have been in focus of several translation studies and those on language contrasts dealing with English and German. Recent years have seen an increase in the number of works employing corpus-based methods for their analysis. However, multilingual stud-

Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT), pages 158–167,
Lisbon, Portugal, 17 September 2015.

ies are mostly concerned with individual phenomena in particular genres, see e.g. (Bührig and House, 2004) for particular cohesive conjunctions or adverbs in prepared speeches, (Zinsmeister et al., 2012) for abstract anaphora in parliament debates, and (Taboada and Gómez-González, 2012) for particular coherence relations. The latter, however, considers two modes: spoken and written, and states that the differences between modes are more prominent than between languages. Kunz and Lapshinova-Koltunski (2015) and Kunz et al. (2015) show that distributions of different discourse phenomena are not only mode- but also genre-dependent. The authors show this for a number of textual phenomena, analysing structural and functional subtypes of coreference, substitution, discourse connectives and ellipsis. Their dataset includes several genres, and they are able to identify contrasts and commonalities across languages (English and German) and genres with respect to the subtypes of all textual phenomena under analysis, showing that these languages differ as to the degree of variation between individual genres. Moreover, there is more variation in the realisation of discourse devices in German than English. The authors attested the main differences in terms of preferred meaning relations: a preference for explicitly realising logico-semantic relations by discourse markers and a tendency to realise relations of identity by coreference. Interestingly, similar meaning relations are realised by different subtypes of discourse phenomena in different languages and genres.

2.2 Discourse properties in human and machine translation

Cross-lingual contrasts stated on the basis of non-translated data are also of great importance for translation. Kunz et al. (2015) suggest preferred translation strategies on the basis of contrastive interpretations for the results of their quantitative analysis, which show that language contrasts are even more pronounced if we compare languages per genre. These contrasts exist in the features used for creating textual relations. Therefore, they suggest that, for instance, when translating popular science texts from English into German translators should more extensively use linguistic means expressing textual relations. Overall, they claim that translators should use more explicit devices translating from English into German, e.g. demonstrative pronouns should be used more often instead of personal pronouns (e.g. *dies/das* instead of *es/it*). The opposite translation strategies should be used when translating from German to English.

However, studies of translated language show that translators do not necessarily apply such strategies. For instance, Zinsmeister et al. (2012) demonstrate that translations in general tend to preserve the source language anaphor's categories, functions and positions, which results in the *shining through* effect (shining through of the source language preferences, see (Teich, 2003)) in both translation directions. Additionally, due to the tendency to explicate textual relations, translators tend to use more nominal coreference instead of pronominal one. *Explicitation* (tendency of translations to be more explicit than their sources, see (Vinay and Darbelnet, 1958) and (Blum-Kulka, 1986)) along with *shining through* belong to the characteristics of translated texts caused by peculiarities of translation process. A number of works on discourse connectives, e.g. (Becher, 2011; Bisiada, 2014; Meyer and Webber, 2013) and (Li et al., 2014), show implicit/explicit discourse expression divergence in both human and machine translation. There are several studies that attempt to incorporate information on discourse relations or other discourse properties into MT, see for instance, those by Le Nagard and Koehn (2010), Hardmeier and Federico (2010) and Guillou (2012), or those presented within the first DiscoMT workshop, see (Webber et al., 2013). Most of them employ parallel corpora, thus, the approximation of the target language is based on translations, which, however, possess characteristics that differ them from non-translated texts originally written in a target language, also in terms of discourse properties. This paper will consider discourse-related characteristics that differ translation from non-translated texts, and also differentiate human from machine translations.

3 Methodology

3.1 Data

As we focus on variation of discourse phenomena in English and German, as well as English-German translations, our data should contain both English-German parallel texts and non-translated comparable texts in German. Furthermore, as we are also interested in linguistic variation in terms of genre, the texts should be from different genres. For this reason, we had to dismiss the typical corpora used in MT, e.g. Europarl (Koehn, 2005) or TED talks, as translated texts in these resources are not comparable. The latter contains multilingual subtitles which are produced under different restrictions than those of translations. We also expect that some of the phenomena under analysis might be omitted in the subtitles, as this is recommended in the guidelines[1]. So, we select two corpora which contain English-German parallel and comparable texts from different genres. English and German originals (EO and GO) were extracted from CroCo (Hansen-Schirra et al., 2012), whereas German translations originate from the VARTRA corpus (Lapshinova-Koltunski, 2013), as it contains multiple translations of the CroCo English originals produced both manually and automatically (HU and MT).

The whole dataset totals 406 texts which cover seven genres: political essays (ESS), fictional texts (FIC), instruction manuals (INS), popular-scientific articles (POP), letters to shareholders (SH), prepared political speeches (SP), and tourism leaflets (TOU). The decision to include this wide range of genres is justified by the need for heterogeneous data for our experiment. The number of words per genre in comprises ca. 36 thousand tokens. We tag both English and German data with the TreeTagger tools (Schmid, 1994).

3.2 Feature selection

Linguistic relations between textual elements help recipients in their cognitive interpretation as to how different thematic concepts are connected. These relations are indicated by particular structures that language producers employ, e.g. grammatical items such as connectives, personal and demonstrative pronouns, substitute forms, elliptical constructions and lexical items, such as nouns, verbs and adjectives. As already mentioned in Section 1 above, we will analyse discourse relations, coreference and modality.

For discourse relations, we will analyse connectives classified according to the semantic relations they convey. Our classification is based on semantic relations defined by Halliday and Hasan (1976) and includes additive (relation of addition, e.g. *and, in addition, moreover*), adversative (relation of contrast/alternative, e.g. *yet, although, by contrast*), causal (relation of causality/dependence, e.g. *because, therefore, that's why*), temporal (temporal relation between events such as *after, afterwards, at the same time*) and modal relations (expressing rather a pragmatic meaning, in which evaluation of the speaker is involved, e.g. *unfortunately, surely*).

Demonstrative and personal pronouns (such as *this, that, she, his, theirs, it*, etc.) will serve as triggers of coreference. We also consider distributions of general nouns, e.g. *plan, case, fact*, which commonly function as abstract anaphora (Zinsmeister et al., 2012). For the analysis of modality, we consider frequencies of modal verbs grouped according to the modal meanings defined by Biber et al. (1999): permission (*can/could, may/might*), volition (*will, would, shall*) and obligation (*must, ought to, should, need to, have got to, suppose to*).

feature pattern	**discourse property**
permission obligation volition	modality
additive adversative causal temporal modal	discourse relations
general.nouns perspron dempron	coreference

Table 1: Features under analysis

The set of 11 selected features is outlined in Table 1. The first column denotes the extracted and analysed feature patterns, the second represents the corresponding discourse property. For the extraction of the frequencies of these feature patterns, we use a number of regular expressions based on string, part-of-speech and chunk tags, as well as further constraints, e.g. position in a sentence or in a text. Frequency information is collected both per text, and per subcorpus (e.g. per genre in a certain language).

3.3 Methods

For our analysis, we use exploratory and also unsupervised classification (automatic clustering) techniques which will allow us to observe differences between groups of texts and subcorpora, and also to discriminate between them on the basis of discourse features described in Section 3.2.

We apply correspondence analysis (CA) (Venables and Smith, 2010; Baayen, 2008; Greenacre,

[1] See the subtitling guidelines `http://translations.ted.org/wiki/How_to_Compress_Subtitles`

2007) that is conceptually similar to principal component analysis (PCA), with the difference that the data is scaled so that rows and columns are treated equivalently. Thus, this technique will help us to see not only which variables (e.g. languages or genres) have similarities, but also possible correlation of these variables with discourse features contributing to these similarities, as distances between between dependent and independent variables are calculated. These distances are then represented in a two-dimensional map, and the larger the differences between subcorpora or texts, the further apart they are on the map. Likewise, dissimilar categories of discourse phenomena are further apart. Proximity between subcorpora and discourse features in the merged map is as good an approximation as possible of the correlation between them. In computing this low-dimensional approximation, CA transforms the correlations between rows and columns of our table into a set of uncorrelated variables, called principal axes or dimensions. These dimensions are computed in such a way that any subset of *k* dimensions accounts for as much variation as possible in one dimension, the first two principal axes account for as much variation as possible in two dimensions, and so on. In this way, we can identify new meaningful underlying variables, which ideally correlate with such variables as language or genre, indicating the reasons for the similarities or differences between these subcorpora. The length of the arrows in the graph indicates how pronounced a discourse feature is, see (Jenset and McGillivray, 2012) for details. The position of the points in relation to the arrows indicates the relative importance of a feature for a subcorpus. The arrows pointing in the direction of an axis indicate a high correlation with the respective dimension, and thus, a high contribution of the feature to this dimension.

The results of automatic clustering will indicate differences and similarities between the languages (English and German) and their varieties (genres). Moreover, we can also discover differences between non-translated and (manually or automatically) translated texts. We decide for unsupervised techniques, in favour of different genres contained in our data, and supervised classification performs better with single genre data, so that in a supervised scenario, we would need to perform several classification tasks. We apply *hierarchical cluster analysis* (HCA), see (Hothorn and Everitt, 2014) and (Everitt et al., 2011). This clustering technique is connectivity-based as its core idea is that objects are more related to nearby objects than to objects farther away. Objects, in our case texts and subcorpora, are connected to form clusters based on their distance measured here on the basis of the feature distributions. We calculate the distance by the Euclidean distance which is one of the most straightforward and generally accepted ways of computing distances between objects in a multi-dimensional space. The results of hierarchical clusters are represented graphically in a dendrogram, which is a branching diagram that represents the relationships of similarity among a group of entities. The arrangement of the branches tells us which texts/subcorpora (on leaves) are most similar to each other. The height of the branch points indicates how similar or different they are from each other. Ward's method (also called Ward's minimum variance method) is employed to perform clustering. This method minimises the total within-cluster variance after merging.

The main drawback of this technique is that the number of clusters needs to be specified in advance. Therefore, we apply a technique based on bootstrap resampling, with the help of which we are able to produce *p-value*-based clusters, i.e. that are highly supported by the data will have large *p-values*[2]. The output dendrogram demonstrates two types of p-values: AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) value. AU p-value, which is computed by multi-scale bootstrap resampling, is a better approximation to unbiased p-value than BP value computed by normal bootstrap resampling.

4 Analyses

4.1 Discourse properties in English and German

First, we analyse English and German non-translated texts, to define the differences between these languages in terms of discourse properties. We perform CA on the subset of data containing originals only. In the first step, the dataset is labelled with text IDs only (e.g. EO_001, GO_010, etc.).

In Table 2, we present the Eigenvalues calculated for each dimension to assess how well our

[2]We use pvclust() package available in the R environment (version 3.0.2; (Team, 2013)).

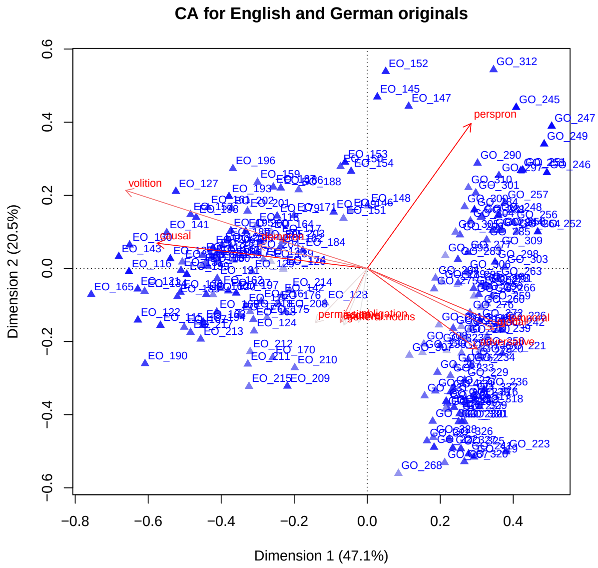

Figure 1: Variation of discourse phenomena across languages

data is represented in the graph[3]. The cumulative value for dimensions allow us to analyse how well our data is represented in the graph.

dim	value	%	cum%
1	0.109830	47.1	47.1
2	0.047842	20.5	67.6
3	0.018943	8.1	75.7
4	..		
Total:	0.233192	100.0	

Table 2: Contribution of dimensions for variation across languages

We plot the results in a two-dimensional graph in Figure 1, representing the first two dimensions, which explain 67.60% (cumulative value) of the data inertia. The second dimension although covering only 20,50% is also important for our analysis if we want to explain more than 50% of the data variation. The rest of inertia remain unexplained with the two-dimensional representation[4].

Concerning dimension 1 (47,10% of inertia), we see a clear distinction between English and German texts (along the x-axis on the left and on the right from zero respectively). So, the distinction along this dimension reflects language contrasts in the use of particular discourse features, i.e. different types of discourse relations via connectives for German, and coreference via demonstrative pronouns, modal meaning of volition and causal logico-semantic relations for English. The assumption is that the second dimension indicates distinction between genres available in our dataset, which is not seen in the data labelled with text IDs only.

For the sake of the visualisation of results, we perform the same analyses labelling our dataset with genres, and also reducing it to subcorpora corresponding to different genres and languages (e.g. EO_ESS containing all texts of English political essays, etc.), see the resulting plot in Figure 2.

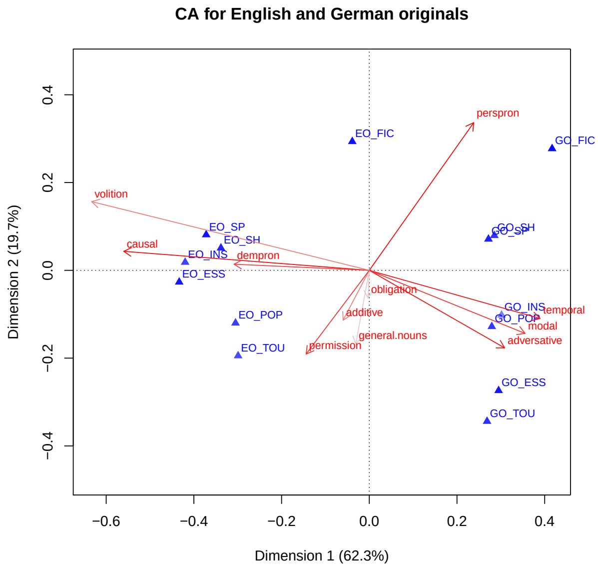

Figure 2: Variation of discourse phenomena across genres

This time, we achieve a cumulative value of 82%, with the first dimension covering over 60% of the data variance, see Table 3.

dim	value	%	cum%
1	0.103453	62.3	62.3
2	0.032665	19.7	81.9
3	0.012870	7.7	89.7
4	..		
Total:	0.166179	100.0	

Table 3: Contribution of dimensions for variation across genres

As in the previous graph, this dimension still indicates language contrasts in the dataset, with the same features contributing to these differences. The second dimension (the *y*-axis) clearly indicates language-independent differences in genres:

[3]'dim' lists dimensions, 'value' – Eigenvalues converted to percentages of explained variation in '%' and calculated as cumulative explained variation with the addition of each dimension in 'cum'.

[4]This means that we are not able to explain ca. 30% of the variation in our data, which might indicate differences to further parameters, e.g. according to individual authors or translators.

tourism, essays and popular-scientific texts grouping together below zero (with additives, modality and general nouns as features), and fiction, political speeches and letters to shareholders above zero. The features of instruction manuals seem to be language-dependent, as the English and the German INS subcorpora are positioned on the opposite axis sides. Fictional texts of both languages are positioned at the edge of the genre axis, with personal pronouns contributing to this grouping, which coincides with the results obtained by Kunz and Lapshinova-Koltunski (2015) and Kunz et al. (2015) showing that fiction is best distinguished from the other genres for both languages with supervised classification techniques.

Automatic clustering deliver similar results, see Figure 3, with the exception of English fictional texts, which are classified along with the German fictional texts into the cluster of German subcorpora.

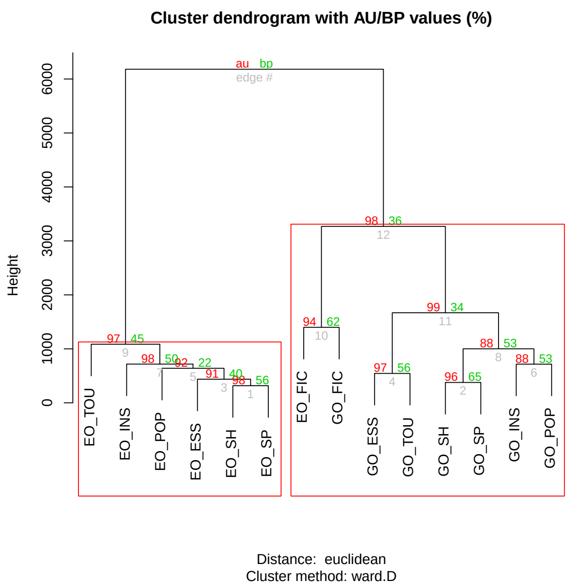

Figure 3: Classification of English and German subcorpora

4.2 Originals and translations

In the next step, we include translated texts into our analysis. The translation data is labelled with HU and MT, indicating manual or automatic method of translation, whereas digits indicate translation variants. Thus, MT1 and MT2 are produced with two different SMT systems, and HU1 and HU2 were produced by two different groups of translators. The results of the bootstrap resampling[5] suggests two classes in our data, illustrated in Figure 4.

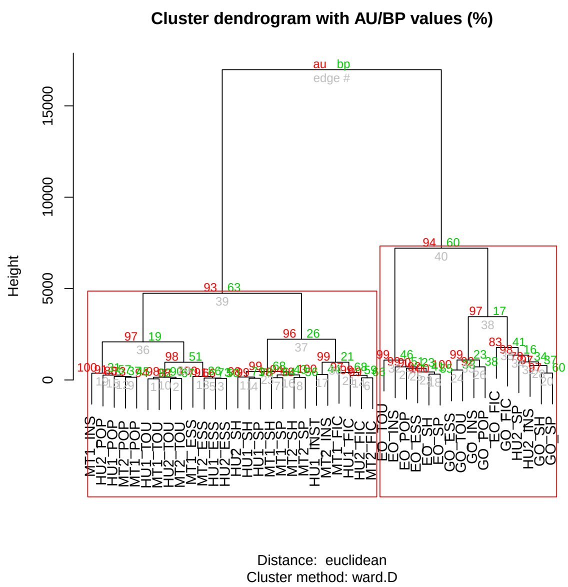

Figure 4: Classification of originals and translations

As seen from the graph, our dataset is clustered into originals (on the right side) and translations (on the left side), which is apparently the most prominent difference in this data. This coincides with the statements of the theory of *translationese*, see (Gellerstam, 1986) or (Baker, 1993), that translations have their specific feature differing them from the source texts and comparable originals in the target language. A number of studies have shown that these features can be used to automatically discriminate between translated and non-translated texts, such as (Baroni and Bernardini, 2006; Ilisei et al., 2010; Koppel and Ordan, 2011). Our results show that this discrimination is also possible with discourse features, which means that translations differ from originals also in these properties.

The only exceptions in our results are manually produced translations of political speeches (HU2-SP) and instruction manuals (HU2-INS) classified together with political speeches and letters to shareholders originally written in German. Most of the smaller clusters within the bigger 'non-translated' class are grouped rather according to languages than genres, e.g. political essays, tourism texts, manuals and popular-scientific arti-

[5]We achieve a good classification performance with an average error rate of 0,06.

cles.

Next, we want to prove if the observed difference between originals and translations is dependent on the source or the target language (which would indicate the phenomenon of shining through or normalisation). For this reason, we perform two classification experiments applying the same clustering technique and including German translation data and their English sources in the first experiment (Figure 5), and the same German translations together with German comparable non-translated texts in the second (Figure 6). The results show that in both cases, the data is separated into translations and originals, with the same two subcorpora as exceptions. So, no shining through/normalisation effect can be detected.

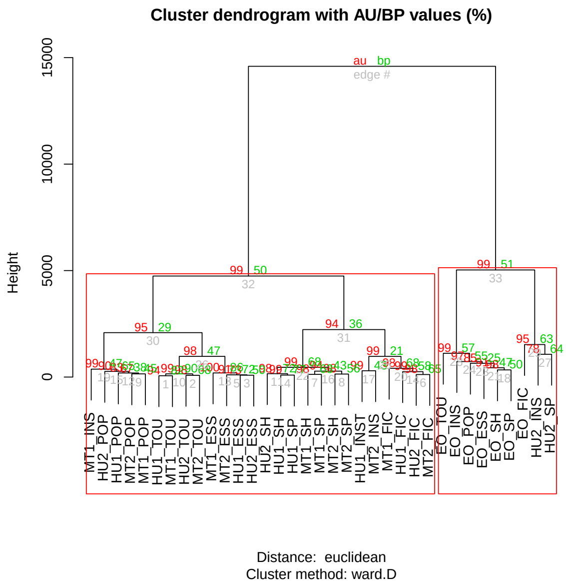

Figure 5: German translations and non-translated English source texts

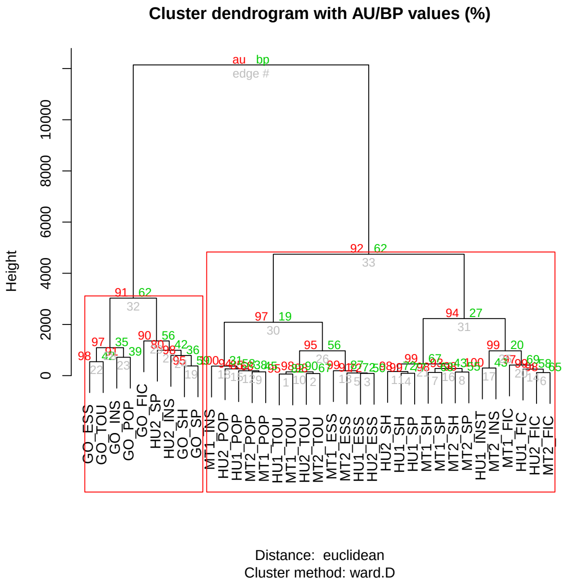

Figure 6: German translations and comparable German non-translated texts

4.3 Human and machine translations

Finally, we perform classification on the data subset containing translations only. The resulting dendrogram in Figure 7 reveals four heterogeneous classes of translations, all containing both manually and automatically produced outputs. The two human translations that were classified with the non-translated data in previous experiments in Section 4.2 form a cluster on their own. This is the only cluster containing one type of translations in the whole data subset. The other three clusters consist of a mixture of human and machine translations. They presumably form genre-sensitive clusters, as we observe groupings of translations of the same genres on smaller cluster nodes.

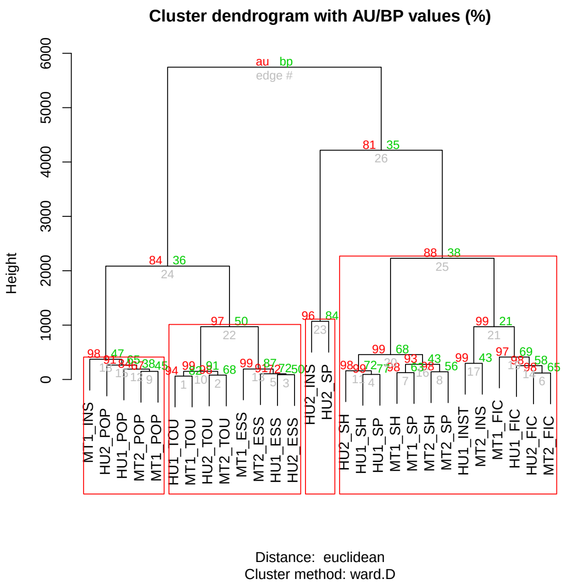

Figure 7: Human and machine translations

On the one hand, this suggests that genre is more prominent than translation method, i.e. there are more differences between various genres than between human and machine translations in the data under analysis, if discourse properties are concerned. On the other hand, the results may also indicate that discourse features are more informative in genre classification than in the distinction into human vs. machine. Similar results were shown by Zampieri and Lapshinova-Koltunski (2015) who were able to achieve better results in the classification between genres than between translation methods, operating with delexicalised n-grams and using supervised classification techniques. Therefore, we claim that the distributions of the discourse features under analysis are genre-dependent, which coincides with the results of the previous analyses within a number of multilingual genre studies.

As seen in the analyses above (see Figures 4, 5, 6 and 7), political speeches and letters to shareholders are always clustered together in translated data. Similar observations were also made in (Lapshinova-Koltunski, inpress) for a different set of features. According to Neumann (2013), these two registers seem to be closer in English than in German, and so, their commonalities in our translation data might indicate the influence of the source texts. However, CA performed on German and English originals reveal that these register are similar not only within each language, but also cross-lingually, as they are situated on the same level of the y-axis, see Figure 2. As a result, translations also reveal these similarities.

5 Conclusion and Discussion

We have demonstrated an example of a corpus-based analysis of discourse properties in a multilingual dataset which contains both translated and non-translated texts, using exploratory and automatic clustering techniques. The results show that discourse-related features vary depending on the languages and genres involved. Languages, even such closely related ones as English and German, have different preferences in the usage of discourse properties, which are also prone to interlingual variation in terms of genres. This knowledge on contrasts will be valuable not only for contrastive linguistics and translation studies, but also for natural language processing including statistical MT, as it is available in form of frequency-based information and can be used for language models. The observed variation of discourse properties is also influenced by the nature of the texts (translated vs. non-translated). Both human and machine translations have constellations of discourse properties different from those of their underlying originals, and from comparable non-translated texts in the target language.

Comparing machine-translated texts with those translated by humans, we stated that genre-membership of translations determines more prominent differences between them than the methods they were translated with (manual vs. automatic). This points to the fact that machine translations resemble rather human translations than non-translated texts in both the source and the target languages, if discourse features are considered. On the one hand, this confirms the hypothesis of *levelling out* indicating that individual translated texts are more alike than individual original texts, in both source and target languages[6]. On the other hand, our results conform to those obtained by Rabinovich and Wintner (2015) who show that multi-genre data is more difficult to be classified with translationese (translation-specific) features.

Furthermore, the results seem to contradict the findings in (Guzman et al., 2014), which used discourse information to develop automatic MT evaluation metrics. However, we believe that the differences in the outcome are caused by the nature of the dataset: translations in the present study originate from multiple genres, whereas Guzman et al. (2014) use news texts only. Intralingual variation in both English and German imply that if a model is applicable for a certain genre in one language, it is not necessarily applicable to a different genre of the same language, as the distributions of the underlying phenomena differ (sometimes) tremendously.

The contrasts between translated and non-translated texts suggest that we need more research on how to incorporate discourse-based language models induced from comparable and not parallel data. In this way, we might achieve a closer approximation of machine translation to non-translated texts in a target language. This is relevant not only for the development of machine translation systems but also for their evaluation, as the similarities between a reference and an MT output might be confounding in the quality judgement, if discourse phenomena are concerned. In the future, experiments could be planned that apply the present results for the development and evaluation of MT. Moreover, it would be interesting to learn if the differences between translated and original text affect perception of the quality of the text, for which experiments involving human judgements are required.

[6]Variation in individual translators is not considered.

References

R. Harald Baayen. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press.

Mona Baker. 1993. Corpus linguistics and translation studies: Implications and applications. In G. Francis Baker M. and E. Tognini-Bonelli, editors, *Text and Technology: in Honour of John Sinclair*, pages 233–250. Benjamins, Amsterdam.

Marco Baroni and Silvia Bernardini. 2006. 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.

Viktor Becher. 2011. *Explicitation and implicitation in translation. A corpus-based study of English-German and German-English translations of business texts*. Ph.D. thesis, Universität Hamburg.

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Longman, Harlow.

Mario Bisiada. 2014. Lösen sie Schachtelsätze möglichst auf: The impact of editorial guidelines on sentence splitting in german business article translations. *Applied Linguistics*, 3.

Shoshana Blum-Kulka. 1986. Shifts of cohesion and coherence in translation. In Juliane House and Shoshana Blum-Kulka, editors, *Interlingual and intercultural communication*, pages 17–35. Gunter Narr, Tübingen.

Kristin Bührig and Juliane House. 2004. Connectivity in translation: Transitions from orality to literacy. In J. House and J. Rehbein, editors, *Multilingual Communication*, pages 87–114. Benjamins, Amsterdam.

Brian S. Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. 2011. *Cluster Analysis*. Wiley series in probability and statistics. Wiley.

Martin Gellerstam. 1986. Translationese in Swedish novels translated from English. In L. Wollin and H. Lindquist, editors, *Translation Studies in Scandinavia*, pages 88–95. CWK Gleerup, Lund.

Michael J. Greenacre. 2007. *Correspondence analysis in practice*. Chapman & Hall/CRC, Boca Raton.

Liane Guillou. 2012. Improving pronoun translation for statistical machine translation. In *EACL 2012, 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, April 23-27, 2012*, pages 1–10.

Francisco Guzman, Shafiq Joty, Lluís Màrquez, and Preslav Nakov. 2014. Using discourse structure improves machine translation evaluation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 687–698. Association for Computational Linguistics.

M.A.K. Halliday and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman, London, New York.

Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner. 2012. *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. de Gruyter, Berlin, New York.

Christian Hardmeier and Marcello Federico. 2010. Modelling pronominal anaphora in statistical machine translation. In *IWSLT (International Workshop on Spoken Language Translation); Paris, France; December 2nd and 3rd, 2010.*, pages 283–289.

Christian Hardmeier. 2014. *Discourse in statistical machine translation*. Ph.D. thesis, Uppsala: Acta Universitatis Upsaliensis.

Torsten Hothorn and Brian S. Everitt. 2014. *A Handbook of Statistical Analyses Using R*. Chapman & Hall/CRC Press, 3rd edition.

Iustina Ilisei, Diana Inkpen, Gloria Corpas Pastor, and Ruslan Mitkov. 2010. Identification of translationese: A machine learning approach. In *Computational linguistics and intelligent text processing*, pages 503–511. Springer Berlin Heidelberg.

Gard B. Jenset and Barbara McGillivray. 2012. Multivariate analyses of affix productivity in translated english. In Michael P. Oakes and Meng Ji, editors, *Quantitative Methods in Corpus-Based Translation Studies*, pages 301–324. John Benjamins.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Conference Proceedings: the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. AAMT.

Moshe Koppel and Noam Ordan. 2011. Translationese and its dialects. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011)*, June.

Kerstin Kunz and Ekaterina Lapshinova-Koltunski. 2015. Cross-linguistic analysis of discourse variation across registers. *Special Issue of Nordic Journal of English Studies*, 14(1):258–288.

Kerstin Kunz, Stefania Degaetano-Ortlieb, Ekaterina Lapshinova-Koltunski, Katrin Menzel, and Erich Steiner. 2015. Gecco – an empirically-based comparison of english-german cohesion. In G. De Sutter, I. Delaere, and M.-A. Lefer, editors, *New Ways of Analysing Translational Behaviour in Corpus-Based Translation Studies*. Mouton de Gruyter. TILSM series.

Ekaterina Lapshinova-Koltunski. 2013. Vartra: A comparable corpus for analysis of translation variation. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora*, pages 77–86, Sofia, Bulgaria, August. Association for Computational Linguistics.

Ekaterina Lapshinova-Koltunski. *in press*. Linguistic features in translation varieties: Corpus-based analysis. In G. De Sutter, I. Delaere, and M.-A. Lefer, editors, *New Ways of Analysing Translational Behaviour in Corpus-Based Translation Studies*. Mouton de Gruyter. TILSM series.

Ronan Le Nagard and Philipp Koehn. 2010. Aiding pronoun translation with co-reference resolution. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, WMT '10, pages 252–261, Stroudsburg, PA, USA. Association for Computational Linguistics.

Junyi Jessy Li, Marine Carpuat, and Ani Nenkova. 2014. Cross-lingual discourse relation analysis: A corpus study and a semi-supervised classification system. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 577–587, Dublin, Ireland, August 23-29.

Thomas Meyer and Bonnie Webber. 2013. Implicitation of discourse connectives in (machine) translation. In *Proceedings of the Workshop on Discourse in Machine Translation*, pages 19–26, Sofia, Bulgaria, August. Association for Computational Linguistics.

Thomas Meyer, N. Hajlaoui, and Andrei Popescu-Belis. 2015. Disambiguating discourse connectives for statistical machine translation. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 23(7):1184–1197, July.

Stella Neumann. 2013. *Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German*. De Gruyter Mouton, Berlin, Boston.

Ella Rabinovich and Shuly Wintner. 2015. Unsupervised identication of translationese. *Transactions of the Association for Computational Linguistics*, 3.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *International Conference on New Methods in Language Processing*, pages 44–49, Manchester, UK.

Maite Taboada and MLA Gómez-González. 2012. Discourse markers and coherence relations: Comparison across markers, languages and modalities. *Linguistics and the Human Sciences*, 6 (1-3):17–41.

R Core Team. 2013. R: A language and environment for statistical computing. Technical report, R Foundation for Statistical Computing, Vienna, Austria.

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

William N. Venables and David M. Smith. 2010. *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics*.

Jean P. Vinay and Jean Darbelnet. 1958. *Stylistique Comparée du Français et de l'Anglais. Méthode de Traduction*. Didier, Paris.

Bonnie Webber, Andrei Popescu-Belis, Katja Markert, and Jörg Tiedemann, editors. 2013. *Proceedings of the Workshop on Discourse in Machine Translation*. Association for Computational Linguistics, Sofia, Bulgaria, August.

Marcos Zampieri and Ekaterina Lapshinova-Koltunski. 2015. Investigating genre and method variation in translation using text classification. In Petr Sojka, Ales Horák, Ivan Kopecek, and Karel Pala, editors, *Text, Speech and Dialogue - 18th International Conference, TSD 2015, Plzen, Czech Republic, Proceedings*, Lecture Notes in Computer Science. Springer.

Heike Zinsmeister, Stefanie Dipper, and Melanie Seiss. 2012. Abstract pronominal anaphors and label nouns in german and english: selected case studies and quantitative investigations. *Translation: Computation, Corpora, Cognition*, 2(1).