

# Chapter 9

## Analysing the dimension of mode in translation

Ekaterina Lapshinova-Koltunski

Universität des Saarlandes

The present chapter applies text classification to test how well we can distinguish between texts along two dimensions: a text-production dimension that distinguishes between translations and non-translations (where translations also include interpreted texts); and a mode dimension that distinguishes between spoken and written texts. The chapter also aims to investigate the relationship between these two dimensions. Moreover, it investigates whether the same linguistic features that are derived from variational linguistics contribute to the prediction of mode in both translations and non-translations. The distributional information about these features was used to statistically model variation along the two dimensions. The results show that the same feature set can be used to automatically differentiate translations from non-translations, as well as spoken texts from the written texts. However, language variation along the dimension of mode is stronger than that along the dimension of text production, as classification into spoken and written texts delivers better results. Besides, linguistic features that contribute to the distinction between spoken and written mode are similar in both translated and non-translated language.

### 1 Introduction

In the present contribution, we analyse translation as a product which possesses a number of linguistic characteristics expressed in its linguistic features. These linguistic features make translation look different from other language products. Translation variation is influenced by various dimensions such as language, register, text production and expertise (Lapshinova-Koltunski 2017). They are related



to the constraint dimensions as defined by Kotze (2019: p. 346) who sees translation as a constrained language variety. These varieties are probabilistically conditioned by various interacting dimensions that allow for modelling their variation. Mode<sup>1</sup> and text production are amongst the five dimensions described by Kotze (2019: p. 346). The focus of this paper is on the variation in English-to-German translation that involves the dimension of mode, i.e. variation between spoken and written language production. We believe that such variation is manifested by the linguistic features of written and spoken translations (also referred to as ‘translations’ and ‘interpretations’, respectively), e.g. preferences for modality meanings, proportion of nominal or verbal phrases and others. These features should allow us to analyse and model the dimensions involved. Methodologically, we focus on the quantitative distributions of these linguistic features reflected in the lexico-grammar of texts.

In the following, we analyse language variation in translation products that include both translations and interpretations. Our focus is on **mode** in translation products – differences between English-to-German translations vs. interpretations. We also analyse differences between translated and non-translated texts in German. These differences correspond to the variation along the dimension of **text production**. Based on existing studies in the area of translationese, interpretese and variational linguistics (see §2) we expect that the variation along the dimension of mode should be stronger than that along the dimension of text production. We are interested in the linguistic features that contribute to the distinction between spoken and written mode in both translated and non-translated language.

The remainder of the paper is organised as follows: §2 provides an overview of the related work and theoretical background; we give details on the data and methods used in our analyses in §3; §4 is dedicated to our results and analysis; we conclude and point to some issues for discussion and future work in §5.

## 2 Related work and theoretical background

### 2.1 Translationese

We rely on studies on translationese (Baker 1993; Toury 1995; Bernardini & Ferraresi 2011; Teich 2003: among numerous others) showing that translated texts have certain linguistic characteristics in common which differentiate them from original, non-translated texts. These differences, however, do not point to the

---

<sup>1</sup>Kotze (2019: p. 346) calls this constraint ‘Modality and register’.

quality of the texts, as claimed by Gellerstam (1986) and empirically shown by Kunilovskaya & Lapshinova-Koltunski (2019). Translationese is rather a statistical phenomenon, and the differences are reflected in the distribution of lexico-grammatical, morpho-syntactic and textual language patterns that can be organised in terms of more abstract categories often called features of translations. These include *simplification* (Toury 1995), *explicitation* (Olohan & Baker 2000; Øverås 1998), *normalisation* and *shining-through* (Bernardini & Ferraresi 2011; Teich 2003; Scott 1998) and *convergence* (Laviosa 2002). Since the differences between translated and non-translated texts are of statistical character, they can be uncovered automatically. Recent studies on translationese employ automatic detection techniques using various feature constellations. One of the first works in this area is (Baroni & Bernardini 2006). They use n-grams for word forms, lemmas and parts-of-speech (POS) which represent lexical and grammatical features in a supervised scenario<sup>2</sup> to differentiate between translated and non-translated texts. Ilisei et al. (2010) use a number of simplification-related features to successfully differentiate between translated and non-translated texts with machine-learning algorithms. A number of translationese indicators have been applied in an unsupervised approach to automatic classification between translations and originals by Volansky et al. (2015). Linguistically interpretable features were used by Kunilovskaya & Lapshinova-Koltunski (2020), who automatically differentiate translated Russian and German from originals in both languages. They proceed bottom-up in their feature definition and try to identify translationese effects based on the results of corpus analysis.

## 2.2 Variational linguistics

We refer to studies in variational linguistics, such as Systemic Functional Linguistics (SFL, Halliday 2004; Halliday & Matthiessen 2014) and genre or register studies (Biber 1995; Neumann 2013). Following these studies, language varies according to the context of use. To account for a functional organisation of language, the framework offers contextual configurations, i.e. three variables characterising the level of context: *Field*, *Tenor* and *Mode* of discourse.<sup>3</sup> These variables

---

<sup>2</sup>Supervised machine learning, also referred to as text classification, is an approach for discovering groupings in multivariate data sets. In a supervised scenario, we know what groupings we can expect in the data, and the question is whether the data under analysis support these groupings (see Baayen 2008: p. 118). In an unsupervised scenario, we do not know what groupings exist in the data and an algorithm tries to identify any groupings by extracting features and patterns on its own.

<sup>3</sup>Note that Mode of discourse does not correspond to mode of production that we use to differentiate between spoken and written text. Mode of discourse is related the role of the language in the interaction.

correspond to sets of specific lexico-grammatical features. Field of discourse is realised in term patterns or functional verb classes. Tenor of discourse is realised in stance used by speakers or modality expressed by modal verbs. Mode of discourse relates to Theme-Rheme structure and cohesive relations. Linguistic features inspired by SFL and genre or register studies have been used in the analysis of contextual variation of translated texts. For instance, Evert & Neumann (2017) apply them for intralingual and cross-lingual variation in both translated and non-translated text. Lapshinova-Koltunski & Martínez Martínez (2017) use various categories of cohesive relations (related to the parameter of Mode) to automatically differentiate between spoken and written texts in English and German. This is one of the few works known to us that analyses differences between spoken and written texts with machine learning techniques. The authors succeed in automatically identifying the dimension of mode in non-translated texts.

### 2.3 Interpretese

In terms of the dimension of mode in translation, there are fewer studies on interpretese (Kajzer-Wietrzny 2012; Defrancq et al. 2015; He et al. 2016; Bernardini et al. 2016; Ferraresi & Miličević 2017; Dayter 2018; Bizzoni & Teich 2019). They show that interpreted texts possess linguistic features that differentiate them not only from translated texts but also from other language products. In our work, we aim to analyse the differences not only between interpreted and translated texts, but also between interpreted, non-interpreted and non-translated texts. With this goal in mind, we follow work by Shlesinger & Ordan (2012) who claim that modality (corresponding to our notion of mode dimension) exerts a stronger effect than ontology (corresponding to our notion of text production). This means that the dimension of mode (i.e. whether a text is spoken or written) has more influence than the dimension of text production (whether a text is a translation or an original).

## 3 Methodology

### 3.1 Hypotheses and research questions

Our research questions are based on our assumptions given in §1 above.

**Research Question 1 (RQ1)** First of all, we are interested in language variation along the dimension of text production. So, we would like to find out if we can automatically differentiate between translations and non-translations independently of the mode production (whether spoken or written).

**Research Question 2 (RQ2)** This research question is related to the analysis of mode in translation – differences between written and spoken translations. We aim to find out if we can automatically detect mode in translation.

**Research Question 3 (RQ3)** We would like to know if it is easier to automatically detect text production (differentiate between translations and non-translation) or mode (spoken and written translation) with an assumption that variation along the dimension of mode is stronger than that along the dimension of text production.

**Research Question 4 (RQ4)** We are also interested in the linguistic features that contribute to the distinction between spoken and written mode. Specifically, we want to find out if the same features are responsible for the variation between translations and interpretations and between written and spoken texts.

### 3.2 Corpus resources

For our analyses, we use written and spoken data that is derived from the European Parliament, so that all the subcorpora belong to the same register. We include transcribed interpretations and translations (spoken and written translations) from English into German (INTER and TRANS) and comparable German originals – transcriptions of European Parliament speeches by native speakers of German and published written speeches in German (GO-SP and GO-WR). The spoken part (transcribed speeches in German and interpretations from English into German) is taken from EPIC-UdS (Karakanta et al. 2019), whereas the written part (published written speeches in German and official translations from English into German) is taken from Europarl-UdS (Karakanta et al. 2018). We provide details on the size of the subcorpora in terms of number of texts (txt), sentences (sent) and tokens (token) in Table 9.1. As seen in the table, the ‘spoken’ subcorpora are much smaller.

All the texts in the subcorpora at hand were automatically annotated with information on token, lemma and part-of-speech based on the Universal Dependency framework (Nivre et al. 2019; Straka & Straková 2017). The texts are encoded in CWB and can be queried with the help of Corpus Query Processor (CQP, Evert & Team 2019), which is a part of the Corpus Workbench (CWB, Evert & Hardie 2011). They are also available in CQPWeb<sup>4</sup> supported by CLARIN-D.

---

<sup>4</sup><http://corpora.clarin-d.uni-saarland.de/cqpweb>

Table 9.1: Size of the subcorpora under analysis

subcorpus	txt	sent	token
TRANS	575	169,016	3,994,177
INTER	137	3,409	61,631
GO-SP	165	4,076	59,896
GO-WR	1072	427,775	8,954,768
<b>TOTAL</b>	<b>1,949</b>	<b>604,276</b>	<b>13,070,472</b>

These annotations facilitated the extraction of features for our analysis as described in §3.3. The accuracy of our feature extraction is thus dependent on the accuracy of the automatic annotation. The respective model performance is 99.9% for words, 80.9% for sentence borders, 91.7% for universal parts-of-speech and 95.4% for lemmas.<sup>5</sup>

### 3.3 Features

In our approach, we use a set of features derived from variational linguistics (see §2.2 above). As already mentioned above, the frameworks offer three context parameters for language variation that correspond to various lexico-grammatical patterns. Table 9.2 illustrates the features used, as well as the language patterns they represent within a text. The first column in the table contains the corresponding contextual parameter of variation, the second column includes examples of features formulated in abstract categories, and the third column shows examples of language patterns serving as operationalisations for the features.

Overall, we use 17 lexico-grammatical patterns. Four of those are patterns related to the Field of discourse (see Table 9.2). They are associated with the abstract categories of processes and participants and are linguistically realised in nouns and verbs, the distribution of content words and also *ung*-nominalisations. The next four patterns are included within the parameter of Tenor. They are related to the roles and attitudes of participants, and are realised linguistically in modality expressed by modal verbs such as *can*, *may*, *must* that we group according to their meanings (3 patterns). Tenor is also related to evaluation used by speakers to convey personal attitude to the given information, e.g. evaluative patterns like *very important*, *it is important to say*. They represent the fourth patterns in this parameter. The final nine patterns are related to Mode of discourse,

<sup>5</sup>See [http://ufal.mff.cuni.cz/udpipe/models#universal\\_dependencies\\_20\\_models](http://ufal.mff.cuni.cz/udpipe/models#universal_dependencies_20_models) for details.

Table 9.2: Features under analysis

parameter	feature	language pattern
FIELD	participants and processes	nominal and verbal parts-of-speech, content words, <i>ung</i> -nominalisations
	modality	modal meanings: obligation, permission, volition
TENOR	evaluation	evaluative patterns ( <i>more importantly/ it is important to say</i> )
MODE	textual cohesion	personal and demonstrative pronouns; general nouns ( <i>fact, plan</i> ); conjunctions; logico-semantic relations: additive, adversative, causal, temporal, modal

i.e. the role and function of language in a particular situation, the symbolic organisation of a text. They are realised as cohesive relations at the textual level, for instance coreference via pronouns (2 patterns) or general nouns (1 pattern), distribution of conjunctions (1 pattern) or discourse relations via conjunctions (5 patterns). All these features were used in previous works on translationese (see e.g. Lapshinova-Koltunski 2019; 2017).

The frequencies of these features are automatically extracted from the corpora. We use the functionality of Corpus Query Processor mentioned above. This query tool allows definition of language patterns in the form of complex regular expressions based on string and part-of-speech restrictions. The query tool delivers text instances along with their frequencies in the texts and subcorpora in which they occur. The extracted distributional information is saved in matrices for further use for statistical analysis.

### 3.4 Methods

We use Weka (Witten et al. 2011), an open source tool for statistical analysis and visualisation for our analyses. To answer the first two research questions (RQ1 and RQ2), we apply text classification using Support Vector Machines (SVM, Vapnik & Chervonenkis 1974; Joachims 1998) with a linear kernel. Classification with SVM is a supervised scenario in machine learning. We label our data with the information on classes represented in our case by text production (translations vs. non-translations) and mode (spoken vs. written), collect the information on the

language patterns outlined in §3.3 from the corpora described in §3.2, and see if our corpus data support the predefined classes.

We apply separate binary classification tasks for text production (RQ1) and mode (RQ2). The result of a linear SVM is a hyperplane (a line separating two classes) that separates the classes as best as possible, and allows a clear interpretation of the results. The classes defined in this study include translations and non-translations in the first classification task, and spoken and written modes in the second. The performance scores of classifiers are judged in terms of precision, recall and F-measure. They are class-specific and indicate the results of automatic assignment of class labels to certain texts.

To answer the third research question (RQ3), we compare the scores resulting from the two classifications in RQ1 and RQ2. If the scores are higher in the second classification task, variation along the dimension of mode is stronger than the variation along the dimension of text production (in line with our assumption).

We use methods of feature selection to answer the fourth question (RQ4). Attribute selection derived from machine learning is used to automatically select attributes (the language patterns we use) that are most relevant to the predictive modeling problem (prediction of a class membership). In the data, there is always a mixture of attributes with some being more relevant for making predictions and the others being less relevant. The process of selecting attributes in the data helps to reduce their number to those relevant for the specific prediction task. The attribute evaluator is the technique by which each attribute in the dataset is evaluated in the context of the output class (mode in our case). We use the best-first strategy (the best attribute is added at each round) which uses an iterative algorithm (starts with an arbitrary solution to a problem and attempts to find a better solution at every step). This is a correlation-based technique that evaluates the value of a subset of attributes by considering the individual predictive ability of each of these attributes along with the degree of redundancy between them. Subsets of language patterns that are highly correlated with the class (mode) and at the same time have low intercorrelation are preferred over others (see Hall 1998: for more details). We then compare the lists of resulting language patterns for the class in the two data subsets (mode in non-translation and mode in translation). Our assumption is that if there are any overlaps in the lists, this would indicate that the same/similar linguistic features are responsible for the prediction of mode in both translated and non-translated texts.



## 4 Results

### 4.1 RQ1

In the RQ1 analyses, we define three classification tasks. All classes are defined on the basis of the two text production types under analysis – translations/ interpretations and non-translations/-interpretations in German. In the first classification task, we automatically separate written translations (TRANS) from written non-translations (GO-WR). In the second classification task, we automatically separate spoken translations (INTER) from spoken originals (GO-SP). And finally, in the third classification task, we do not differentiate the mode but attempt to assign both translations and interpretations one class (TRANS+INTER) and both written and spoken originals (GO-WR+GO-SP) – the other. The performance of the classifier that automatically separates two juxtaposed classes is evaluated with a 10-fold cross-validation step. We judge the performance scores in terms of precision, recall and F-measure. These scores are specific for each class (text production type) and indicate the results of automatic assignment of production type labels to certain texts in our data. In the case of precision, we measure how many cases in the data correspond with the positive labels given by the classifier. For example, there are 137 spoken translations in our data. If the classifier assigns INTER labels to 137 texts, and all of them really belong to the subcorpus of spoken translations, then we will achieve a precision of 100%. If 37 texts turned out to be non-translations in German and were wrongly classified into the INTER class, we would have a precision of 73% only. With recall, we measure if all translated texts were actually assigned to the INTER class. So, if we have 137 translated texts, we would have the highest recall if all of them are assigned the INTER label. If only 100 out of 137 available in the data were assigned to the INTER class (and the rest to the GO-SP class), we would have a recall score of 73% of. F-measure combines both precision and recall and is given by their harmonic mean. The results of the classification performance (in terms of precision, recall and F-measure) are presented in Tables 9.3–9.5 below. Figure 9.1 provides bar plots of the weighted average of the F-Measure for the three classification tasks.

Overall, we achieve an accuracy of 86.5%, with an average F-measure of 85.7% in the first classification task for translations and non-translations. As seen from Table 9.3, non-translated texts are better identified by the classifier than the translated ones (F-measure of 90.5% vs. 76.6%). However, translations are identified with better precision (97.3% vs. 83.4%), whereas the texts originally written in German achieve higher recall (99.1% vs. 63.1%). This means that more texts in

Table 9.3: Classification results for the first text production type distinction in %

	Precision	Recall	F-Measure
TRANS	97.3	63.1	76.6
GO-WR	83.4	99.1	90.5
Weighted average	88.2	86.5	85.7

the dataset were labelled by the model as originals, with more translations being wrongly recognised as originals than originals being wrongly recognised as translations. The Weka output data show that 212 translations (out of 575) were labeled as non-translations. According to the Weka output data, around 5.4% of the translations were erroneously labelled as interpretations.

In the second text production classification task (interpretations vs. speeches originally produced in German), we achieve an accuracy of 74.5% with the average F-measure of 74.0%, pointing to the fact that text production distinction in the spoken texts is harder to make than in the written ones in the dataset at hand. The scores in Table 9.4 reveal that again, translations are recognised with better precision than non-translations (77.8% vs. 72.7%), but recall is higher for non-translations: 85.5% vs. 61.3%. This means that more texts in the dataset were recognised by the model as original speeches, and thus, more translations were wrongly recognised as originals than originals were wrongly recognised as translations, as we also observed in the first case.

If we combine the spoken and the written data to differentiate between translations and non-translations, we achieve an accuracy of 80.04% and an average F-measure of 77.7% (see Table 9.5). These overall scores are lower than in the first task (distinction between written translations and non-translations) and higher than in the second one (distinction between spoken translations and non-

Table 9.4: Classification results for the second text production type distinction in %

	Precision	Recall	F-Measure
INTER	77.8	61.3	68.6
GO-SP	72.7	85.5	78.6
Weighted average	75.0	74.5	74.0

Table 9.5: Classification results for the third text production type distinction in %

	Precision	Recall	F-Measure
TRANS+INTER	98.5	46.1	62.8
GO-WR+GO-SP	76.2	99.6	86.4
Weighted average	84.4	80.0	77.7

translations), which was foreseeable as the data in this task is a mixture of the first two. However, we observe an overall increase in the precision for translations along with an overall increase of recall for non-translations. The analysis of the confusion matrix shows that only five non-translations were erroneously classified into the class of translations, whereas more than a half of translations (53.9%) were erroneously labelled as non-translations. In other words, the translations in our data seem to be readily but inappropriately recognised as non-translations by the non-translation class, whereas non-translations are not accepted as translations by the modelled translation class. This indicates that non-translations represent a more diverse class, displaying more variation than translated texts, with the latter being a subset of non-translations in terms of the features underlying classification. In terms of translationese, this points to convergence of translations.

The results in Table 9.5 suggest that it is easier to model non-translated texts regardless of the mode they belong to (F-measure of 86.4%). Although written originals achieve the best result (F-measure of 90.5%), mixing them with spoken non-translations (whose F-measure equals 78.6%) results in a drop of 4.1% against the result for the written mode and an increase of 7.8% against the spoken mode. For translations, mixing both modes results in an F-measure of 62.8% with a drop of 13.8% against the written mode (76.6%) and a drop of 5.8% against the spoken mode (68.6%).

The results of the three classifications suggest that we can automatically tease apart translations from non-translations regardless of the mode production. At the same time, the task is easier when only written texts are involved.

## 4.2 RQ2

We perform the same analysis steps for the differentiation between spoken and written modes as we did for translation and non-translation in §4.1. We again decide for a three-fold task in the mode analysis: (1) classification of non-translated

Table 9.6: Classification results for the first mode distinction in %

	Precision	Recall	F-Measure
GO-SP	80.5	100.0	89.2
GO-WR	100.0	96.3	98.1
Weighted average	97.4	96.8	96.9

spoken and written texts (GO-SP vs. GO-WR); (2) classification of translated spoken and written texts (INTER vs. TRANS) and (3) classification of spoken and written texts with both translations and non-translations taken together (GO-SP+INTER vs. GO-WR+TRANS). In the last task, we do not sort texts according to the text production type, defining the task as one of finding the overall variation along the dimension of mode.

We achieve an overall accuracy of 96.77% with an average F-measure of 96.9% in the classification into spoken and written non-translations (see Table 9.6). Interestingly, written texts are better classified than the spoken ones (98.1% vs. 89.2% of F-measure). At the same time, we observe asymmetries in precision and recall: the classification of spoken texts delivers 80.5% for precision with 100% recall, whereas the classification of written texts works with perfect precision (100%) but with lower recall (96.3%). This means that some written originals were erroneously recognised as spoken texts, but none of the spoken texts were recognised as written texts. This indicates that some written texts in our data may contain features considered specific to spoken language.

The mode distinction in translations also achieves high accuracy (95.7%) with an average F-measure of 95.8% (See Table 9.7). Again, we observe a higher F-measure for the written translations than for the spoken ones (97.2% vs. 89.8%). Similarly to the first classification task, interpretations are recognised with better recall (100% vs. 94.6%) and translations are identified with better precision (100%

Table 9.7: Classification results for the second mode distinction in %

	Precision	Recall	F-Measure
INTER	81.5	100.0	89.8
TRANS	100.0	94.6	97.2
Weighted average	96.4	95.6	95.8

Table 9.8: Classification results for the third mode distinction in %

	Precision	Recall	F-Measure
GO-SP+INTER	81.6	99.7	89.7
GO-WR+TRANS	99.9	95.9	97.9
Weighted average	97.1	96.5	96.6

vs. 81.5%). The confusion matrix shows that around 5.4% of the translations were erroneously labelled as interpretations.

In the third classification task, we achieve 96.5% of accuracy and an F-measure of 96.6%. Similarly to the other mode distinction tasks, written texts, regardless of their text production type, achieve a better F-measure than the spoken ones (97.9% vs. 89.7%), with higher precision observed for the written texts (99.9% vs. 81.6%) and a higher recall for the spoken ones (99.7% vs. 95.9%), see Table 9.8. Mixing both text production types for the mode distinction task results in an intuitively insignificant drop in the observed scores.

The results show that we can automatically detect mode in translation, and the results of such a classification are comparable with the results on mode distinction in non-translated German. We achieve very good classification results in all tasks on mode distinction.

### 4.3 RQ3

We compare the three F-measure scores resulting from the three classification tasks within RQ1 – differentiation between translation and non-translation<sup>6</sup> with the three F-measure scores from the three classification tasks within RQ2 – differentiation between spoken and written texts.<sup>7</sup> For this, we summarise the results of all these classification tasks in Figures 9.1 and 9.2. The first figure contains average F-measure scores for the text production type distinction, whereas the second figure illustrates the F-measure scores for the mode distinction.

As seen from the graphs, it is easier to detect mode than text production type in our dataset given the same feature set. These results confirm our assumption that variation along the dimension of mode is stronger than that along the dimension of text production.

<sup>6</sup>We use the weighted average F-measure from Tables 9.3, 9.4, 9.5.

<sup>7</sup>We use the weighted average F-measure from Tables 9.6, 9.7, 9.8.

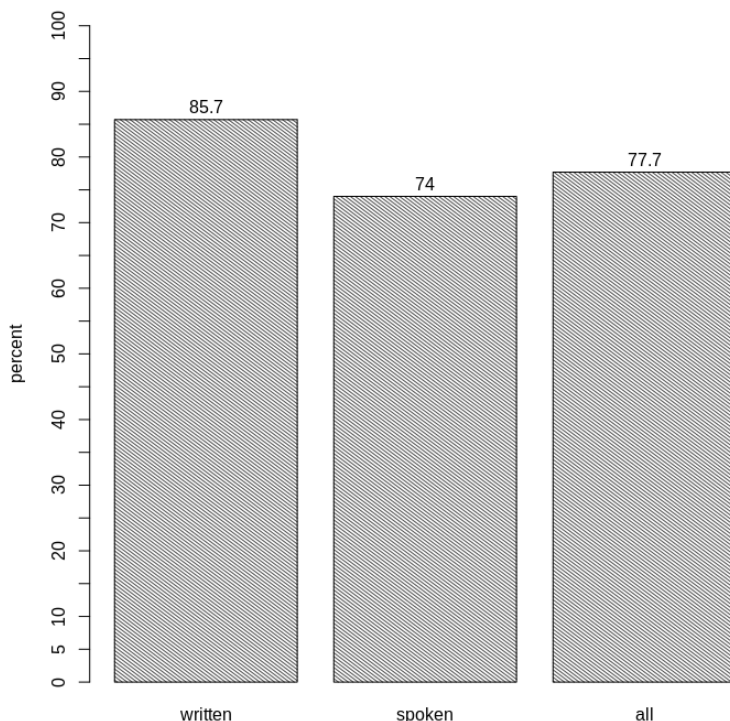


Figure 9.1: Classification results for text production distinction in % (weighted average)

#### 4.4 RQ4

As explained in §3.4, we use automatic attribute selection to test if the same language patterns are responsible for prediction of mode in both translated and non-translated texts. In both cases, the language patterns should contribute to the classification of the two classes: spoken and written.

We start with the evaluation of the language patterns relevant for the mode prediction task in non-translations. We use cross-validation with 10 folds which records in how many folds each of the attributes (our language patterns) appeared in the best subset found. We select attributes that appeared in at least one fold, which results in a list containing 10 language patterns:<sup>8</sup> content words (10),

---

<sup>8</sup>The figure in brackets indicates the number of folds the feature appears in.

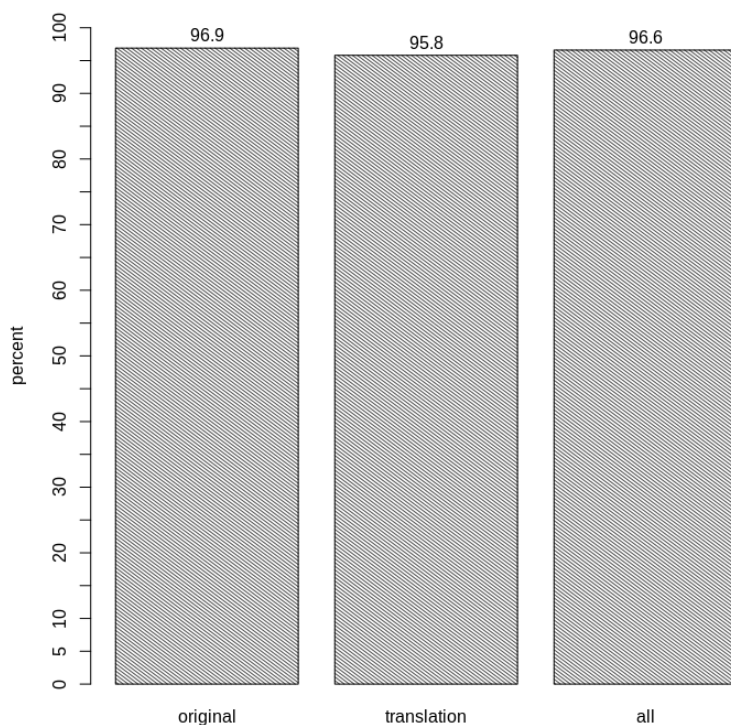


Figure 9.2: Classification results for mode distinction in % (weighted average)

nominal parts-of-speech (10), verbal parts-of-speech (9), *ung*-nominalisations (10), obligation (7), additive relations (2), adversative relations (7), modal relations (2), personal pronouns (10) and demonstratives (9).

The same procedure is applied on the dataset of translations. The list of language patterns contributing to the mode distinction here contains seven items: content words (4), nominal parts-of-speech (10), verbal parts-of-speech (3), *ung*-nominalisations (10), temporal relations (10), personal pronouns (4) and demonstratives (2).

For a better visualisation, we outline the language patterns selected for the mode distinction in both non-translated and translated texts in Table 9.9. We also relate them to the more abstract feature categories as well as contextual parameters introduced earlier (see Table 9.2 in §3.3 for an overview).

Table 9.9: Features contributing to the mode distinction

non-translated			translated		
par.	feat.	lang.pattern	lang.pattern	feat.	par.
Field	participants, processes	content words	content words	participants, processes	Field
Field	participants	nominal pos	nominal pos	participants	Field
Field	processes	verbal pos	verbal pos	processes	Field
Field	processes	<i>ung-nom.</i>	<i>ung-nom.</i>	processes	Field
Teno	modality	obligation			
Tenor	cohesion	pers. pron.	pers. pron.	cohesion	Mode
Mode	cohesion	dem. pron.	dem. pron.	cohesion	Mode
Mod <sub>c</sub>	cohesion	additive			
Mod <sub>c</sub>	cohesion	adversative			
			temporal	cohesion	Mode

As seen in Table 9.9, the two lists have an overlap of six language patterns, while the first list contains three language patterns not included in the second list (modal verbs of obligation, additive and adversative relations). However, the second list is not entirely a subset of the first one, as it contains one language pattern which is not included in the first list (temporal relations). We mark the non-overlaps in grey in the table. In terms of abstract linguistic features, participants, processes and cohesion contribute to the mode distinction in both translated and non-translated texts, which corresponds to the contextual parameters of Field and Mode. However, in the texts originally produced in German, there is also modality corresponding to Tenor, which is not distinctive for mode in the translations. It is also interesting to see that although discourse relations contribute to the mode distinction in both translations and non-translations, they differ in the logico-semantic types in each list.

Since the majority of the features overlap (6/10 and 6/7), we suggest that the same features (especially if interpreted in terms of abstract categories) are responsible for the variation between translations and interpretations and between written and spoken texts. The overlap in the features common for the distinction of mode may be traced back to the register the texts in the dataset belong to – they are all speeches from the parliamentary debates.



## 5 Conclusion and discussion

The present study focuses on the variation in English-to-German translation along the mode dimension. Translation variation is reflected in the linguistic features that we were able to analyse with language patterns derived from variational linguistics. We extracted the distribution of these patterns in spoken and written texts that included both texts originally spoken or written in German, and translations and interpretations. This distributional information was used to statistically model variation along the text production dimension (translations vs. non-translation) and the mode dimension (spoken vs. written). Our results show that we are able to automatically tease apart translation from non-translations, as well as spoken texts from the written texts using the same feature set. However, it turned out to be easier to automatically differentiate between spoken and written texts regardless of their production type, which confirms our assumption that language variation along the dimension of mode is stronger than that along the dimension of text production. We are also able to find out which linguistic features contribute to the distinction between spoken and written mode in both translated and non-translated language.

This brings our findings in accordance with Shlesinger & Ordan (2012)'s claim that mode exerts a stronger effect than text production. This means that the difference between spoken and written texts is stronger than that between translations and non-translations. In this way, the interpretations in our dataset show more similarities to the speeches originally spoken in German than to the written translation, making interpretations more 'spoken' than 'translated'.

At the same time, we realise that our study also has a number of limitations. First of all, we use a feature set inspired by variational linguistics. Although it has been applied in the analysis of translationese in a number of previous studies, it was originally developed for the analysis of register variation that also includes variation along the dimension of mode. However, many of the language patterns in our set are extensively applied in the analysis of translationese (e.g. cohesive markers) as well.

Another drawback of the present study is the limitation of the corpus data – it includes political speeches only. Yet, whereas there are many translation corpora which could be used for such an analysis, it is hard to find comparable interpreted data.

In the future, we should extend the features and the data to further investigate the specifics of translated and interpreted texts. It will also be interesting to have a closer look at the features contributing to the mode distinction and perform a qualitative analysis of these features.

## Abbreviations

RQ	research question	ung-nom.	ung-nominalisation
pos	part-of-speech	par.	parameter
pers. pron.	personal pronoun	feat.	feature
dem pron.	demonstrative pronoun	lang.pattern	language pattern

## References

- Baayen, R. Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 233–250. Amsterdam: John Benjamins.
- Baroni, Marco & 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. DOI: 10.1093/lc/fqi039.
- Bernardini, Silvia & Adriano Ferraresi. 2011. Practice, description and theory come together: Normalization or interference in Italian technical translation? *Meta* 56. 226–246.
- Bernardini, Silvia, Adriano Ferraresi & Maja Miličević. 2016. From EPIC to EPTIC—exploring simplification in interpreting and translation from an intermodal perspective. *Target* 28. 61–86.
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Bizzoni, Yuri & Elke Teich. 2019. Analyzing variation in translation through neural semantic spaces. In *Proceedings of the 12th Workshop on Building and Using Comparable Corpora (BUCC) at RANLP-2019*. Varna, Bulgaria: ACL.
- Dayter, Daria. 2018. Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). *FORUM* 16(2). 241–264.
- Defrancq, Bart, Koen Plevoets & Cédric Magnifico. 2015. Connective items in interpreting and translation: Where do they come from? In Jesús Romero-Trillo (ed.), *Yearbook of corpus linguistics and pragmatics*, 195–222. New York: Springer International Publishing.
- Evert, Stefan & Andrew Hardie. 2011. Twenty-first century corpus workbench: Updating a query architecture for the new millennium. In *Proceedings of the corpus linguistics 2011 conference*. Birmingham, UK: University of Birmingham.

- Evert, Stefan & Stella Neumann. 2017. The impact of translation direction on characteristics of translated texts : A multivariate analysis for English and German. *Empirical Translation Studies: New Methodological and Theoretical Traditions* 300. 47.
- Evert, Stefan & The CWB Development Team. 2019. *CQP query language tutorial*. Tech. rep. Version Version 3.4.15.
- Ferraresi, Adriano & Maja Miličević. 2017. Phraseological patterns in interpreting and translation. Similar or different? In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies. New methodological and theoretical traditions*, vol. 300 (Trends in Linguistics. Studies and Monographs [TiLSM]), 157–182. Berlin: Mouton de Gruyter.
- Gellerstam, Martin. 1986. Translationese in Swedish novels translated from English. In L. Wollin & H. Lindquist (eds.), *Translation studies in Scandinavia*, 88–95. Lund: CWK Gleerup.
- Hall, Mark A. 1998. *Correlation-based feature subset selection for Machine learning*. University of Waikato. (Doctoral dissertation).
- Halliday, Michael A.K. 2004. *An introduction to functional grammar*. London: Arnold.
- Halliday, Michael A.K. & Christian M.I.M. Matthiessen. 2014. *Halliday's introduction to functional grammar*. 4th edn. London: Routledge.
- He, He, Jordan Boyd-Graber & Hal Daumé III. 2016. Interprete vs. translationese: The uniqueness of human strategies in simultaneous interpretation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 971–976. San Diego, California: Association for Computational Linguistics. DOI: 10.18653/v1/N16-1111. <http://www.aclweb.org/anthology/N16-1111>.
- Ilisei, Iustina, Diana Inkpen, Gloria Corpas Pastor & Ruslan Mitkov. 2010. Identification of translationese: A supervised learning approach. In Alexander Gelbukh (ed.), *Proceedings of CICLing-2010*, vol. 6008 (LNCS), 503–511. Heidelberg: Springer.
- Joachims, Thorsten. 1998. Text categorization with support vector machines: Learning with many relevant features. In Claire Nédellec & Céline Rouveirol (eds.), *Machine learning: ECML-98*, 137–142. Berlin: Springer.
- Kajzer-Wietrzny, Marta. 2012. *Interpreting universals and interpreting style*. Unpublished PhD thesis. Poznan: Uniwersytet im. Adama Mickiewicza. (Doctoral dissertation).
- Karakanta, Alina, Katrin Menzel, Heike Przybyl & Elke Teich. 2019. Detecting linguistic variation in translated vs. interpreted texts using relative entropy. In *Empirical investigations in the forms of mediated discourse at the European*

- parliament, thematic session at the 49th Poznan linguistic meeting (PLM2019), Poznan.*
- Karakanta, Alina, Mihaela Vela & Elke Teich. 2018. EuroParl-UdS: Preserving and extending metadata in parliamentary debates. In *ParlaCLARIN workshop, 11th Language Resources and Evaluation Conference (LREC2018)*. Miyazaki, Japan. [http://lrec-conf.org/workshops/lrec2018/W2/pdf/10\\_W2.pdf](http://lrec-conf.org/workshops/lrec2018/W2/pdf/10_W2.pdf).
- Kotze, Haidee. 2019. Converging what and how to find out why: An outlook on empirical translation studies. In Lore Vandevoorde, Joke Daems & Bart Defrancq (eds.), *New empirical perspectives on translation and interpreting* (Routledge Advances in Translation and Interpreting Studies), 333–371. London: Routledge.
- Kunilovskaya, Maria & Ekaterina Lapshinova-Koltunski. 2019. Translationese features as indicators of quality in English-Russian human translation. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, 47–56. Varna: Incoma Ltd. <https://www.aclweb.org/anthology/W19-8706>.
- Kunilovskaya, Maria & Ekaterina Lapshinova-Koltunski. 2020. Lexicogrammatic translationese across two targets and competence levels. In *Proceedings of LREC-2020*. Marseille, France.
- Lapshinova-Koltunski, Ekaterina. 2017. Exploratory analysis of dimensions influencing variation in translation: The case of text register and translation method. In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies: New methodological and theoretical traditions*, vol. 300 (TILSM series), 207–234. TILSM series. Berlin: Mouton de Gruyter.
- Lapshinova-Koltunski, Ekaterina. 2019. Exploring linguistic differences between novice and professional translators with text classification methods. In Lore Vandevoorde, Joke Daems & Bart Defrancq (eds.), *New empirical perspectives on translation and interpreting* (Routledge Advances in Translation and Interpreting Studies), 215–239. London: Routledge.
- Lapshinova-Koltunski, Ekaterina & José Manuel Martínez Martínez. 2017. Statistical insights into cohesion: Contrasting English and German across modes. In Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds.), *Contrasting English and other languages*, 130–163. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Laviosa, Sara. 2002. *Corpus-based translation studies, theory, findings, application*. Amsterdam: Rodopi.
- Neumann, Stella. 2013. *Contrastive register variation. A quantitative approach to the comparison of English and German*. Berlin, Boston: Mouton de Gruyter.

- Nivre, Joakim et al. 2019. *Converging what and how to find out why an outlook on empirical translation studies*. Tech. rep. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. 333–371. <http://hdl.handle.net/11234/1-1983>.
- Olohan, Maeve & Mona Baker. 2000. Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1. 141–158.
- Øverås, Linn. 1998. In search of the third code. An investigation of norms in literary translation. *Meta* 43. 557–570.
- Scott, Maria Nelia. 1998. *Normalisation and readers' expectations: A study of literary translation with reference to Lispector's A Hora Da Estrela*. doctoral dissertation. Liverpool: University of Liverpool. (Doctoral dissertation).
- Shlesinger, Miriam & Noam Ordan. 2012. More spoken or more translated?: Exploring a known unknown of simultaneous interpreting. *Target* 24. 43–60.
- Straka, Milan & Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 shared task: Multilingual parsing from raw text to universal dependencies*, 88–99. <http://www.aclweb.org/anthology/K17-3009>.
- Teich, Elke. 2003. *Cross-linguistic variation in system and text. A methodology for the investigation of translations and comparable texts*. Berlin: Mouton de Gruyter.
- Toury, Gideon. 1995. *Descriptive translation studies – and beyond*. Amsterdam: John Benjamins.
- Vapnik, Vladimir N. & Alexey J. Chervonenkis. 1974. *Theory of pattern recognition*. Moscow: Nauka.
- Volansky, Vered, Noam Ordan & Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities* 30(1). 98–118. DOI: 10.1093/llc/fqt031.
- Witten, Ian H., Eibe Frank & Mark A. Hall. 2011. *Data mining: Practical Machine learning tools and techniques: Practical Machine learning tools and techniques* (The Morgan Kaufmann Series in Data Management Systems). Amsterdam: Elsevier Science.