Pragmatic information in translation: a corpus-based study of tense and mood in English and German

Anita Ramm[†], **Ekaterina Lapshinova-Koltunski**[‡], **Alexander Fraser**[†]
[†]Center for Information and Language Processing, LMU Munich
[‡]Saarland University

Abstract

Grammatical tense and mood are important linguistic phenomena to consider in natural language processing (NLP) research. We consider the correspondence between English and German tense and mood in translation. Human translators do not find this correspondence easy, and as we will show through careful analysis, there are no simplistic ways to map tense and mood from one language to another. Our observations about the challenges of human translation of tense and mood have important implications for multilingual NLP. Of particular importance is the challenge of modeling tense and mood in rule-based, phrase-based statistical and neural machine translation.

1 Introduction

This paper analyzes tense and mood in English and German from the perspective of the data commonly used to train MT systems or to model tense/mood, namely freely available bilingual texts. The need for a thorough analysis of tense/mood in parallel texts arises from the fact that there is a high degree of variation between the two languages resulting in a many-to-many relation in the tense/mood translation between English and German. Particularly, frequently occurring unintuitive tense correspondences and the low frequency of the many tense/mood combinations is problematic for different NLP tasks using parallel corpora. We study the correspondences in a large English-German parallel corpus and explain them from the point of view of different pragmatic factors – contextual constraints in terms of genre/user preferences or textual properties, and tense interchangeability. We compare English and German morpho-syntactic tense sets suffering from tense correspondence gaps in both directions and discuss the impact of translation process on the tense/mood variability in our data. Finally, we take a look to the modeling of tense and mood for machine translation pointing to important features needed to transfer tenses between languages. Our analysis indicates that bilingual modeling of tense and mood cannot be properly done by considering solely lexical/syntactic features, e.g. words, POS tags, etc., also supported by the previous work (Ye et al., 2006). Instead, incorporation of pragmatic information is required, which is currently not directly accessible to most NLP systems. We summarize the pragmatic information required and provide a list of available tools for automatic annotation with the respective information, which will be of direct use in future efforts to solve this difficult modeling task. In the following, we present theoretical issues and related work (Section 2), quantitative analysis on the usage of the tense/mood correspondences in English-German parallel data and their modeling in the context of MT (Section 3), summarizing the findings in Section 4.

2 Theoretical issues and related work

2.1 Contrasts in English and German tense and mood systems

As known from contrastive grammars (König and Gast, 2012; Hawkins, 2015), English and German share a common ground of six morpho-syntactic tenses: present/*Präsens*, simple past/*Präteritum*, present perfect/*Perfekt*, past perfect/*Plusquamperfekt*, future I/*Futur I* and future II/*Futur II*. We summarize those with examples in both languages in Table 1, which we created to show the correspondence between these languages. In English, each of the tenses has a progressive variant. The German tense system does not have an explicit marking of the progressive aspect. But German has a larger set of subjunctive tense forms. While a few of

Morph. tense	English		German	
	Synt. tense	Example	Synt. tense	Example
present	present simple	(I) read	Präsens	(Ich) lese
	present progressive	(I) am reading		
	present perfect	(I) have read	Perfekt	(Ich) habe gelesen
	present perfect progressive	(I) have been reading		
	future I	(I) will read (I) am going to read	Futur I	(Ich) werde lesen
	future I progressive	(I) will be reading (I) am going to be reading		
	future II	(I) will have read	Futur II	(Ich) werde gelesen haben
	future II progressive	(I) will have been reading		
past	past simple	(I) read	Präteritum	(Ich) las
	past progressive	(I) was reading		
	past perfect	(I) had read	Plusquam-perfekt	(Ich) hatte gelesen
	past perfect progressive	(I) had been reading		
present*	conditional I	(I) would read	Konjunktiv II	(Ich) würde lesen
past*	conditional I progressive	(I) would be reading		
	conditional II	(I) would have read	Konjunktiv II	(Ich) hätte gelesen
	conditional II progressive	(I) would have been reading		
present*			Konjunktiv I	(Er) lese (Er) werde lesen

Table 1: List of the tenses in English and German in active voice. The table indicates the tense correspondences in terms of their morpho-syntactic structure.

them have direct morpho-syntactic counterparts in English, most of them correspond to indicative tenses in English. The meaning of a specific tense form may considerably vary too. We summarize the contrasts related to the meaning of the English and German tenses described by König and Gast (2012) in Table 2. This description refers to different aspects such as the time reference (*past, futurate, future*, etc.) and relation to the moment of utterance (*resultative, universal, narrative*). In other words, the (non-)parallelism of the respective tenses can be established by considering specific semantic properties of a given verb and the utterance that the respective verb occurs in. Different aspects in the English tense system have different impacts on the use of tenses. For instance, in contrast to the simple present tense, the present progressive can be used in the futurate context. In German, *Präsens* can almost always be used to refer to the future. The English progressive tense lacks direct counterparts in German and is therefore translated into a number of different German tenses.

English and German also differ greatly with respect to the grammatical mood. In German, the subjunctive is expressed in the verbal morphology and interacts with the German tense system changing the time of an utterance. German distinguishes between two subjunctive morpho-syntactic forms: *Konjunktiv I* and *Konjunktiv II*. The latter is used in the context of conditional and contrafactual utterances. Usually, sentences with *Konjunktiv II* are composed of at least two clauses. There are, however, also *free factive* occurrences of *Konjunktiv II*, where it occurs in a simple sentence, see Example (1). Such sentences may, for instance, indicate politeness.

(1) Ich **hätte** gern ein Glas Wasser.
I have gladly a glass water.
'I'd like to have a glass of water.'

Both *Konjunktiv I* and *Konjunktiv II* can be used in the context of the reported speech. Note, however, that the use of the subjunctive mood is not grammatically required to signal reported speech. In fact, the two *Konjunktiv* forms and the indicative mood are often used interchangeably in reported speech (Csipak, 2015). For the English subjunctive mood, König and Gast (2012) rather use the term *quasi-subjunctive*, since subjunctive mood in English exists only for the verb *be*. Other forms used in the subjunctive contexts correspond to the infinitives. The German *Präsens* and *Futur I* are interchangeable in many contexts. In the futurate use, *Präsens* is usually combined with a temporal phrase which points to the future; in (2), the adverbial *morgen* provides the respective temporal information. However, the temporal phrase is not always overtly given in a considered sentence: in (3), the verb *kommen* in the present tense refers to the future which is obvious solely by considering the preceding sentence.

(2) Ich **komme** morgen. I come tomorrow.
'I'll come tomorrow.'

(3) **Kommst** du morgen? Ja, ich **komme.**
Come you tomorrow? Yes, I come.
'I'll come tomorrow.'

Another prominent example of tense interchangeability in German is related to the past tenses.

Use	German	English
Präsens/present tense		
non-past	Ich schlafe von 12 bis 7.	I sleep from midnight to seven.
futurate	Morgen weiß ich das.	→ future tense (*I will know that tomorrow.*)
Präteritum/simple past		
past time	Ich schlief den ganzen Tag.	I slept the whole day.
Futur I/future tense		
future time	Ich werde schlafen.	I will sleep. I am going to sleep.
Perfekt/present perfect		
resultative	Jemand hat mein Auto gestohlen.	Someone has stolen my car.
existential	Ich habe (schon mal) Tennis gespielt.	I have played tennis.
hot news	Kanzler Schröder ist zurückgetreten.	Chancellor Schröder has resigned.
universal	→ Präsens (*Ich lebe hier seit 2 jahren.*)	I have lived here for two years.
narrative	Ich bin gestern im Theater gewesen.	→ past tense (*I was in theater yesterday.*)
Futur II/future perfect		
future results	Ich werde das bis morgen erledigt haben.	I will have done this by tomorrow.
Plusquamperfekt/past perfect		
pre-past	Ich hatte geschlafen.	I had slept.

Table 2: Meaning of tenses in English and German (König and Gast, 2012, p. 92)

There are some fine-grained differences between the respective tenses, but at least *Präteritum* and *Perfekt* are interchangeable in many contexts, see Sammon (2002). In fact, the dominance of either of the forms is a matter of author's preference or contextual constraints, see 2.2 below. For instance, *Perfekt* is often used in spoken language, while *Präteritum* is more frequently used in writing. Furthermore, there is a certain lexical preference: auxiliaries and modals are more frequently used in *Präteritum* than in *Perfekt*.

2.2 Contextual constraints

Contextual constraints on the tense/mood usage have been analyzed mostly in a monolingual context. For example, Weinrich (2001) differentiates between two groups of the German tenses: (i) *discussing* (*Präsens*, *Perfekt*, *Futur I*, *Futur II*) and (ii) *narrative* (*Präteritum*, *Plusquamperfekt*, and subjunctives *Konjunktiv I* and *Konjunktiv II*). His classification is relevant for genre differentiation. For instance, the narrative tenses are mostly found in written German (e.g., literary works), while the discussing tenses are more often used in the spoken language. In a multilingual context, there exist a few studies that analyze the role of tense/mood in functional variation of language called register variation. Biber (1995) uses preferences for specific tense and mood as linguistic indicators for specific registers in a number of languages. Neumann (2013) presents a contrastive corpus-based study of English and German (including translations), in which the tense frequency is used among other textual properties to induce the *goal type* of the text (one of the parameters of register variation): argumentation, narration, instruction, etc. She observed that the frequency of present vs. past across texts from different registers expose different (i.e., domain-specific) distributional specifics: past tenses are rather typical for narrative texts, while present tense verbs are more typical for argumentative texts such as political essays, popular science articles, etc. These findings are in line with the classification of tenses proposed by Weinrich (2001). In addition to contextual constraints expressed in genre or register, translation of tenses may also follow a *set of rules* defined for a specific translation project. For instance, the translation guidelines of the European Commission for German require the session minutes or reports be written in the present tense.[1]

2.3 Tense and mood in human translation

Tense and mood were analyzed in previous studies on English-German translation (Teich, 2003; Neumann, 2013). However, a systematic description of tense/mood transformation patterns for this language pair has been missing until our work. At the same time, translation studies provide us with valuable information on how translation process has an impact onto translated texts which, as a result, differ from non-translated texts both in the source (SL) and the target language (TL). These differences are reflected in the features of translated language (Gellerstam, 1986; Baker, 1993). Two of these translation features are important for bilingual modeling of tense and mood: (i) *shining through* and (ii) *normalization*. The former one indicates the closeness of the translation to the source (Teich, 2003), whereas the latter one is related to the tendency to conform (and exaggerate) the patterns typical for the TL (Baker, 1993). We would observe *shining through* in our data if tenses used in the sources are preserved in the transla-

[1] https://ec.europa.eu/info/sites/info/files/german_style_guide_de_0.pdf

tions. While there is much parallelism with respect to tense in the two languages under analysis, many cases may expose a TL-specific usage of tense, which may considerably differ from a form given in the source due to a smaller set of tenses available in the German language system. Following Teich (2003), *shining through* is less prominent and *normalization* is more prominent when translating into a language which has fewer options with respect to a specific grammatical system. This means that our parallel texts may expose a great variation in the tense translation. Finally, parallel corpora represent a concatenation of the translations produced by many different translators. Therefore, we expect that the observed variation in tense translations can be impacted by the preferences of a specific translator.

2.4 Tense and mood in machine translation and NLP

In the context of the rule-base MT, i.e., in EUROTRA (Copeland et al., 1991), translation of tense and mood relies on an interlingua representation to which the SL sentence is mapped, and which is then mapped to the syntax of the TL respectively. This mapping is rule-based and follows a set of manually defined rules which make use of different kinds of information. The rules for English-German formulated within EUROTRA indicate that tense cannot be considered in isolation, but rather in a combination with other related linguistic features such as aspect and *Aktionsart*. Thus, specific modality, as well as voice properties, need to be considered in the bilingual modeling of tense and mood.

Recently, there have been attempts to automatically model tense and mood for different NLP tasks. In the monolingual context, for instance, Tajiri et al. (2012) used a tense classification model for detecting and correcting tense in the texts produced by English learners. In the bilingual context, Ye et al. (2006) presented an empirical study of the features needed to train a classification model for predicting English tenses given the source sentences in Chinese. Gispert and Mariño (2008), Loáiciga et al. (2014) and Ramm and Fraser (2016) presented work on building tense classification models which are used to improve tense choice in statistical MT systems for English-Spanish, English-French and English-German, respectively. While Loáiciga et al. (2014) reported encouraging results, Gispert and Mariño (2008) and Ramm and Fraser (2016) left unanswered questions about the appropriate method and the necessary contextual information for modeling tense and mood in a bilingual context.

3 Analyses

3.1 Tense and mood in human translation

Data and tools Since one of our aims is to serve the task of machine translation, our contrastive analysis of tense and mood in English and German relies on the parallel corpora provided for WMT15 shared tasks on machine translation (Bojar et al., 2015). We make use of the News corpus (news articles, 272k sentences), the Europarl corpus (1,9 mio. sentences) and the Crawl corpus, a large collection of mix-domain bilingual documents retrieved from the Internet (2,4 mio. sentences). In addition, we also consider Pattr[2], a medical corpus (1,8 mio. sentences). In this way, we have a constellation of various domains (as they are called in NLP) or registers/genres (as they are called in the studies described in 2.2 above). The corpora are tokenized with a standard tokenizer provided with the SMT toolkit *Moses* (Koehn et al., 2007) and parsed with the Mate parser (Bohnet and Nivre, 2012) which provides dependency parse trees for both languages, and, for German, morphological analysis of words. Both sides of the parallel corpora are annotated with tense, mood and voice information using the TMV annotator (Ramm et al., 2017). The English-German verb pairs annotated with the respective information are then extracted by (i) automatically computing word alignment of the parallel texts with Giza++ (Och and Ney, 2003) and (ii) identifying pairs of VCs from the aligned, annotated parallel data (see example in Figure 1). In our analyses, we do not differentiate between translation directions, because we are interested in all transformations possible for the analyzed language pair.

Indicative tense As already mentioned in Section 2.1, the English progressive tenses are translated into a number of different German tenses. Figure 2 illustrates the frequency distribution of the English present perfect (progressive) in our data. It is striking that both English tense forms correspond to three different German tenses in

[2] http://www.cl.uni-heidelberg.de/statnlpgroup/pattr/de-en.tar.gz

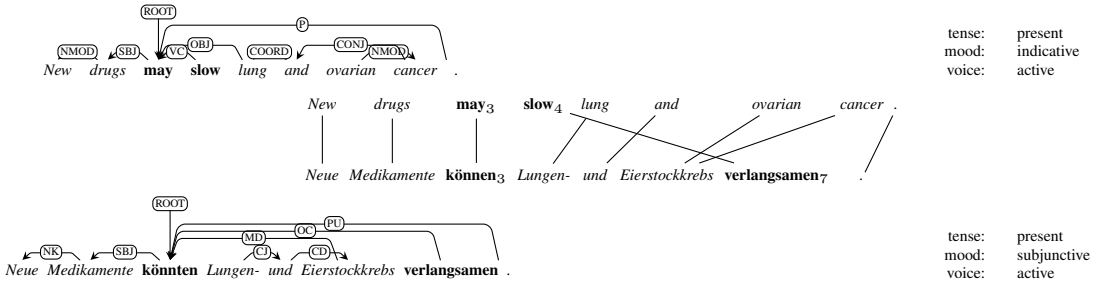

Figure 1: Word-aligned, parsed English-German parallel sentence pair with TMV annotations. Parallel VC: *may slow* ↔ *können verlangsamen*, tense/mood pair *present/indicative* ↔ *present/subjunctive*.

most cases: *Präsens*, *Perfekt* and *Präteritum* whereby *Perfekt* is the most prominent equivalent. Considering the two German past tenses together, it becomes clear that both present perfect tense forms correspond to one of the German past tenses more often than the present tense does. Progressiveness also seems to have a large impact on the translation of present perfect into German: ca. 77% of the non-progressive forms corresponds to one of the German past tenses, whereas 56% of the progressive cases do so. In other words, the progressive present tense still prefers to be transferred into one of the German past tenses. However, the German *Präsens* more often corresponds to this English tense than to the non-progressive variant.

Subjunctive mood The frequency distribution of the tense correspondences between subjunctive forms based on news texts is shown in Figure 3. As expected, the German *Konjunktiv* tense forms are equivalents of all English indicative tense forms in the dataset at hand. Thereby, the *Konjunktiv II* is a more frequent equivalent than the *Konjunktiv I*. Assumed that the conditional and counter-factual situations in English are described with conditional forms, it is quite unexpected that the other English tense forms more often correspond to the German *Konjunktiv II* (used to indicate conditional contexts) than to *Konjunktiv I* (used to indicate reported speech) in our translation data. A possible explanation for this is that in the news data, *Konjunktiv II* is more often used to express reported speech than the *Konjunktiv I* form. When expressing non-factual events, English conditionals can be seen as direct counterparts of the German *Konjunktiv II*: Figure 4 shows that *Konjunktiv II* is the most frequent equivalent for all four English conditional tense forms in our data. Further frequent counterparts are *Präteritum* for the conditional I and *Perfekt* for the conditional II.

Finite vs. non-finite verbal complexes Our data shows that the usage of non-finite VCs in the two languages varies considerably. For instance, in the News corpus, 16.7% of all VCs in English are non-finite, while this is the case for only 7.9% of the German VCs. Similar ratio is also given in the Europarl corpus in which 18.2% of the English VCs and 6.2% of the German VCs, respectively, are non-finite. Figure 5 indicates that the major part of the English non-finite VCs have German finite VCs as equivalents. These translation equivalents pose an interesting problem in the context of MT. When translating from English to German, MT needs to generate a finite clause for the given non-finite source clause. Particularly, it needs to generate a finite German VC in a tense form for which there is no obvious evidence in the source.

Tense interchangeability In the data from the News corpus, we identified 190 occurrences of the auxiliary *sein (to be)* in one of the composed past tenses (i.e., *Perfekt* and *Plusquamperfekt*) in active voice in contrast to 10,247 occurrences in the simple past tense *Präteritum*. This lexical preference is also given for a few additional full verbs in German as shown by counts derived from the Crawl corpus: *denken* (819 vs. 354), *stehen* (3083 vs. 98), *geben* (7220 vs. 1523), *ziehen* (1565 vs. 145). The data suggests the same preference also for the passive voice: *denken* (184 vs. 38), *geben* (1517 vs. 395), *ziehen* (380 vs. 78).

Contextual specifics Direct comparison of the tense frequencies extracted from the German data (Figure 6) shows variation in the usage of tenses in different domains (or registers). Being the most frequent tense form in all corpora, *Präsens*, however, differs in its relative frequency. For instance,

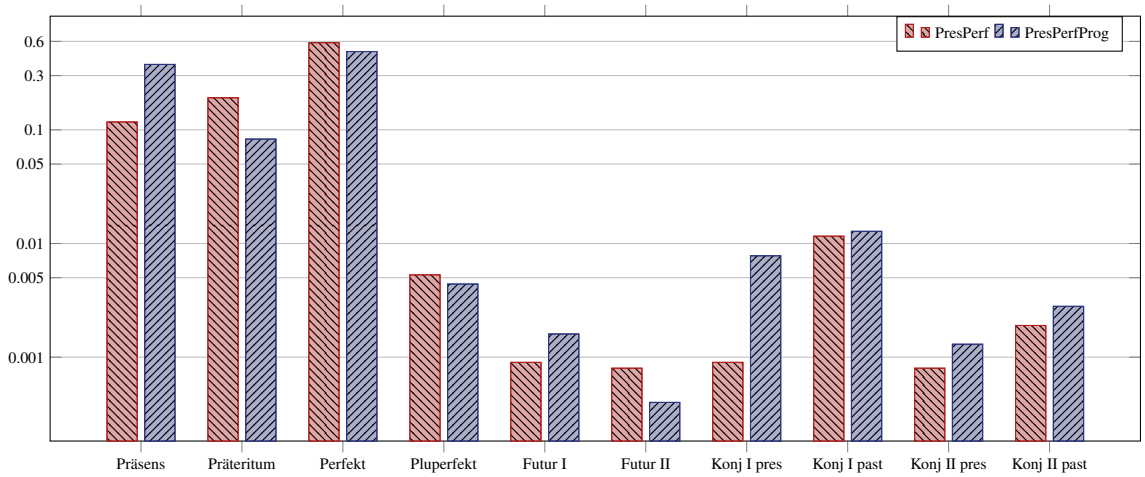

Figure 2: German correspondences of the English present prefect (progressive) in the Europarl corpus.

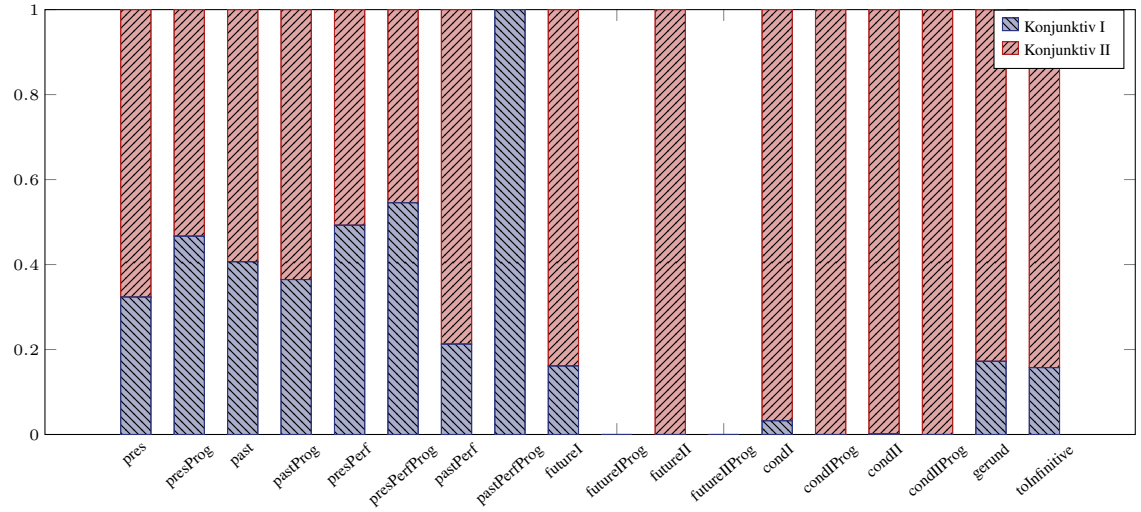

Figure 3: English correspondences of the German *Konjunktiv* tenses in the Europarl corpus.

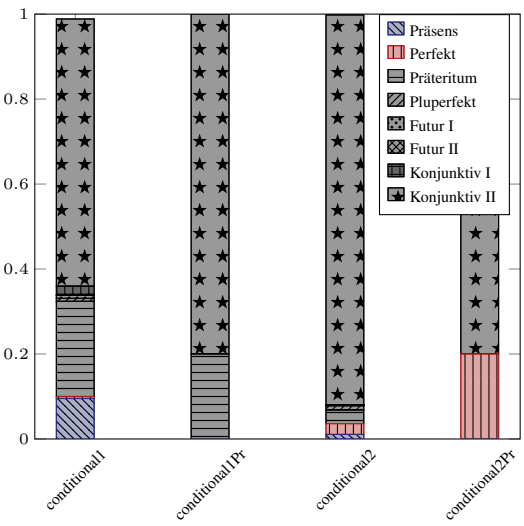

Figure 4: German correspondences of the English conditionals in the News corpus.

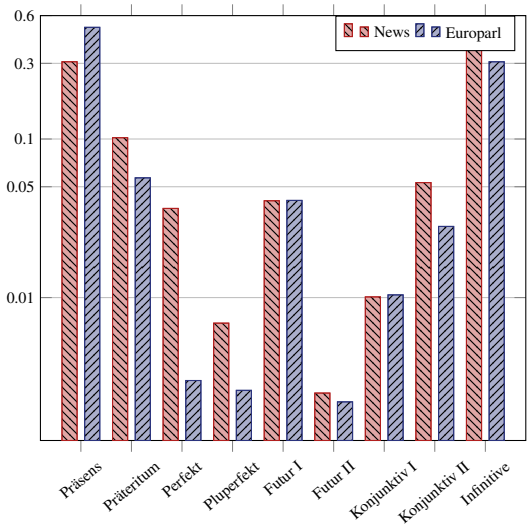

Figure 5: German correspondences of the English non-finite VCs – gerunds and to-infinitives.

in News, the relative frequency of *Präsens* is 9% lower than in Europarl (0.66 vs. 0.75), while *Präsens* represents 97% of the tense forms in Pattr (medical texts). We also observe variation in the past tense use within the respective corpora. For instance, in Europarl, *Präteritum* and *Perfekt* have almost equal relative frequency (0.08 and 0.10, respectively), whereas News clearly prefers the narrative tense *Präteritum* over the discussing tense *Perfekt* (0.19 vs. 0.08, respectively).

3.2 Modeling tense and mood

Many-to-many relation Corpus analyses of human translations presented in Section 3.1 show that the respective monolingual, as well as bilingual linguistic tense-related specifics in English and German result in a many-to-many relation. Figure 7 illustrates this relation on the basis of distributions of tense transformation patterns derived from our data. A formal description of the respective many-to-many relation requires knowledge on different linguistic levels: lexical, syntactic and semantic/pragmatic.

One of the reasons for the many-to-many relation is the different granularity of the tense systems in the two languages. While there are tenses in English which do not have a direct counterpart in German, some tense forms in German do not have a direct counterpart in English (*Konjunktiv I*) either.

Tense/mood-related contextual features For automatic modeling of tense and mood, textual characteristics discussed in the preceding sections need to be mapped to the specific contextual information overtly given in a sentence. The respective contextual features are summarized in Table 3. Many of these features can be derived from parsed and POS-tagged data. However, some of them require access to other annotation tools, as well as lexical databases which include information about semantic properties of the words/ Automatic annotation of the temporal ordering can be done with the tool TARSQI (Verhagen et al., 2005). Information about tense, mood and voice of the VCs in the English texts can be obtained with the TMV annotator (Ramm et al., 2017). Information about Aktionsart in terms of state, event and progress can be gained from the output of the tool Sitent (Friedrich and Palmer, 2016). Currently, no tools are publicly available for automatic identification of conditionals in English, which is important for translation of the German subjunctive mood. However, the set of syntactic rules described by Olivas et al. (2005) can be reused to identify the respective contexts in English. To our knowledge, there are no publicly available tools for automatic annotation of texts with genre and/or domain information, although there has been ongoing research in this area (Santini, 2007; Sharoff et al., 2010; Petrenz, 2014; Biber and Egbert, 2016). As seen from Table 3, textual properties carrying pragmatic information represent their own subtasks in NLP. Tools for annotation of the respective information are mostly based on classification models that use many different subtask-related information. While predicted annotations are correct in many cases, they may also be erroneous, consequently having negative impact on training a tense/mood translation model. Instead of using outputs of many different tools requiring a complex processing pipeline, one might train a model directly with the features used to train models for predicting each of the relevant textual properties.

The information on the distribution of various tenses and mood in the bilingual data is also important. Therefore, training data should be carefully preselected to account for this specific distribution.

4 Discussion and Conclusion

The paper describes a contrastive analysis of English and German tense and mood by means of parallel data. We provide an overview of the categories available in both language systems, point to the existing asymmetries providing corpus-based evidence from human translations and formulate assumptions on their impact on MT. Our translation data shows considerable amount of variation of tense/mood translation between the two languages. Translations also vary a lot leading to many unexpected tense/mood correspondences. The observed variation may be explained by a number of different factors which are not only related to the differences on lexical/syntactic level between the considered languages, but also to a number of pragmatic factors, including the process of translation. We also show that modeling tense/mood for MT requires additional information beyond the morpho-syntactic properties of the source, and we discuss tools for obtaining this information, which can (and should) be used in fu-

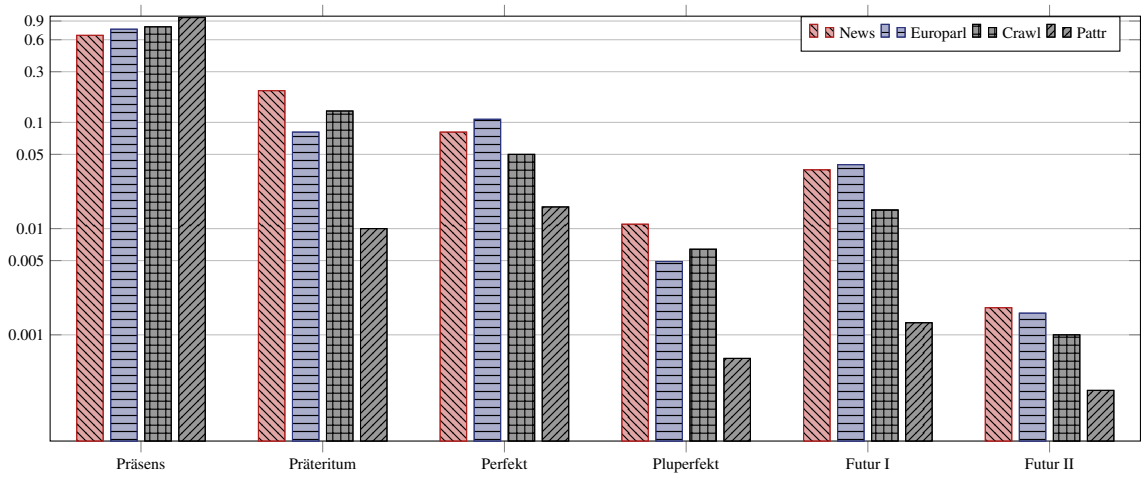

Figure 6: Relative frequencies of the indicative active tense forms in four different German corpora.

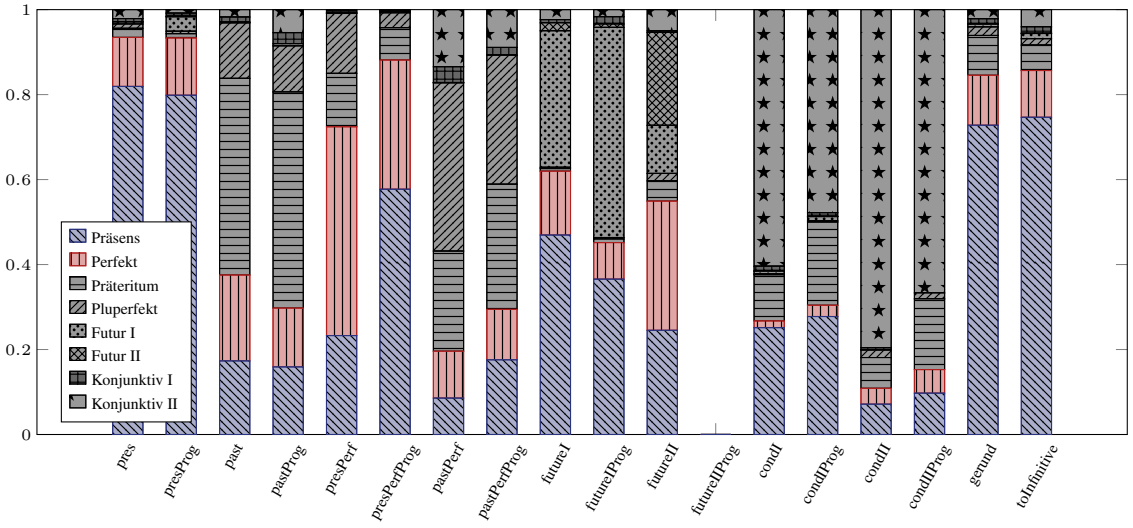

Figure 7: Distribution of tense translations derived from the News, Europarl and Crawl corpus.

Textual property	Lexical/syntactic level	Annotation tool availability
Tense	VC, main verb tense, mood, voice temporal expressions (NPs and PPs): head, preposition, adjective, adverb temporal ordering	POS tagging and parse trees TMVannotator (Ramm et al., 2017) TARSQI (Verhagen et al., 2005) POS tagging + parse trees TARSQI (Verhagen et al., 2005)
Aspect	auxiliary (combination)	POS tagging and parse trees + mapping rules
Aktionsart	event/state/progress subject NP: determiner, quantifier number mass, count	sitent (Friedrich and Palmer, 2016) parse trees semantic properties POS tagging WordNet
Domain/ genre		-
Reported speech		QSample (Scheible et al., 2016)
Conditional clauses		-

Table 3: Mapping of the different textual properties to the corresponding lexical/syntactic levels. Column *Tool availability* lists tools for automatic annotation of the English texts with the respective information.

ture modeling research. Beyond directly improving the modeling, an interesting future consideration would be to give the translation system user control of document-level tense and mood choices (e.g., by introducing a parameter for how choices for tense and mood in indirect speech should be made).

References

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.

Douglas Biber. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press.

Douglas Biber and Jesse Egbert. 2016. Using grammatical features for automatic register identification in an unrestricted corpus of documents from the open web. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 2.

Bernd Bohnet and Joakim Nivre. 2012. A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing. In *In Proceedings of EMNLP-CoNLL*, Jeju, Korea.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi. 2015. Findings of the 2015 workshop on statistical machine translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 1–46, Lisbon, Portugal. Association for Computational Linguistics.

Charles Copeland, Jacques Durand, Steven Krauwer, and Bente Maegaar. 1991. The Eurotra linguistic specifications. Technical report, Office for Official Publications of the Commission of the European Communities, Brussels/Luxembourg. Studies in Machine Translation and Natural Language Processing 1.

Eva Csipak. 2015. *Free factive subjunctives in German*. Doctoral thesis. Niedersächsische Staats- und Universitätsbibliothek Göttingen.

Annemarie Friedrich and Alexis Palmer. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of ACL*, Berlin, Germany.

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.

Adrià de Gispert and Jose B. Mariño. 2008. On the impact of morphology in English to Spanish statistical MT. *Speech Communication*, 50(11-12):1034–1046.

J.A. Hawkins. 2015. *A Comparative Typology of English and German: Unifying the Contrasts*. Routledge Library Editions: The English Language. Taylor & Francis.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of ACL, demonstration session*, Prague, Czech Republic.

Ekkehard König and Volker Gast. 2012. *Understanding English-German constrasts*. Number 29 in Grundlagen der Anglistik und Amerikanistik. Erich Schmidt Verlag.

Sharid Loáiciga, Thomas Meyer, and Andrei Popescu-Belis. 2014. English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling. In *Proceedings of the The 9th Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland.

Stella Neumann. 2013. *Contrastive register variation. A quantitative approach to the comparison of English and German*. Trends in Linguistics. Studies and Monographs. De Gruyter Mouton.

Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

José A. Olivas, Cristina Puente, and Andrea Tejado. 2005. Searching for causal relations in text documents for ontological application. In *Proceedings of ICAI*, Las Vegas, Nevada, USA.

Philipp Petrenz. 2014. *Cross-Lingual Genre Classification*. Ph.D. thesis, School of Informatics, University of Edinburgh, Scotland.

Anita Ramm and Alexander M. Fraser. 2016. Modeling verbal inflection for English to German SMT. In *Proceedings of WMT*, Berlin, Germany.

Anita Ramm, Sharid Loáiciga, Annemarie Friedrich, and Alexander Fraser. 2017. Annotating tense, mood and voice for English, French and German. In *Proceedings of ACL, demonstration session*, Vancouver, Canada.

Geoff Sammon. 2002. *Exploring English grammar*. Cornelson Verlag.

Marina Santini. 2007. *Automatic Identification of Genre in Web Pages*. Doctoral thesis. University of Brighton.

Christian Scheible, Roman Klinger, and Sebastian Padó. 2016. Model Architectures for Quotation Detection. In *Proceedings of ACL*, Berlin, Germany.

Serge Sharoff, Zhili Wu, and Katja Markert. 2010. The Web Library of Babel: evaluating genre collections. In *Proceedings of LREC*, Malta.

Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. 2012. Tense and aspect error correction for ESL learners using global context. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL): Short Papers - Volume 2*, Jeju Island, Korea.

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

Marc Verhagen, Inderjeet Mani, Roser Saurí, Robert , Knippen, Jess Littman, and James Pustejovsky. 2005. Automating Temporal Annotation with TARSQI. In *Proceedings of ACL, demonstration session*, Ann Arbor, Michigan, USA.

Harald Weinrich. 2001. *Tempus. Besprochene und erzählte Welt*, 6 edition. C.H.Beck.

Yang Ye, Li Fossum, Victoria, and Steven Abney. 2006. Latent Features in Automatic Tense Translation between Chinese and English. In *Proceedings of the Seventh SIGHAN Workshop on Chinese Language Processing*, Sidney, Australia.