Discontinuous Verb Phrases in Parsing and Machine Translation of English and German

Sharid Loáiciga & Kristina Gulordava
University of Geneva
Département de Linguistique
Centre Universitaire d'Informatique
sharid.loaiciga@unige.ch, kristina.gulordava@unige.ch

Abstract

In this paper, we focus on the verb-particle (V-Prt) split construction in English and German and its difficulty for parsing and Machine Translation (MT). For German, we use an existing test suite of V-Prt split constructions, while for English, we build a new and comparable test suite from raw data. These two data sets are then used to perform an analysis of errors in dependency parsing, word-level alignment and MT, which arise from the discontinuous order in V-Prt split constructions. In the automatic alignments of parallel corpora, most of the particles align to NULL. These mis-alignments and the inability of phrase-based MT system to recover discontinuous phrases result in low quality translations of V-Prt split constructions both in English and German. However, our results show that the V-Prt split phrases are correctly parsed in 90% of cases, suggesting that syntactic-based MT should perform better on these constructions. We evaluate a syntactic-based MT system on German and compare its performance to the phrase-based system.

Keywords: verb-particle split, discontinuous phrases, test suite, parsing, machine translation

1. Introduction

Discontinuous phrases are syntactic constructions which have proved to be hard for natural language processing (NLP) tasks. A multi-word expression (MWE), for instance, is hard to identify if its elements are separated by other words in a sentence (Sag et al., 2002). Discontinuous phrases can give rise to long and non-projective dependencies which pose problems for dependency parsing (McDonald and Nivre, 2011). But machine translation (MT) is probably the NLP task where discontinuity poses the most evident problems. Most of the modern MT systems are based significantly on phrase-based alignment models. While phrase-based MT manages very well the translation of frequent continuous phrases and their reordering, it performs poorly on discontinuous phrases.

In this paper, we focus on one particular discontinuous verb phrase construction. Specifically, we look at the verb-particle (V-Prt) split in English and German, such as in the examples below:

(1) take$_V$ the shoes off$_{Prt}$

(2) macht$_V$ schon wieder blau$_{Prt}$

From a syntactic perspective, these two languages behave very differently: in German, the particle position is clause-final and a particle can be separated from its verb by an embedded clause, while in English a particle can be separated from the verb only by its direct object and a very long split is impossible. However, from a MT perspective, these constructions are rather similar as they involve the same type of lexical items (verbs and particles) which must be translated in connection to each other for a correct output. Indeed, verb-particle split phrases are hard to translate as illustrated in (3)-(4) (translated using Google Translate), where the same type of error is produced.

(3) EN We are going to put$_V$ a video out$_{Prt}$ about this.
FR Nous allons mettre une vidéo sur à ce sujet.

(4) DE Nächste Woche händigt$_V$ er mir die Schlüssel aus$_{Prt}$.
EN Next week he handed to me from the key.

Except for the system of Galley and Manning (2010), which handles discontinuous phrases during decoding, phrase-based MT shows limitations in translating discontinuous phrases. Hierarchical MT, on the other hand, deals with the construction implicitly (Chiang, 2005; Chiang, 2007; Kaeshammer, 2015), but it is syntax-based MT models which are designed to tackle this problem intently, in particular for German (Sennrich and Haddow, 2015). Syntax-based MT relies on syntactic structure of the source language which is mapped to the target language. For these approaches, the correct parsing and identification of discontinuous verb phrases in the source language is therefore essential.

In the first part of the paper, we evaluate the parsing accuracy of particle-verbs in English and German using a widely-used state-of-the-art parser (Bohnet et al., 2013). The manual analysis of several hundred parsed sentences allowed us to evaluate parsing performance in the two languages and also extract a gold subset of V-Prt split constructions with correct parses for English. The set of verb-particle split constructions in English is, to our knowledge, the first publicly available and syntactically annotated corpus of these constructions and presents a valuable resource for linguistic analyses. In the second part of the paper, we manually analyse the alignment and MT quality of discontinuous V-Prt phrases for the English-French and German-English language pairs based on the translations of our gold test suites in English and German, correspondingly.

2. Data and methods

2.1. Corpora

For German, the test suite created by Schottmüller and Nivre (2014) is analysed. This test suite is composed of 236 sentences comprising 59 different particle verbs in their finite (split) form and in their non-finite form (118 total) and

59 non-particle verbs with a synonym meaning to the particle verbs, again in their finite and non-finite forms (118 total). Additionally, we looked at particle verbs at a large scale using the English-German IWSLT 2014 data (Cettolo et al., 2014), specifically the training set, composed of 171,721 TED Talks segments. Although both German corpora are parsed, a manual inspection of the results is reported for the verbs from the test suite only.

For English, we created a small corpus of verb-particle split constructions from scratch. We used the English-French word-level aligned version of the IWSLT 2014 data provided by the Second DiscoMT Workshop (Hardmeier et al., 2015). We used the English side for extraction of candidate sentences. The details of test suite development from the IWSLT data are reported in the section 3.2. In addition, we used the provided bitexts to evaluate the alignments between the particle verbs and their translations. Note that the TED corpus consists of talks transcripts. It contains therefore sentences of natural (but not spontaneous) speech.

The sizes of the corpora are summarized in the Table 1.

	Corpus	Size
German	Test suite	236 sentences
	IWSLT 2014	171,721 segments
English	(created) Test suite	157 sentences
	IWSLT 2014	179,404 segments

Table 1: Sizes of the corpora used in the evaluation (test suite) and for data extraction and MT training (IWSLT 2014).

2.2. Tools

To parse both English and German, we used the joint part-of-speech tagger and dependency parser of Bohnet et al. (2013) from the Mate tools package. This is a transition-based parser which demonstrated state-of-the-art performance on a number of languages and which is very fast and convenient to use since it does not require any pre-processing of the data. We used the pre-trained models for English and German available online.[1] For bidirectional word-level alignment of the English-German data, we used Giza++ (Och and Ney, 2003).

3. Parsing evaluation

3.1. German

German parsing turned out to be very good, with an accuracy of $\approx 96\%$. Although this figure comes from the evaluation of the test suite sentences, which are shorter and somewhat simpler than real corpus sentences. Table 2 shows a summary of the results of the evaluation.

Moreover, in the case of finite verbs, which present the V-Prt split, we measured the distance in tokens between the verb and the particle. The intuition here is that in a MT context, the farther away these two dependencies are, the less likely the alignment will be accurate and therefore the more likely poor translations will be produced. For the verbs from the test suite, distances between 2 and 4 are the most frequent. This result is confirmed by the 17,694 V-Prt split constructions found in the IWSLT data. In addition, this data shows that a verb and its particle might be up to 50 tokens apart.

Category	#	%
split-forms		
Dependency and PoS-tag are correct	54	23
Dependency and PoS-tag are wrong		
- Particle not identified	5	2
non-split forms		
Dependency and PoS-tag are correct	173	73
Dependency and PoS-tag are wrong		
- Main verb not identified	4	2
Total	236	100

Table 2: Analysis of German parsing.

3.2. English

For English, to our knowledge, there exist no publicly available corpus of verb-particle split constructions. We therefore created a new test suite comparable in size to the one in German. Our approach was different from Schottmüller and Nivre (2014): instead of creating the sample sentences, we mined sentences containing V-Prt phrases from the parsed IWSLT English transcriptions. This approach allowed us to obtain a corpus of verb-particle-split constructions as occurring in natural speech and evaluate the parsing performance on these constructions at the same time.

We first parsed the entire IWSLT corpus. We then extracted the sentences containing candidate verb-particle-split constructions. As we aimed for high precision, we constrained the identification of verb-particle constructions using the verb and the particle PoS tags and the obligatory head-child dependency between them. Furthermore, we extracted only the cases where a verb and a particle are separated by at least two words. This condition ensures that we obtain verb-particle-split constructions hard for MT, where a verb and a particle are separated by a noun phrase.

This procedure extracted 773 candidates. We then evaluated 362 cases out of them to estimate the parsing accuracy and extract a gold subset. The results are reported in Table 3. Overall, parsing attachment accuracy is relatively high (90%), only in 10% of the cases the head of a particle is incorrectly identified. This accuracy is comparable to that on German for split-forms (Table 2).

Note that an unambiguous identification of verb-particle phrases is not straight-forward. In some cases the particle can be analyzed as an adverb or a preposition in constructions such as *'move that part across'* or *'take the arena down to the village'*. Other constructions include double particles as in *'roll it back up'* and small clauses *'keep the lights off'*. For this paper, we have focused on 157 cases which we identified as canonical cases of verb-particle split constructions. We leave a more detailed linguistic analysis of the full set of verb-particle constructions for future work.

[1] https://code.google.com/p/mate-tools/downloads/list

Category	#	%
Dependency is correct		
- Unambiguous particle	157	43
- Particle as Adv	44	12
- Particle as Other	125	35
- *all*	*326*	*90*
Dependency is wrong	36	10
Total	362	100

Table 3: Analysis of English parsing.

Verb alignment	Prt alignment	#	%
	Prt → Prt	22	9
V → V	Prt → NULL	22	9
	Prt → other	11	5
V → NULL	Prt → V	1	0
V → other	Prt → NULL	2	1
	Prt → V	1	0
Non-split forms			
V → V		158	67
V → other		19	8

Table 4: Error analysis summary of word-level German-English alignment of particle-verbs in the test suite.

4. Word alignment

4.1. German

Here we present a manual error analysis of the 236 sentences from the Schottmüller and Nivre (2014) test suite and their word-level alignment from German to English. Since parallel data is required for word-level alignment, we translated the test suite into English. Aiming at high quality alignment, the test suite sentences were concatenated with supplementary data (Europarl data from WMT15 (Bojar et al., 2015), summing up to 2,056,965 sentences in each side of the corpus).

Table 4 presents the types of alignment patterns that we found. Approximately two-thirds of all verbs are aligned well. However, looking closely only at the 59 finite (split) V-Prt phrases present in the test suite, we can note that only 22 are properly aligned with their English translation. When aligned with English, a German particle-verb instance can correspond either to a particle-verb as well or to a single form. Only when there is a particle in both sides, the alignment is good (5), but this is not the case when a particle verb corresponds to a single form in English. Indeed, in 11 cases, the particles are aligned with other tokens in the English sentence (6), and in 22 cases, they are aligned with NULL (7).

(5) a. reg ab → *calm down*
b. rennt weg → *runs away*

(6) a. drehe um → *do pancake*
b. gebe aus → *grade*

(7) a. nimmt ab → *decreases* ∅
b. atmen ein → *inhale* ∅

NULL alignments such as those illustrated in (7) are potentially detrimental for MT quality since they add weight to incorrect translations. In particular, they affect non-compositional V-Prt phrases. For instance, 'anfangen' means *to start*, while 'fangen' means *catch*. Lining up 'fangen' (no particle) with *to start*, would only add noise to the model.

4.2. English

For English, we analyze the word-level alignment of the 157 V-Prt split cases identified at the parsing stage to point out the source of translation errors. As mentioned before, we look at the automatic alignments between these sentences and their French translations from the DiscoMT data. Table 5 shows the numbers for different types of alignment patterns. First, we checked whether an English verb is aligned to the corresponding French verb (or a part of the verb phrase including auxiliaries). This type of expected alignment occurs in 89% of cases. More interestingly, we observed that the particle is most frequently aligned to NULL (43%), similarly to the case of V-Prt split alignment in German. Another frequent option is the alignment of the particle to the noun in the French verb phrase (21%). The particle is aligned to the same French verb as the English verb in 15% of cases. In the remaining cases, the particle is aligned to some adverb or a preposition which are not connected to the English verb phrase translation. The results show that the particle alignment is less predictable compared to the verb alignment and can be a source of translation errors.

Verb alignment	Prt alignment	#	%
	all	*139*	*89*
	Prt → V	24	15
V → Vx	Prt → N	33	21
	Prt → NULL	68	43
	Prt → X	14	9
	all	*18*	*11*
V → other/NULL	Prt → V	6	4
	Prt → other/NULL	12	8

Table 5: Summary of English-to-French word-level alignment of verb-particle split constructions.

5. Machine Translation

Both phrase-based and syntax-based MT systems were used for evaluating the translation quality of the verb-particle split construction in the German test suite. We evaluated the English test suite as translated by a phrase-based system. However, since there were no readily available syntax-based systems for the translation between English and French, we instead show a comparison with Google Translate[2].

We reproduced the set-up of last year's Workshop on Statistical Machine Translation (WMT15) for training both German to English and English to French phrase-based systems. We relied on the Moses toolkit (Koehn et al., 2007) with GIZA++ (Och and Ney, 2003) for word alignment and used 5-gram language models built with lmplz and all

[2] https://translate.google.com/

monolingual data (Heafield et al., 2013). Approximately 4.5 million sentences were used for training, combining Europarlv7, CommonCrawl and News data. Optimization weights were tuned using Minimum Error Rate Training (MERT) (Och, 2003) and 3,003 sentences of News data.

For the syntax-based translation of the German test suite, we used the University of Edinburgh's system submitted to WMT15 (Williams et al., 2015). This system uses compound splitting as a pre-processing step and synchronous context free grammar (SCFG) with phrase-structure labels on the target side and the generic non-terminal label X on the source side (Durrani et al., 2014).

Importantly, we evaluate the quality of the translations only with respect to the verb-particle phrases, in other words, the translation of the sentence itself is not assessed. Moreover, we evaluate the MT of discontinuous verb-particle phrases qualitatively, focusing on the patterns of obvious errors and largely ignoring subtle differences in more or less acceptable translations. All marginal cases are therefore evaluated as 'good' translations and our results can be seen as an upper bound on the MT performance.

5.1. German

We have evaluated the translation of the German test suite taking into account the different types of forms which the verbs composing it can take. As mentioned earlier, this test-suite is composed of 59 particle verbs in their split (finite) form and in their non-split (non-finite) form, and 59 non-splittable prefix verbs in their finite and non-finite forms as well. The last three categories have been pooled together in Table 6, which contains the summarized results.

Unsurprisingly, the phrase-based system produces more problematic translations of this construction than the syntax-based system, which is known to perform better in the context of the translations between German and English. For instance, the syntax-based system produced more fluent translations in cases where the phrase-based system produced sentences with a German word order as shown in examples (8a) and (8b).

(8) Er wollte den Radfahrer nicht überfahren.
   a. *syntax-based* He did not want to cross the cyclists.
   b. *phrase-based* He wanted the cyclists not run over.

However, looking at the split-forms in particular, it can be noted that both systems have an important number of *wrong verb* translations, 37% the phrase-based system and 27% the syntax-based system. This indicates that the particle is not being considered at translation time, generating a verb which often times correspond to the meaning of the source verb without the particle. This reflects to a great extent the numerous inaccurate word alignments seen in Section 4.1., in particular NULL alignments. Examples (9) and (10) contain very common verbs which are mistranslated in this manner.

(9) Ich schreibe so schnell ich kann zurück.
   a. *phrase-based* I write as quickly as I can.
   b. *reference* I answer as quickly as I can.

(10) Er macht das Fenster auf.
   a. *syntax-based* It makes the window.
   b. *reference* He opens the window.

System	Category	#	%
Phrase-based	*split-forms*		
	Good translations	27	11
	Problematic translations:		
	- no main verb translation	9	4
	- source inserted	1	0
	- wrong verb	22	9
	non-split forms		
	Good translations	130	55
	Problematic translations:		
	- no main verb translation	17	7
	- source inserted	12	5
	- wrong verb	18	8
	Total	236	100
Syntax-based	*split-forms*		
	Good translations	40	17
	Problematic translations:		
	- no main verb translation	2	0
	- source inserted	1	0
	- wrong verb	16	7
	non-split forms		
	Good translations	149	63
	Problematic translations:		
	- no main verb translation	4	1
	- source inserted	13	6
	- wrong verb	11	5
	Total	236	100

Table 6: Error analysis summary of German-to-English MT

For the non-split forms, we observed that many verbs are not translated and the source is inserted instead, as in the example (11). Source insertions happen when the word in question is not seen at training time, and the problem is common in the context of the translation of compounds. This should be alleviated by the compound splitting preprocessing of the syntax-based system. However, we observed similar proportions of source insertions for both systems throughout the evaluation categories.

(11) Er wird sich bald abregen.
   a. *both systems* He will soon be abregen.

In contrast, we noticed an impact of the compound split preprocessing in the case of *wrong verb* translations of non-split forms. 'Splittable' infinitives are noticeable more accurately translated by the syntax-based than by the phrase-based system (11/59 and 18/59 respectively). This means, however, that ca. 20% of these infinitives are wrongly translated.

5.2. English

Overall, the performance of the phrase-based system on the English verb-particle phrases is rather meager: only 20% of the total verb phrases are acceptably translated. Among the 80% of the wrong translations we distinguish two large groups of errors directly linked to the alignment errors we

Category	#	%
Good translations	31	20
Problematic translations	126	80
- only verb is translated	83	53
- both verb and particle are translated	20	13
- source inserted	8	5
- other	15	10

Table 7: Error analysis summary of English-to-French MT

have identified previously, similarly to the German case. First, in 53% of the phrases only the verb is translated and the particle is ignored, yielding a translation which differ in meaning from the source, such as 'take off' translated into French *prendre* (lit. 'take') and not *enlever*. In other 13% of the cases, both the verb and the particle are translated independently of each other, yielding an incorrect literal translation such as *tourner sur* for 'turn on'. Other cases of unknown verbs and completely incorrect translations constitute 15% of the data.

It is worth noting that some cases of acceptable translations and unacceptable literal translations of verbs (where the particle was not translated) differ only with respect to the semantic composition of the verb-particle construction. In cases such as 'take (the shoes) off', mentioned above or 'make (these things) up', the literal translation does not make sense because the verb phrase meaning is non-compositional. However, in cases such as 'bring in' the meaning of the simple verb 'bring' is very similar to the meaning of the bigger phrase and the literal verb-only translation is acceptable. Since in our data, most of the verb-particle phrases are of the first type, the straight-forward phrase-based MT does not suffice as we see from the evaluation figures.

The results of Google Translate, which are very similar to our phrase-based system (i.e. around 50% of translations were identical), suggest that this system is phrase-based too. Compared to our system, Google Translate translates particles more often, leading to worse translation but has better performance at very frequent phrases such as 'turn the lights on'. It follows from this analysis that phrase-based systems are not sufficiently powerful to accurately translate discontinuous verb-particle phrases, and that a more complex syntactic-based translation could be appropriate even for syntactically similar languages such as English and French.

6. Related Work

Lohse et al. (2004) present a corpus-based syntactic study of the V-Prt split orders in English. In this work, 430 verb-split constructions were collected based on semi-automatic extraction from four corpora of written and spoken English. Unfortunately, this corpus was not publicly released. The authors used the collected data to study split versus joint order in connection with some syntactic and semantic factors such as the length of the noun phrase in transitive constructions and the semantic compositionality of verb-particle phrases. Compared to the work of (Lohse et al., 2004), our corpus of verb-particle split constructions in English has an additional layer of data, namely the parallel translations of the TED talks into French. Its application can be therefore two-fold - as a resource for MT evaluation and as a linguistic resource for in-depth analyses of verb-particle constructions in English.

In addition to compiling the German test suite we used in this study, Schottmüller and Nivre (2014) report a manual evaluation of their V-Prt constructions using the commercial systems: Google Translate and Bing Translator. They report 22% of problematic translations among the split-forms, a similar number to the one found in our own evaluation. Unlike this previous work, we also evaluated the errors of parsing and word-level alignment for the V-Prt phrases in the test suite.

In the framework of phrase-based translation the problem of verb-particle split has been tackled by preprocessing the data before training in order to normalize its morpho-syntatic properties. For instance, Collins et al. (2005) and Nießen and Ney (2000) prepend particles to the main verb. In the latter work, in addition, the authors split the compound forms (such as the infinitive forms in our German data). Splitting is a common technique for the treatment of compounds in general which makes all words of the compound known to the system reducing data sparseness (Koehn and Knight, 2003; Stymne, 2008; Weller et al., 2014).

7. Concluding Remarks

Verb-particle split constructions are common in Germanic languages, among which German and English are the most addressed in NLP. Here we have examined this type of constructions in connection to parsing and MT performance. First, we have complemented the German test suite of V-Prt constructions with an English translation and developed a new comparable English data set of V-Prt split constructions mined from unannotated spoken speech transcriptions. These two datasets can be used in future work to test language-independently the performance of MT systems on the frequent and infrequent discontinuous phrases.

Concerning the alignment evaluation, it was shown that particles are not well aligned in the German-English nor in the English-French language pairs. Most of the time, the particles separated from the verb are aligned with NULL which subsequently prompts MT systems to literal translations of the verbs such as that of examples (3)-(4). These errors can also lead to generation problems such as lexical inconsistency when translating in the other direction.

In regard to the MT evaluation, even when looking at controlled and short sentences as those of the German test suite, around 52% of the verb-particle splits were mistranslated by the phrase-based system, without taking into account the meaning shift caused by the particle. The syntax-based system showed better results (32% or errors), confirming that it is better conceived to handle discontinuous constructions. The MT errors are exacerbated when translating English sentences coming from spoken language and amount to 80%. Surprisingly, commercial MT systems do not seem to outperform the MT systems built for this study in translating the V-Prt split phrases.

Finally, while the new test suite of verb-particle constructions in English which we presented in this paper is small, it can be used to bootstrap parsing performance as new gold data in the future work. Despite its small size, the set covers many lexical items which can be useful for training of the parser. A larger set of verb-particle constructions can be then more easily obtained using the improved parsing model.

8. References

Bohnet, B., Nivre, J., Boguslavsky, I., Farkas, R., Ginter, F., and Hajič, J. (2013). Joint morphological and syntactic analysis for richly inflected languages. *Transactions of the Association for Computational Linguistics*, 1:415–428.

Bojar, O., Chatterjee, R., Federmann, C., Haddow, B., Huck, M., Hokamp, C., Koehn, P., Logacheva, V., Monz, C., Negri, M., Post, M., Scarton, C., Specia, L., and Turchi, M. (2015). Findings of the 2015 workshop on statistical machine translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, WMT15, pages 1–46, Lisbon, Portugal. Association for Computational Linguistics.

Cettolo, M., Niehues, J., Stüker, S., Bentivogli, L., and Federico, M. (2014). Report on the 11th IWSLT evaluation campaign. In *Proceedings of the International Workshop on Spoken Language Translation*, IWSLT 2014, Hanoi, Vietnam.

Chiang, D. (2005). A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 263–270, Ann Arbor, Michigan. Association for Computational Linguistics.

Chiang, D. (2007). Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Collins, M., Koehn, P., and Kučerová, I. (2005). Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the ACL*, page 531–540, Ann Arbor. Association for Computational Linguistics.

Durrani, N., Haddow, B., Koehn, P., and Heafield, K. (2014). Edinburgh's phrase-based machine translation systems for wmt-14. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 97–104, Baltimore. Association for Computational Linguistics.

Galley, M. and Manning, C. D. (2010). Accurate non-hierarchical phrase-based translation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 966–974, Los Angeles, California. Association for Computational Linguistics.

Hardmeier, C., Nakov, P., Stymne, S., Tiedemann, J., Versley, Y., and Cettolo, M. (2015). Pronoun-focused MT and cross-lingual pronoun prediction: Findings of the 2015 DiscoMT shared task on pronoun translation. In *Proceedings of the Second Workshop on Discourse in Machine Translation*, DiscoMT 2015, Lisbon, Portugal.

Heafield, K., Pouzyrevsky, I., Clark, J. H., and Koehn, P. (2013). Scalable modified Kneser-Ney language model estimation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia.

Kaeshammer, M. (2015). Hierarchical machine translation with discontinuous phrases. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 228–238, Lisbon, Portugal. Association for Computational Linguistics.

Koehn, P. and Knight, K. (2003). Empirical methods for compound splitting. In *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics*, EACL '03, pages 187–193, Budapest. Association for Computational Linguistics.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoli, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C. J., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Lohse, B., Hawkins, J. A., and Wasow, T. (2004). Domain minimization in english verb-particle constructions. *Language*, pages 238–261.

McDonald, R. and Nivre, J. (2011). Analyzing and integrating dependency parsers. *Computational Linguistics*, 37(1):197–230.

Nießen, S. and Ney, H. (2000). Improving SMT quality with morpho-syntactic analysis. In *Proceedings of the 18th International Conference on Computational Linguistics*, COLING 2000, pages 1081–1085.

Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29:19–51.

Och, F. J. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 160–167, Stroudsburg, PA. Association for Computational Linguistics.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A. A., and Flickinger, D. (2002). Multiword expressions: A pain in the neck for nlp. In *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing '02, pages 1–15, London, UK. Springer-Verlag.

Schottmüller, N. and Nivre, J. (2014). Issues in translating verb-particle constructions from German to English. In *Proceedings of the 10th Workshop on Multiword Expressions*, MWE, pages 124–131, Gothenburg, Sweden. Association for Computational Linguistics.

Sennrich, R. and Haddow, B. (2015). A joint dependency model of morphological and syntactic structure for statistical machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, EMNLP 15, pages 2081–2087, Lisbon, Portugal. Association for Computational Linguistics.

Stymne, S. (2008). German compounds in factored statistical machine translation. In *Proceedings of the 6th International Conference on Natural Language Process-*

ing, GoTAL-08, Gothenburg.

Weller, M., Cap, F., Müller, S., Schulte im Walde, S., and Fraser, A. (2014). Distinguishing degrees of compositionality in compound splitting for Statistical Machine Translation. In *Proceedings of the 1st Workshop on Computational Approaches to Compound Analysis*, pages 81–90, Dublin, Ireland.

Williams, P., Sennrich, R., Nadejde, M., Huck, M., and Koehn, P. (2015). Edinburgh's syntax-based systems at wmt 2015. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 199–209, Lisbon. Association for Computational Linguistics.