

How to Account for Idiomatic German Support Verb Constructions in Statistical Machine Translation

Fabienne Cap¹, Manju Nirmal², Marion Weller^{1,2}, Sabine Schulte im Walde²

¹ CIS, Ludwig-Maximilian University of Munich – cap@cis.uni-muenchen.de

² IMS, University of Stuttgart – (nirmalmu|wellermn|schulte)@ims.uni-stuttgart.de

Abstract

Support-verb constructions (i.e., multiword expressions combining a semantically light verb with a predicative noun) are problematic for standard statistical machine translation systems, because SMT systems cannot distinguish between literal and idiomatic uses of the verb. We work on the German to English translation direction, for which the identification of support-verb constructions is challenging due to the relatively free word order of German. We show that we achieve improved translation quality for verb-object support-verb constructions by marking the verbs when occurring in such constructions. Additional evaluations revealed that our systems produce more correct verb translations than a contrastive baseline system without verb markup.

1 Introduction

It is widely acknowledged in the NLP community that multiword expressions (MWEs) are a challenge for many NLP applications (Sag et al., 2002), due to their idiosyncratic behaviour at different levels of linguistic description. In this paper we address German support verb constructions (SVCs) in statistical machine translation.¹

Support-verb constructions, also known as *light-verb constructions*,² are multiword expressions combining a verb and a predicative noun. The verb neither contributes its full meaning to the construction, nor is the meaning completely void (Butt,

¹The work presented in this paper is part of the Master’s Thesis of Manju Nirmal, cf. (Nirmal, 2015).

²in German: *Funktionsverbgefüge*

2003; Langer, 2009). For example, the verb *take* does not contribute its full meaning to the SVC *take a bath*, but nevertheless its semantic contribution is different to the verb *make* in the SVC *make a bath* (Butt, 2003). Often, an SVC is close in meaning to a corresponding full verb, e.g., the SVC *make a contribution* is synonymous to the verb *contribute*. Table 1 presents examples for English and for German SVCs and synonymous full verbs, where the predicative nouns are embedded in a noun phrase (V+NP) or a prepositional phrase (V+PP).

	English	
V+NP	make a contribution	contribute
V+PP	take into account	consider
	German	
V+NP	einen Beitrag leisten lit. <i>a contribution achieve</i>	beitragen <i>to contribute</i>
V+PP	in Frage stellen lit. <i>in question put</i>	hinterfragen <i>to question</i>

Table 1: Examples of English and German SVCs.

Support-verb constructions are problematic for phrase-based statistical machine translation (SMT) systems, as these systems consider texts to consist of word sequences, without distinguishing between the literal meanings of the verbs vs. their idiomatic meaning within an SVC. In this paper, we will show that we can achieve improved translation quality by marking the verbs that occur within V+NP SVCs. The marking distinguishes the SVC verbs from independent occurrences of the verb and thus enables the SMT system to learn different translations for the different kinds of occurrences. We focus on German SVCs, which are particularly challenging due to the morphological richness and the relatively free word

order in German. While Carpuat and Diab (2010) included some English SVCs into their pilot study on evaluating MWEs through SMT, to our knowledge there is no other previous work on SVCs in the context of SMT.

2 SVCs in Statistical Machine Translation

Default translation: In SMT, translations are “learned” from parallel data. Out of a set of possible translations derived from that data, the SMT decoder selects the most probable one. Today, most SMT systems translate whole phrases instead of single words, which allows to take some context of the word into account. Moreover, a language model and an reordering model are consulted in order to promote fluent translations. Nevertheless it is often the most frequent translation of a word which is chosen by the decoder. For example, the German verb “*vertreten*” is most often translated as “represent” in the training data. A standard phrase-based SMT system thus considers “represent” as a suitable translation for “*vertreten*”. However, when occurring in the context of an SVC like “*die Auffassung vertreten*”, a translation into “**represent** the view” is clearly wrong. Instead, “*vertreten*” should in this case be translated into “take” in order to yield the correct translation into “**take** the view”. However, this is only one translation scenario. Sometimes, the German SVC is not translated into an English SVC but a different construction. For example, “*Auffassung vertreten*” is often translated as “being of the opinion that”. In other cases, the SVC is identical in both languages: e.g. “*Rolle spielen*” - “play a role”.

Dazu leistet die Effizienz des Vermittlungsverfahrens einen substanziellen Beitrag . <i>To that make the effectiveness of the codecision procedure a substantial contribution.</i>
The effectiveness of the codecision procedure has made a substantial contribution in this case .

Table 2: Example for a non-adjacent SVC.

Non-adjacent SVCs: If the verb and its object are directly adjacent, a phrase-based system with sufficient coverage of the SVC in question is likely to correctly translate the SVC as one phrase. However, if the verb appears isolated, which is not un-

common in German, it is much more difficult for the SMT system to recognize that the verb should not be translated by its “default”, but by the SVC-specific translation. The example in Table 2 illustrates that several words may occur between the components of the SVC “*Beitrag leisten*”.

Note that some SVCs allow for more intervening words than others. In Table 3, the comparison of the average distance between the verb and the noun within the two SVCs “*Beitrag leisten*” and “*Rechnung tragen*” shows considerable differences.

SVC	distance
Beitrag leisten <i>to make a contribution</i>	5.44
Rechnung tragen <i>to account for</i>	2.62

Table 3: Average distance of SVC components.

The mean distances are derived from 3,549 occurrences of the SVC “*Beitrag leisten*” and 1,868 occurrences of the SVC “*Rechnung tragen*” within the Europarl corpus (Koehn, 2005). They were calculated by subtracting the lower position in the sentence from the higher position for either the noun or the verb. Whenever the verb and the noun occurred directly adjacent, the score yields “1”.

Methodology: In order to enable the SMT system to distinguish occurrences of a verb within an SVC from independent occurrences, we add a special markup to the verbs occurring in an SVC. By introducing this markup, the translations for independent verbs with a literal meaning are separated from those of verbs occurring in an SVC context. Thus, the SMT system can learn the SVC-translation of a verb not only when it occurs directly adjacent to the noun, but also for SVCs with many intervening words between the components. In such cases, the SVC is chopped and stored in different phrases of the SMT system. For a standard SMT system without markup it is almost impossible to learn the correct translation of the verb.

3 Related Work

MWEs in general: Multiword expressions have been a recurrent focus of attention within theoretical, cognitive, and in the last decade also within computational linguistics: The workshops on multi-

word expressions attached to major CL conferences³ celebrated their 10th anniversary in 2014, and the SIGLEX-MWE has initiated three special issues in NLP journals.

After initial approaches mainly focused on characterising the computationally challenging properties of multiword expressions (such as Sag et al. (2002) and Villavicencio et al. (2005)) and automatically identifying various types of multiword expressions in corpora (such as Baldwin and Villavicencio (2002), Villavicencio (2003) and Bannard (2007) who extracted English particle verbs), the focus of interest moved towards deeper semantic models of specific types of multiword expressions and towards integrating multiword expressions into applications.

Compositionality of MWEs: A wide range of semantic approaches has been concerned with distinguishing degrees of compositionality within multiword expressions, addressing

- noun compounds (Zinsmeister and Heid (2004), Reddy et al. (2011), Schulte im Walde et al. (2013)),
- particle verbs (McCarthy et al. (2003), Bannard (2005), Cook and Stevenson (2006), Ramisch et al. (2008)),
- light-/support-verb constructions (Bannard (2007), Fazly et al. (2007), Fazly et al. (2009)),
- various MWE types (Lin (1999), Katz and Giesbrecht (2006), Fazly and Stevenson (2008), Evert (2009))

The most prominent approach exploring measures of association strength within multiword expressions, in order to distinguish literal from collocational interpretations, is probably (Evert, 2005).

Addressing the compositionality of multiword expressions is a crucial ingredient for lexicography and NLP applications, to know whether the expression should be treated as a whole, or through its parts, and what the expression means. Examples of applications that have profited from integrating the semantics of multiword expressions are Part-of-Speech Tagging (Constant and Sigogne, 2011), Parsing (Wehrli, 2014), Information Retrieval (Acosta et al., 2011), and SMT (Carpuat and Diab, 2010; Weller et al., 2014), see below for details.

³<http://multiword.sourceforge.net>

MWEs in SMT: Previous work regarding multiword expressions in SMT can be divided into *static* approaches, where the training data is modified in order to facilitate a standard SMT system to learn suitable MWE translations and *dynamic* approaches where the modification takes place in the phrase table of the SMT system.

Static approaches include (Lambert and Banchs, 2005), who first extract bilingual – English and Spanish – MWEs based on parsed data and then merge them into “super-tokens”, which later is treated as a unit by the SMT system. Similarly, Carpuat and Diab (2010) merge parts of English MWEs extracted from lexica into larger units in order to improve English to Arabic SMT. In addition, they increase the maximal phrase size from 5 in conventional systems to 10 words per phrase. More recently, Cholakov and Kordoni (2014) described a static approach to handle English phrasal verbs – extracted from lexical resources – for translation into Bulgarian, where the particles are usually not separated from the verbs.

While static approaches have shown to improve translation quality, they do not allow for context-dependent decisions on how to translate MWEs. Instead of modifying MWEs in the training data, *dynamic* approaches handle MWEs directly in the phrase table of the SMT system. Ren et al. (2009) present an approach to handle bilingual Chinese - English MWEs. These are extracted from domain-specific parallel text and then added as separate phrases to the training data. In a subsequent step, the resulting phrase table is then annotated with a boolean variable indicating the presence or absence of an MWE. This approach was then taken one step further by Carpuat and Diab (2010), who worked with longer phrases and indicated not only the presence, but also the number of MWEs in each phrase. Finally, Cholakov and Kordoni (2014) further improved the dynamic approach in that they, in addition to the number of MWEs in a phrase, also encoded linguistic features of the phrasal verbs they investigated, like transitivity or separability.

In terms of translation quality, both static and dynamic approaches performed more or less equally well, except for (Cholakov and Kordoni, 2014), who found considerable improvements for the dynamic approach incorporating linguistic features.

DE	“Sie wollen herausfinden, welche Rolle der Riesenplanet bei der Entwicklung des Sonnensystems gespielt hat.” <i>they wanted to find out, what role the giant-planet for the development of-the solar-system played has.</i>
EN	“They want to find out what role the giant planet has played in the development of the solar system .”

Table 4: German word order allows for many intervening words between a verb and its object, here: “*Rolle gespielt*”.

Relation to the presented work In this paper, we pursue a static approach, i.e. we modify the training data of the SMT system, but leave the system itself as it is. We extract MWEs directly from the parallel training data (like Lambert and Banchs (2005) and Ren et al. (2009)) using parsed data (to account for the flexible word order of German) and word association measures (similarly to Ren et al. (2009)). In contrast to previous static approaches, where the MWEs were joined together to form a single unit, we only mark the verb of a support verb construction. We have shown with the example of “*Beitrag leisten*” above that German word order allows for many intervening words between the two components. Joining German MWEs together may thus lead to highly fluent sentences.

4 Extraction and Markup of SVC verbs

This section provides more details on our methodology. The general procedure is done in five steps, with steps 1–4 explained in the following subsections, and step 5 described in Section 5:

1. extract verb-object pairs (on lemma-level) from the parsed training data
2. identify SVCs (on lemma-level) in this set using standard word association measures
3. create several SVC sets with different degrees of idiomaticity
4. re-visit the training data and mark the verbs of SVCs (on token-level) accordingly
5. run SMT systems trained on data with verb markup based on the different SVC sets (cf. Section 5)

4.1 Verb-Object Pair Extraction To obtain a set of SVCs, we first extract verb-object pairs from dependency-parsed data. In a second step, all of these potential SVCs are scored and ranked by association measures. The SVC candidates with the highest association scores constitute the set of SVCs to be marked in both the parallel training data as well

as in the data to be translated.

For extracting the SVC candidates, we follow the extraction method outlined in Scheible et al. (2013) who describe a set of guidelines to induce the complete set of argument and adjunct phrases from dependency-parses (Bohnet, 2010). While in this study we focus on *verb-object* pairs, our extraction method allows for an easy extension to also cover other types of SVCs, such as *preposition+noun+verb* triples.

The example given in Table 4 illustrates the need for parsed data when working with German: due to the flexible word order already illustrated in Section 2, verb and object are often not adjacent, but allow for the insertion of several phrases (*[the giant planet]_{SUBJ} [for the development [of the solar system]_{PP}]_{PP}*) or sub-ordinate clauses between them. Furthermore, parsed data allows for an extraction of verb-object pairs on lemma-basis in order to generalise over the morphological variants of verbs and nouns. From the example in Table 4, we would extract the verb-object lemma pair “*Rolle spielen*”.

4.2 Identification of SVCs The resulting list of SVC verb-object candidate pairs does not only contain idiomatic SVCs, but also literal verb-object combinations. In order to identify the subset of SVCs, we measure the association strength between the verb and the object. For this, we opted for the often-used *log-likelihood* measure implemented in the UCS-toolkit (Evert, 2005). Assuming that verb-object pairs with a high association score are likely to be idiomatic, we rank the SVC candidate pairs according to their association scores.

4.3 Datasets Based on the ranked list of verb-object pairs by a word association measure, we decided to investigate different thresholds to the log-likelihood scores in order to identify idiomatic SVCs among the set of verb-object pairs and thus approximate different degrees of idiomaticity. We set these thresholds at log-likelihood scores of 1,000, 500,

	training		testing	
	types	token	types	token
all	30,6572	1,102,166	794	881
freq \geq 5	25,610	713,734	461	537
LL \geq 1000	338	181,818	58	94
LL \geq 500	693	240,369	95	139
LL \geq 350	1,024	271,908	120	168
LL \geq 250	1,473	304,148	142	191

Table 5: Number of SVCs in the training data and test set when applying different log-likelihood (LL) thresholds.

350 and 250. Note that the degree of idiomaticity decreases with the loglikelihood score, while the amount of noise in form of literal verb-object pairs being erroneously taken for SVCs increases. Nevertheless, we performed no manual cleaning of these lists. According to the various thresholds, we obtained different sets of presumably idiomatic verb-object pairs to be marked for the SMT system, and all pairs occurring in the sets are considered SVCs.

Table 5 shows the number of all extracted verb-object pairs from the German part of the parallel data, and the number of pairs with a frequency ≥ 5 . Note that we discarded verb-object pairs with a frequency < 5 as we consider these to be too sparse to be translated adequately by an SMT system. Table 5 also shows the sizes of the resulting sets of SVCs, both for the training data and the test data.

4.4 Verb Markup For each of the SVC sets given in Table 5, the training data is re-visited and all verbs occurring within SVCs receive a special markup. Generally speaking, we follow here the same procedure as for the extraction. If a verb-object pair occurs in the list of SVCs, the verb is marked by adding the string “_SVC” to the verb. It is important to note that, while the list of SVCs is lemmatized, we keep the inflected verb form in the training data. By introducing this markup, independent verbs with a literal sense are distinct from verbs occurring in SVCs. The SMT system can thus distinguish these two types of verbs and learn different translations for them. The example given in Table 6 illustrates a marked occurrence of “*geleistet*” (in the context of “*Beitrag leisten*” (= “make a contribution”) as opposed to an independent occurrence, where “*geleistet*” should be translated literally into “achieved”.

In addition to annotating the source-side part of the DE-EN training data, we also need to annotate the

SVC	Das hat einen wichtigen Beitrag geleistet _SVC. <i>This has an important contribution made.</i>
	This has made an important contribution .
other	Ich glaube , dass sie sehr viel Gutes geleistet hat . <i>I believe, that it very much good achieved has.</i>
	I believe that it has achieved a great deal of good .

Table 6: Illustration of SVC markup on verbs.

source-side part of the data to be translated, i.e. the data set for parameter tuning and the test set on which we evaluate our systems.

5 SMT Experiments

In order to assess the impact of our SVC verb markup, we trained one baseline SMT system without markup and 4 different systems with our markup (one for each idiomaticity threshold, cf. Table 5). Each of our SMT experiments consists of the following steps:

1. add SVC verb markup to the parallel training data (as described in Section 4)
2. train the SMT system, including word alignment, construction of a phrase-table and a re-ordering table
3. tune translation parameters using minimum error rate training
4. translate the test set and evaluate the output against one human reference translation

In the following we give details on the data sets we used and some further technical details on our SMT systems. Apart from differing SVC verb markup, all systems are trained identically.

5.1 SMT training data

We trained our systems on data from the annual shared task for statistical machine translation, all of which are accessible for free download.⁴ For training, we take the training data from the shared task of 2009, which consists of roughly 1.5 million sentences composed of mainly Europarl (Koehn, 2005) and some news data. The English language model is trained on the monolingual training data of the 2009 shared task, which roughly consists of 22 million sentences. For parameter tuning, we used the test set of the shared task 2013 and for testing the most recent test set of 2014 (\sim 3,000 sentences each).

⁴www.statmt.org

Experiments	BLEU	BLEU
	tuning1	tuning2
Baseline	20.43	20.49
Exp1000	21.04	21.01
Exp500	21.08	21.01
Exp350	20.86	20.89
Exp250	20.85	20.84

Table 7: BLEU scores on the 2014 testset.

5.2 System Details

We used the Moses toolkit (Koehn et al., 2007) to train standard phrase-based systems with default configurations. We trained an English 5-gram language model using KenLM (Heafield, 2011). For tuning the feature weights, we applied batch-mira with -safe-hope (Cherry and Foster, 2012). In order to ensure stable tuning, we performed two subsequent tuning procedures with identical starting conditions and report on results for both of them.

6 Evaluation

In order to evaluate the translation quality of our systems in comparison to each other and also to a baseline without any markup, we performed a standard MT evaluation using the BLEU metric. In addition, we also performed a semi-automatic evaluation with a focus on verb translations.

6.1 Automatic MT Evaluation

It is common practise to evaluate the performance of an SMT system by comparing its output to one (or more) human reference translations. We follow this line and calculate BLEU scores (Papineni et al., 2001) for each of our systems. Our testset is taken from the 2014 shared task on statistical machine translation ($\sim 3,000$ words). We tested all BLEU scores for statistical significance using pairwise bootstrap resampling with sample size 1,000 and a p-value of 0.05⁵. Results are given in Table 7. Compared to the baseline, we found that all of our systems containing verb markup for SVC verbs lead to a significant improvement in terms of BLEU.

The fact that all investigated sets of automatically identified SVCs improve the translation quality in the same magnitude shows that no manual filtering

⁵Code to be obtained from www.ark.cs.smu.edu/MT

System	# sentences with at least one full verb
Reference	2,712
Baseline	2,378
Exp1000	2,412
Exp500	2,413
Exp350	2,412
Exp250	2,411

Table 8: Number of sentences produced by the systems, which contain at least one full verb.

of the SVC sets is required to improve translation quality. Even though the sets certainly contain literal verb-object pairs, their markup does not seem to decrease translation quality. In future experiments, we will investigate the effect of manual filtering the SVC lists on translation quality.

6.2 Improved Verb Translations

In addition to the standard evaluation using BLEU scores, we investigated the effect of the SVC verb markup on verb translations in general. In the past, we often observed that verbs are missing in the SMT output. Due to their primary role in the understanding of a sentence, each missing verb translation has a severe effect on the perception of translation quality of humans. In Table 8, we give the number of sentences in which at least one full verb has occurred (note that auxiliary verbs were discarded in this evaluation). From these absolute numbers, it can be seen that each of our systems produces more verbs when compared to the baseline.

In a subsequent evaluation, we compared verb translations separately for each sentence, taking the reference translation the baseline translation and the output of one of our systems (Exp250) into account. The results of this evaluation are given in Table 9. It can be seen that, compared to the baseline, our system yields more verbs that match the reference translation on lemma level (3,648 vs. 3,505).

system	lemma matches the reference		
Baseline	X		X
Exp250		X	X
#verbs	3,505	3,648	2,436

Table 9: Overview of verb counts. 'X' indicates a verb matching the reference verb on lemma-level.

(a) baseline: no verb translation, Exp250: correct translation of the SVC verb.

input	Sie wollen herausfinden, welche Rolle der Riesenplanet bei der Entwicklung des Sonnensystems gespielt hat. <i>They wanted to find out, what role the giant-planet for the development of-the solar-system played has.</i>
reference	They want to find out what role the giant planet has played in the development of the solar system.
baseline	You want to find out what role the <i>Riesenplanet</i> in the development of the solar system.
Exp250	They want to find out what role the <i>Riesenplanet</i> played in the development of the solar system.

(b) baseline: default translation of the verb, Exp250: SVC translation of the verb.

input	“Ich vertrete die Auffassung , dass eine hinreichende Grundlage fr eine formelle Ermittlung besteht, sagte er. <i>I take the view that a sufficient basis for a formal investigation exists, said he.</i>
reference	“I am of the opinion that a sufficient basis exits” for a formal investigation, he said.
baseline	I represent the view that a sufficient basis for a formal investigation is, he said.
Exp250	I take the view that a sufficient basis for a formal investigation is, he said.

(c) baseline & Exp250: same verb translation, but Exp250 with better noun translation.

input	UBS gab diese Woche bekannt , dass sie Schritte gegen einige ihrer Mitarbeiter unternommen habe <i>UBS announced this week, that they action against some of their employees taken have</i>
reference	UBS said this week it had taken action against some of its employees.
baseline	UBS was announced this week that they take steps against some of their staff have done.
Exp250	UBS was announced this week that they take action against some of their staff, after.

Table 10: Comparison of translation outputs from the baseline and Exp250.

Recall that this verb evaluation happened with respect to the verbs occurring in the reference set. We already have seen from the improved BLEU scores that our systems are more similar to the reference translation than the Baseline system. While BLEU scores are calculated on exact matches, the verb evaluation in Table 9 has shown that we produce also more verbs matching the reference on lemma level (thus abstracting over morphological variants). But even this number can only be seen as an approximation of the translation quality. Ideally, a later evaluation would include the German source sentence in the evaluation and reflect whether or not the present verb is a correct translation of the German verb or not (independent of which lexeme the human reference translator chose).

Finally, in Table 10, we give some interesting examples of SVC translations in the context of the whole sentence in which they occurred. In Table 10(a), our system was able to produce the SVC verb that was missing in the baseline translation. In contrast, the baseline produced a verb in Table 10(b), but instead of the SVC verb, a default translation of the verb was produced. This example is particularly interesting as the correct translation of the SVC by our system has no positive effect on the BLEU score, as the human reference translator chose a different construction to

translate the SVC. Finally in Table 10(c) we give an example where all systems produced the correct verb (though in a different tense form than the reference), but in addition, our system also yielded an improved translation of the SVC noun. The examples in Table 10 cannot be considered to be more than random samples, not strong enough to draw further conclusions from them. However, they show that a more detailed manual evaluation of the translation quality may reveal even more significant improvements of our systems.

6.3 Translation Probabilities

In this section, we study the effects of the verb markup on the resulting translation probabilities. By marking whether a verb appears in an SVC context or not, we expect to see a difference in the respective translation options and probabilities. Table 11 shows entries for translations and the respective probabilities for the verb *treffen* which often occurs in SVCs such as *Entscheidung treffen* (to make/take a decision), *Wahl treffen* (to make a choice) or *Vorkehrungen treffen* (to take precautions).

In the baseline, the predominant translation options are related to *meet*, with a second literal meaning represented by *hit*. Options for translating SVCs (e.g. *make/take*) are listed as well, but their trans-

Baseline		Exp1000			
<i>treffen</i>		<i>treffen</i>		<i>treffen_SVC</i>	
prob	transl.	prob	transl.	prob	transl.
0.295	meeting	0.315	meeting	0.237	take
0.105	meetings	0.112	meetings	0.176	make
0.086	take	0.074	take	0.032	will
0.059	make	0.048	make	0.022	decide
0.036	meet	0.039	meet	0.019	taken
0.013	be	0.012	be	0.012	reach
0.011	hit	0.012	hit	0.009	will take
0.010	affect	0.011	adopt	0.009	will make
0.010	adopt	0.011	affect	0.009	to take
0.007	taken	0.007	taken	0.009	to make

Table 11: Top-ranked translation possibilities (with probabilities) for the verb *treffen* in the baseline, and its unmarked and marked variants in Exp1000. Valid translations are highlighted.

lation probabilities are considerably smaller. The top-ranked translation possibilities for *treffen* in the Exp1000 system do not differ much from those in the baseline, but the probabilities for the literal translations (highlighted) are higher than those in the baseline, whereas the probabilities for translations in an SVC context are slightly lower compared to the baseline. We assume that the entries *make/take* for the literal translations of *treffen* were derived from usages in SVCs not listed in the set of SVCs on which the annotation for this system was based – keep in mind that 1000 was the highest of the thresholds used and thus resulted in a list of SVCs with a high level of idiomaticity.

When looking at the translation options for *treffen* in an SVC context, we find that there is a considerable change in comparison to the baseline and non-markup entries: translation options for the literal meaning (*meet/hit*) are no longer top-ranked, but instead there are verbs with a light meaning allowing for the respective English SVCs to be realized. While there are a number of variations of the same lemma (*take, taken, will take, to take*), there is also some lexical variation (*take/make/reach [a decision]*) and also one full verb (*decide*) equivalent to one of the SVCs in question.

The comparison of the translation options for the different uses of *treffen* in table 11 illustrates how the verb markup applied to verbs within an SVC separates between the literal translation(s) and those appropriate for an SVC context.

On a sidenote, the selectional preferences of the different usages of *treffen* also reflect its respective meaning: when used with the default meaning of *to meet*⁶, the typical object is likely to be a person whereas in the usage as part of an SVC, the object is an abstract concept like *decision* or *choice*.

7 Conclusion and Future Work

We presented an approach to handle SVCs in an German–English SMT system. By marking verbs that occur within an SVC on the source-side, literal translation options are separated from those appropriate in an SVC context. We investigated different degrees of idiomaticity which all lead to significant improvements in BLEU. An additional evaluation of verbs confirmed that the systems with SVC-markup produced more verbs than the baseline and that also an increased amount of verbs matched with the reference translation.

We assume that our strategy of marking the (limited) set of light verbs is not running risk of introducing data sparsity, but the question of how to decide on an optimal set of SVCs remains to be studied more thoroughly in future work. Moreover, we may want to further distinguish the verb markup: while the current markup separates literal translations from SVC-appropriate translations, we could in the future explicitly distinguish translations of *different* SVCs that share the same verb in the source language, but might need different translations as for example in *Massnahme ergreifen* (lit: “to grasp measures”, “to take measures”) and *Flucht ergreifen* (lit: “to grasp escape”, “to escape”).

An extension to different language pairs would also be interesting – the presented approach can easily be extended to other languages as long as enough data is available as a basis to extract a set of SVCs.

Acknowledgements

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 644402. Moreover, it was funded by the DFG grants *Distributional Approaches to Semantic Relatedness* and *Models of Morphosyntax for Statistical Machine Translation*, and the DFG Heisenberg Fellowship SCHU-2580/1.

⁶The same applies to the second literal meaning *to hit*.

References

- Otavio Acosta, Aline Villavicencio, and Viviane Moreira. 2011. Identification and treatment of multiword expressions applied to information retrieval. In *Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World*, pages 101–109, Portland, Oregon, USA.
- Timothy Baldwin and Aline Villavicencio. 2002. Extracting the unextractable: A case study on verb particles. In *Proceedings of the Sixth Conference on Computational Natural Language Learning*, pages 98–104, Taipei, Taiwan.
- Collin Bannard. 2005. Learning about the meaning of verb–particle constructions from corpora. *Computer Speech and Language*, 19:467–478.
- Colin Bannard. 2007. A Measure of Syntactic Flexibility for Automatically Identifying Multiword Expressions in Corpora. In *Proceedings of the ACL Workshop on A Broader Perspective on Multiword Expressions*, pages 1–8, Prague, Czech Republic.
- Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 89–97, Beijing, China.
- Miriam Butt. 2003. The Light Verb Jungle. Working Papers in Linguistics 9, Harvard University.
- Marine Carpuat and Mona Diab. 2010. Task-based evaluation of multiword expressions: A pilot study in statistical machine translation. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 242–245.
- Colin Cherry and George Foster. 2012. Batch Tuning Strategies for Statistical Machine Translation. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, volume 12, pages 34–35.
- Kostadin Cholakov and Valia Kordoni. 2014. Better statistical machine translation through linguistic treatment of phrasal verbs. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.
- Matthieu Constant and Anthony Sigogne. 2011. MWU-aware part-of-speech tagging with a CRF model and lexical resources. In *Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World*, pages 49–56, Portland, Oregon, USA.
- Paul Cook and Suzanne Stevenson. 2006. Classifying particle semantics in English verb-particle constructions. In *Proceedings of the ACL/COLING Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 45–53, Sydney, Australia.
- Stefan Evert. 2005. *The statistics of word cooccurrences: Word pairs and collocations*. Ph.D. thesis, University of Stuttgart, Germany.
- Stefan Evert. 2009. Corpora and Collocations. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook.*, volume 2 of *Handbooks of Linguistics and Communication Science*, chapter 58, pages 1212–1248. Mouton de Gruyter, Berlin.
- Afsaneh Fazly and Suzanne Stevenson. 2008. A distributional account of the semantics of multiword expressions. *Italian Journal of Linguistics. Alessandro Lenci (guest editor): "From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science"*, 20(1).
- Afsaneh Fazly, Suzanne Stevenson, and Ryan North. 2007. Automatically learning semantic knowledge about multiword predicates. *Language Resources and Evaluation*, 41:61–89.
- Afsaneh Fazly, Paul Cook, and Suzanne Stevenson. 2009. Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics*, 35(1):61–103.
- Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries. In *Proceedings of the EMNLP Workshop on Statistical Machine Translation*, pages 187–197.
- Graham Katz and Eugenie Giesbrecht. 2006. Automatic identification of non-compositional multi-word expressions using Latent Semantic Analysis. In *Proceedings of the ACL/COLING Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 12–19, Sydney, Australia.
- 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 the 45th Annual Meeting of the Association for Computational Linguistics, Demonstration Session*, pages 177–180.
- Philipp Koehn. 2005. Europarl: A parallel corpus for Statistical Machine Translation. In *Proceedings of the 10th Machine Translation Summit*, pages 79–86.
- Patrik Lambert and Rafael Banchs. 2005. Data-inferred multi-word expressions for statistical machine translation. In *Proceedings of Machine Translation Summit X*, pages 396–403.
- Stefan Langer. 2009. Funktionsverbgefüge und automatische Sprachverarbeitung. Habilitationsschrift, Universität München.

- Dekang Lin. 1999. Automatic identification of non-compositional phrases. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 317–324, Maryland, MD.
- Diana McCarthy, Bill Keller, and John Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 73–80, Sapporo, Japan.
- Manju Nirmal. 2015. Studying the effect of annotating idiomaticity of support verb constructions in statistical machine translation. Master’s thesis, University of Stuttgart.
- Kishore A. Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2001. BLEU: A method for automatic evaluation of Machine Translation. Technical Report RC22176 (W0109-022), IBM Research Division, Thomas J. Watson Research Center.
- Carlos Ramisch, Aline Villavicencio, Leonardo Moura, and Marco Idiart. 2008. Picking them up and figuring them out: Verb-particle constructions, noise and idiomaticity. In *Proceedings of the 12th Conference on Computational Natural Language Learning*, pages 49–56. Association for Computational Linguistics.
- Siva Reddy, Diana McCarthy, and Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218, Chiang Mai, Thailand.
- Zhixiang Ren, Yajuan Lü, Jie Cao, Qun Liu, and Yun Huang. 2009. Improving statistical machine translation using domain bilingual multiword expressions. In *Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications*, pages 47–54.
- Ivan A Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Computational Linguistics and Intelligent Text Processing*, pages 1–15. Springer.
- Silke Scheible, Sabine Schulte im Walde, Marion Weller, and Max Kisselew. 2013. A compact but linguistically detailed database for German verb subcategorisation relying on dependency parses from a web corpus. In *Proceedings of the 8th Web as Corpus Workshop*, Lancaster, UK.
- Sabine Schulte im Walde, Stefan Müller, and Stephen Roller. 2013. Exploring Vector Space Models to Predict the Compositionality of German Noun-Noun Compounds. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 255–265, Atlanta, GA.
- Aline Villavicencio, Francis Bond, Anna Korhonen, and Diana McCarthy. 2005. Introduction to the special issue on multiword expressions: Having a crack at a hard nut. *Computer Speech & Language*, 19(4):365–377.
- Aline Villavicencio. 2003. Verb-particle constructions and lexical resources. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 57–64, Sapporo, Japan.
- Eric Wehrli. 2014. The relevance of collocations for parsing. In *Proceedings of the 10th Workshop on Multiword Expressions*, pages 26–32, Gothenburg, Sweden.
- Marion Weller, Fabienne Cap, Stefan Müller, Sabine Schulte im Walde, and Alexander Fraser. 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.
- Heike Zinsmeister and Ulrich Heid. 2004. Collocations of complex nouns: Evidence for lexicalisation. In *Proceedings of Konvens*, Vienna, Austria.