VOLIMET: A Parallel Corpus of <u>L</u>iteral and <u>M</u>etaphorical <u>V</u>erb-Object Pairs for English–German and English–French

Prisca Piccirilli[1] and **Alexander Fraser[2,3]** and **Sabine Schulte im Walde[1]**

[1]Institute for Natural Language Processing, University of Stuttgart, Germany
[2]School of Computation, Information and Technology, TUM, Germany
[3]Munich Centre for Machine Learning, Germany
{prisca.piccirilli,schulte}@ims.uni-stuttgart.de

Abstract

The interplay of cultural and linguistic elements that characterizes metaphorical language poses a substantial challenge for both human comprehension and machine processing. This challenge goes beyond monolingual settings and becomes particularly complex in translation, even more so in automatic translation. We present VOLIMET, a corpus of 2,916 parallel sentences containing gold standard alignments of metaphorical verb-object pairs and their literal paraphrases, e.g., *tackle/address question*, from English to German and French. On the one hand, the parallel nature of our corpus enables us to explore monolingual patterns for metaphorical vs. literal uses in English. On the other hand, we investigate different aspects of cross-lingual translations into German and French and the extent to which metaphoricity and literalness in the source language are transferred to the target languages. Monolingually, our findings reveal clear preferences in using metaphorical or literal uses of verb-object pairs. Cross-lingually, we observe a rich variability in translations as well as different behaviors for our two target languages[1].

1 Introduction

Metaphor is a figurative device which allows us to understand and experience one (typically abstract) domain in terms of another (typically more concrete) domain (Lakoff and Johnson, 1980). For example, in the sentence *I'll <u>tackle</u> the challenging problem of metaphors in translation*, the abstract domain of dealing with a problem is expressed in terms of the more concrete domain of physically seizing and throwing down something/someone. Metaphorical language has long been recognized as a challenge for both human understanding and machine processing (Tong et al., 2021) and is not confined to monolingual settings. It extends into
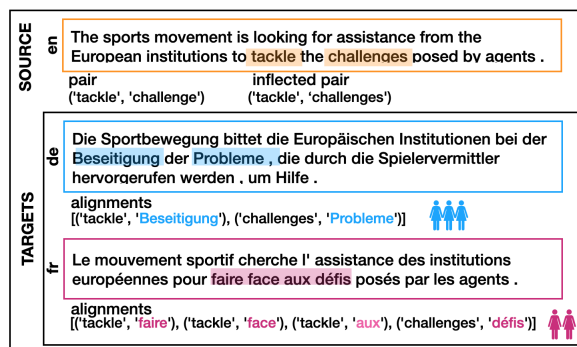

Figure 1: Example of gold standard alignments for a source English sentence containing the metaphorical verb-object *tackle challenge* to German and French.

cross-lingual territory, particularly in the realm of translation, where metaphors represent a hard nut to crack: they are not only very flexible in their structures and meanings, but also strongly depend on the involved languages and cultures (Schäffner, 2004; Kövecses, 2010). While the effort to automate the translation of figurative language using machine translation (MT) systems is underway, limited MT research explores the **contrast between metaphorical and literal language in translation** and its potential effect on translatability (van den Broek, 1981) and variability (Tong et al., 2021) in language production and generation.

To bridge this gap, we create VOLIMET, a parallel corpus of English–German and English–French sentences containing gold standard alignments of paraphrased metaphorical and literal uses of verb-object (VO) pairs (see example in Fig. 1). The corpus provides insights on the translation of metaphorical VO pairs and their corresponding literal paraphrases from the source language (SL) English, to the target languages (TL) French and German. For instance, given the metaphorical VO *tackle question*, is its literal VO counterpart *address question* equally frequent in natural language? How is it translated into other languages

---

[1]All data and guidelines are available at https://github.com/priscapiccirilli/VOLIMET

and to which extent is the metaphoricity transferred or preserved in the translation from SL to TL? How many different translations of this VO do humans produce, i.e., do we find one-to-one or one-to-many mappings between source and target languages?

In this paper, we present a comprehensive account of the corpus construction process and perform extensive monolingual and cross-lingual analyses. Monolingually, we seek to uncover patterns in the use of metaphorical vs. literal VOs. We find that considering the verb and its object as a unit provides a more nuanced representation than considering the verb on its own, crucial for accurate automatic processing. Cross-lingually, we observe a rich tapestry of translation variability, indicating one-to-many mappings between the source VOs and their target translations, where both metaphorical and literal uses are prevalent in the TLs. We further uncover differences in translations between the TLs, highlighting the need for flexible MT systems capable of reproducing this diversity.

Overall, our parallel corpus is meticulously crafted to encapsulate all these intricacies and represents a key resource in the endeavor to tackle the challenges posed by metaphorical language. In the future, it will also be of great use for machine translation research on metaphors.

2 Related Work

Translation Studies Metaphorical language represents an extremely common phenomenon (Shutova and Teufel, 2010) and has been of interest in translation studies when prescribing conditions for translating metaphors (van den Broek, 1981; Schäffner, 2004). As of today, the three translation modes from van den Broek (1981) remain the core choices in TL translations of SL metaphors: (1) a translation "sensu stricto" as in *le jour tombe–der Tag fällt* (lit. *the day falls*), which might lead to a semantic anomaly or innovation if the metaphor vehicles in SL and TL differ, (2) an onomasiological translation referred to as "substitution" where the SL and TL vehicles are translation equivalents sharing the same tenor, as in *le jour tombe–die Nacht bricht (her)ein* (lit. *the night falls in*); and (3) a discursive, non-metaphorical translation "paraphrase" as in *le jour tombe–es wird Abend* (lit. *it is becoming night*).

Machine Translation MT research incorporating figurative language has mainly been restricted to studies on the translation of structurally or semantically less flexible expressions, such as idioms (Huet and Langlais, 2013; Fadaee and Monz, 2018) and multi-word-expressions (e.g., noun compounds such as *flea market*; particle verbs such as *give up*; support verb constructions such as *play a role*) (Carpuat and Diab, 2010; Gamallo et al., 2019).

Cognitive Linguistics Stefanowitsch (2008) and Martin (2008) provide evidence for the cognitive function of metaphors in contrast to their literal counterparts, by demonstrating that people tend to use metaphors to explicate things. Metaphorical language tends to also be more emotionally-loaded than literal language (Citron and Goldberg, 2014; Mohammad et al., 2016; Piccirilli and Schulte im Walde, 2022) and may influence the way people conceptualize the world (Thibodeau and Boroditsky, 2011). Overall, there is empirical evidence for differences in using metaphorical in contrast to literal language, which we explore from a cross-lingual perspective in this work.

NLP Research has mainly focused on metaphor *detection* (Mu et al., 2019; Dankers et al., 2020) and *interpretation* (Bizzoni and Lappin, 2018; Mao et al., 2018), with the predominant idea to generate literal paraphrases for metaphorical expressions. More recently and more closely related to the current interest of this present work, we built a dataset of verb–object and subject–verb metaphorical vs. literal expressions used in large context and collected via crowd-sourcing annotations (Piccirilli and Schulte im Walde, 2021). In further work, we compared adapted computational models for discourse metaphor/literal interactions; the results from the human judgements showed the equal importance of metaphorical and literal usages, a behavior that computational models fail at mimicking (Piccirilli and Schulte im Walde, 2022). This reinforces the necessity for a more nuanced approach and attests limitations of word representations for metaphorically-used language.

Overall, rich interdisciplinary research offers insights on metaphors in monolingual settings, but less so in cross-lingual settings. Our work contributes to filling this gap by looking at the contrast of metaphorical and literal language, both from a monolingual and cross-lingual perspective.

3 Creating VOLIMET

We create VOLIMET, a comprehensive linguistic resource comprising various components that are

necessary to enhance not only our understanding of metaphorical language use but also its effect on translated text. VOLIMET encompasses an **extensive collection of English metaphorical and literal VO pairs**, as in *tackle vs. address question*, thoughtfully curated to provide paraphrases of one another (§3.1). VOLIMET also significantly enhances its contribution by **offering English sentences featuring these VO pairs**, meticulously extracted from parallel data (§3.2). The parallel nature of VOLIMET allows on the one hand **monolingual analyses** of metaphorical and literal VO pairs in context. On the other hand, it also enables a **cross-lingual exploration** of how these VOs are translated into German and French, making VOLIMET the first resource of metaphorical and literal VOs and their respective translations.

3.1 VO Pairs: Collection

At its core, VOLIMET consists of a set of metaphorical and literal verb-object pairs, which we (i) obtained from previous work and (ii) semi-automatically augmented.

Original Pairs   As a starting point, we collected a seed of 47 metaphorical VOs and their literal paraphrases from previous work (Mohammad et al., 2016; Shutova, 2010; Piccirilli and Schulte im Walde, 2021; Stowe et al., 2022), cf. Appendix A. For example, the basic sense of the verb *tackle* is used in the context of "to catch and knock down someone who is running",[2] which makes the idea of *tackling a question* physically impossible. The VO *tackle question* was therefore judged as being metaphorical, and *address question* was proposed as its literal paraphrase. Note that because the verb and its object are considered as a unit, there is no semantic ambiguity: no matter the context in which the VO occurs, *tackle question* is always used metaphorically, while *address question* is always used literally. Each VO pair in our original seed is composed of (i) a metaphorical verb and its literal paraphrase (*tackle/address*) and (ii) a direct-object noun (*question*) which makes the pair as a whole (verb-object) considered synonymous.

Extended Pairs   As we expect that our verb pairs naturally occur with more than one common object, we expanded the range of direct objects co-occurring with each of our 47 seed verb pairs. For example, the verbs in the pair *tackle/address* both

subcategorize *question* as a direct object but may also occur with *issue, challenge, matter*, to name just a few. Each of these nouns is not only a direct object of both verbs but also does not affect the paraphrase reading. We minimized human involvement for this task and applied a semi-automatic approach. Assuming that a direct object (*dobj*) occurring with both verbs within the same parsed corpus is likely to be a valid candidate, we automatically extracted all *dobjs* nouns if occurring with both verbs of a verb pair within the ENCOW corpus (Schäfer and Bildhauer, 2012; Schäfer, 2015). We first collected 157,437 additional arguments across our 47 pairs and then applied restrictions to reduce potential noise: We defined a frequency threshold of 15 for each object to occur with each of the two verbs, and discarded extracted nouns of less than two characters or labeled as "unknown" or "proper noun". Finally, we retained the 50 most frequent extracted objects for each verb pair. We automatically obtained 2,325 additional objects, from which we manually selected up to 10 valid candidates per verb pair. On average, each verb pair was augmented with six objects (max=11, min=1)[3], resulting in a total of 297 VO pairs. Note that the verb pairs remain the same, and the augmentation only applies to the nominal objects. For example, the original VO pair *drown/forget trouble* was augmented with the additional objects {*pain, problem, feeling*}. In Appendix A, we provide the original VO pairs and the sets of extended objects.

3.2 VO Pairs: Parallel Sentence Extraction and VO Alignments

The second part of the data collection consists of extracting natural English data containing any of our VO pairs. Because we aim to explore metaphors and their literal counterparts from both a mono- and cross-lingual perspective, we extracted data from parallel corpora.

Source and Target Languages   We chose English as our SL. As our TLs we chose German and French, two high-resource languages.

Parallel Corpus   Using existing parallel corpora such as the Europarl Parallel Corpus (Koehn, 2005) seemed the most straightforward approach, as it offers large amounts of data for our language pairs English–German (*en2de*) and English–French

---

[2] https://dictionary.cambridge.org/dictionary/english/tackle

[3] No additional valid *dobj* was found for four of our VO pairs, namely *push/sell drugs, wear/have smile, flood/saturate market, shipwreck/ruin career*.

	Met. VOs		Lit. VOs		Total
# instances	730	(12.59)	961	(10.92)	1,691
# VO pairs	58	(27)	88	(32)	31
# inflected VOs	135	(2.33)	203	(2.31)	–
Avg. sent. length	30.08	–	34.16	–	–

Table 1: Statistics on extracted monolingual English data: number of instances containing metaphorical and literal VOs (avg. instances per VO), number of extracted VOs (verb-specific) and number of inflected variants (avg. per VO) as well as average sentence length.

(*en2fr*). However, a major limitation of existing large parallel corpora is the (lack of) information regarding the language-pair direction. For example, the Europarl *en2de* corpus does not reliably ensure that the English text is always the actual SL nor that the parallel German text is the TL obtained *from* translating the SL English. This is problematic within the scope of our work: metaphorical language is a specific cognitive and linguistic phenomenon that is language- and culture-specific (Schäffner, 2004; Kövecses, 2010), hence the necessity to be aware of the original SL and the corresponding TLs. This limitation was previously noticed and addressed in Rabinovich et al. (2018), who publicly released a subset of the Europarl corpus providing accurate and reliable indications of translation directions. Their corpus contains 217,344 *en2fr* and 225,089 *en2de* parallel sentences, representing about 16% of the respective original Europarl datasets. We decided to use this corpus to build VOLIMET.

Extraction and Gold Standard Alignments We extracted all parallel sentences in which the source texts contain any of our 297 VO pairs, and performed word alignments using `fast-align` (Dyer et al., 2013). We ideally wanted to automatically obtain translations of the components of the pairs, but the accuracy of automatic alignments was rather sub-optimal, and resulted in many partial alignments. It also missed some crucial linguistic information or provided erroneous alignments; in fact, as soon as the translator took some creative liberty, the aligner generally failed to provide an alignment.

We therefore hired three German and two French speakers to correct potential errors in the automatically-obtained alignments. We defined clear guidelines on what and how to align. Note that we did not correct the word alignments of the *full* sentences, but focused only on the alignments between the SL *verb* and *object* of our VO pairs and their corresponding translations. This was a necessary and valuable step in creating VOLIMET: 66% and 90% of the *en2fr* metaphorical and literal parallel texts, respectively, needed their alignments to be corrected. For the *en2de* parallel texts, 92% and 85% of the metaphorical and literal data, respectively, had their alignments corrected.

Thanks to this human effort, we obtain **gold standard alignments** between metaphorical and literal English VO pairs and all their German and French human-produced translations. We release the annotation guidelines and the gold standard *en2fr* and *en2de* alignments for our metaphorical vs. literal VO pairs at `https://github.com/priscapiccirilli/VOLIMET`.

4 Quantitative Analyses

VOLIMET encompasses close to 3,000 *en2fr* and *en2de* parallel sentences containing a total of 114 metaphorical and literal VO pairs. We first perform in-depth monolingual (§ 4.1) and cross-lingual (§ 4.2) quantitative analyses. Monolingually, we shed light on the frequency of our VOs, their syntactic (non-)fixedness and the contrast in their metaphorical vs. literal usages. Cross-lingually, we look at the variability across translations. Then we explore whether metaphoricity vs. literalness in English is transferred to French and German during the translation process, and how the findings differ between the two TLs.

4.1 Monolingual (English) Analyses

The parallel nature of our corpora enables us to first perform quantitative analyses regarding the use of metaphorical and literal VO pairs in a monolingual setting, namely English. This way, we shed light on properties of metaphorical vs. literal language use in natural language. All statistics are reported in Table 1.

Starting with a set of 297 VOs, we wanted to see how **frequently** they occur in natural language, and whether we observe a **clear distinction between metaphorical VOs and their literal counterparts**. We extracted a total of 1,691 English sentences, 730 of them containing 58 **metaphorical VOs** (27 verbs-only[4]), and 961 sentences containing 88 **literal VOs** (32 verbs-only). We observe imbalances in the frequencies of VOs, e.g., we retrieved only two instances of *shape outcome*, but

---

[4]For example, *tackle question* and *tackle challenge* are 2 metaphorical VOs with 1 verb-only.

169 instances of *tackle problem*. Independently of their metaphoricity, each VO occurred with an average of 2.5 inflections (max=8) regarding both components (verb and/or object), e.g., *follow(ing) activity(ies), cause(s) death(s)*. We did not observe any major differences in sentence length, and the average sentence length is ≈30 words, both in the metaphorical and literal data.

The first 10 most frequent metaphorical and literal VOs represent 84% and 77% of the data, respectively (Table 2). The only pairs that seem to be equally frequent in their metaphorical and literal forms all stem from the same verb pair, i.e., *tackle* vs. *address (problem/question/issue)*. Paraphrased pairs are not equally frequent; in other words, either the metaphorical use or the literal alternative occurs in our data, e.g., *clause debate* is amongst the 10 most frequent VOs while its counterpart *end debate* does not occur once in our data at all. Out of the 58 metaphorical and 88 literal retrieved VOs, 31 of them are actual paraphrased pairs whose frequencies can be compared. We report in Figure 2 the proportions of frequencies for these 31 pairs. As we can see, only six of these pairs show equal frequencies of their metaphorical and literal uses (e.g., *break/end agreement*). There are 13 of them for which the literal use is more frequent than its metaphorical counterpart (e.g., *stimulate/fuel debate*) and 12 of them for which the metaphorical use is more frequent than its literal alternative (e.g., *boost/improve economy*). We will develop this observation in Section 5.

Quite a few additional verbs also display high frequencies when they are considered regardless of their objects. For example, the metaphorical VO *breathe life* is not part of the 10 most frequent metaphorical VOs but the verb *breathe* is, if we gather all its instances regardless of its objects (*life, confidence, value, hope*, etc.).

4.2 Cross-Lingual Analyses

VOLIMET is a valuable resource to exploit the annotations cross-lingually and analyze metaphorical vs. literal properties of translations. In this section, we quantify our findings for each language pair – *en2fr* and *en2de* – containing metaphorical vs. literal VOs. We provide qualitative analyses of these findings in Section 5. A detailed summary of the discussed statistics of VOLIMET can be found in Table 3.

Size and Frequency VOLIMET consists of 1,701 *en2de* and 1,215 *en2fr* parallel sentences, contain-
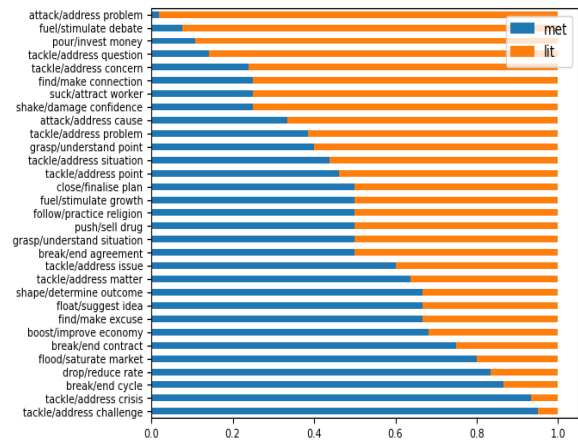

Figure 2: Proportions of VO pairs: metaphorical (blue) vs. literal (orange).

ing 114 of our source VOs. The obtained data is overall balanced for *en2de* regarding the amount of parallel sentences containing source metaphorical and literal VOs. However, the *en2fr* parallel dataset containing source metaphorical VOs is twice as large as the one containing source literal VOs. This is due to the fact that we had more German annotators; we aim to correct more *en2fr* alignments, in order to reach a balanced dataset.

Across the metaphorical and literal parallel datasets and language pairs, we find an average of 11 parallel sentences for each VO. This number however varies greatly across VOs. For instance, for the metaphorical *en2fr* dataset, we obtained only one parallel sentence containing the source VO *break contract* but 102 parallel sentences containing the source VO *find way*.

Met.		Lit.	
VO	verb-only	VO	verb-only
find way	tackle	address problem	address
tackle problem	find	make remark	make
tackle issue	close	address question	pose
tackle challenge	break	address issue	improve
close debate	boost	pose question	invest
tackle question	*float*	make comment	*reduce*
boost economy	*mount*	invest money	*understand*
tackle crisis	attack	improve situation	*get*
break cycle	*breathe*	address concern	stimulate
close case	*shape*	stimulate debate	*cause*

Table 2: The first 10 most frequent metaphorical and literal VOs in the SL English texts, in descending order. Underlined are the metaphorical vs. literal VO pairs that are equally frequent. In *italics* are the verbs that enter the top-10 when considered regardless of their objects.

	Met. VOs				Lit. VOs				Total
	en2de		en2fr		en2de		en2fr		
# parallel sentences	719	(12.40)	394	(8.76)	982	(11.55)	821	(11.40)	2,916
# VO pairs	58	(27)	45	(25)	85	(32)	72	(28)	114/297
# inflected VO pairs	133	(2.29)	103	(2.29)	198	(2.33)	154	(2.14)	–
# total translations	560	(9.66)	296	(6.58)	832	(9.79)	534	(7.42)	–
# unique translations	290	(5.67)	176	(4.44)	417	(5.48)	239	(3.82)	–
# Fig. translations (%)	109	(37.59)	112	(63.64)	128	(30.70)	88	(36.82)	–
# Lit. translations (%)	147	(50.59)	53	(30.11)	245	(58.75)	127	(53.14)	–

Table 3: Statistics on parallel datasets containing literal vs. metaphorical VO pairs: number of parallel sentences, VO pair inflections, all and unique (=type) translations (mean in brackets), the number of VO pairs covered in our datasets (unique verbs in brackets), and whether the respective VO translations were judged figurative or literal by humans (% in brackets). For example, the *en2de* dataset contains 719 **parallel sentences** with 58 **metaphorical VO pairs** (27 **verbs-only**[4]). We obtained 560 **total translations** for these 58 VO pairs (avg. 9.66 **translations per VO**), 327 **unique translations** (avg. 5.67 **translations per VO**), for which 42.41% of the **translations** were judged **figurative** and 57.59% **literal**.

Syntactic Variation Across language pairs and datasets, each VO presents on average two inflections (max=7 for *en2fr*, max=11 for *en2de*), from both components (verb and/or object), e.g., *found/finding excuse(s)*.

Variability in Translation We obtain a large array of translations with an average of eight translations per source VO, irrespective of the metaphoricity and the TL. These numbers are cut in half when looking at the number of unique translations. This still results in large variations in translations as each VO is aligned on average to four different individual translations for each language pair. Out of an average of 280 unique translations across language pairs, we find only 19 *en2fr* and 27 *en2de* translations that are translations of several metaphorical VOs, e.g., *répondre à question* is found as a translation for both *tackle challenge* and *tackle issue*, and 27 *en2fr* and 42 *en2de* translations that are translations of several literal VOs.

Similarly to the number of instances retrieved per VO, the number of unique translations varies across VOs: out of the 302 *en2de* parallel sentences containing the VO *address problem*, we observe up to 79 different (unique) translations. The number of instances per VO is highly correlated with the number of translations (average Spearman's correlation $\rho=0.99$ for *en2de* and $\rho=0.88$ for *en2fr*), i.e., the more a VO appears in natural language, the more (unique) translations are produced. As a matter of fact, none of the source metaphorical VOs for the *en2de* language pair results in only one

translation, and only three source metaphorical VOs produce one *en2fr* translation. A few source literal VOs obtain only one French/German translation (see Appendix B).

	FRENCH			GERMAN	
	Anno2	Anno3		Anno2	Anno3
Anno1	0.36	0.35	Anno1	0.42	0.36
Anno2	–	0.53	Anno2	–	0.43

Table 4: Cohen's κ scores across French and German annotators on judging the figurativeness of French and German translations, respectively, of metaphorical and literal English VOs.

Lost in Translation Beyond the *variability* in translation we described above, it is crucial to also quantify the *diversity* we encounter in translation: is metaphoricity/literalness transferred to the TLs, i.e., are metaphorical vs. literal VOs translated as metaphors vs. literal phrases, respectively?

We presented all unique German and French translations to three German and French native speakers and expert linguists, respectively, and asked them for a binary decision whether they judged each phrase[5] to be figurative[6] or literal. We

---

[5] We do not use the term "translation" in this annotation study, as we want to obtain judgements independently of the corresponding source text.

[6] We use the more general term "figurative language" for this annotation study, as the translations represent different *types* of figurative language, e.g., a metaphor, metonymy, an idiom, etc.

	FRENCH				GERMAN			
	Anno1	**Anno2**	**Anno3**	**Maj. Vote**	**Anno1**	**Anno2**	**Anno3**	**Maj. Vote**
Fig.	145 (41.31%)	178 (50.71%)	199 (56.70%)	178	320 (55.65%)	171 (29.74%)	181 (31.48%)	215
Lit.	206 (58.68%)	173 (49.29%)	152 (43.30%)	173	255 (44.35%)	404 (70.26%)	394 (68.52%)	360
Total		351				575		

Table 5: Human judgements (three expert French and German native speakers, respectively) regarding figurativeness for French (left) and German (right) translations, as well as the majority judgements.

report in Appendix C a detailed description of the annotation instructions.

Table 5 presents the number of translations that are judged figurative vs. literal, across annotators, as well as the majority vote. Overall, out of the 351 French translations, the judgements are rather balanced, i.e., the translations into French do not seem to be *clearly* figurative or literal, and this observation holds across annotators. The picture is different for translations into German: 2/3 of them are judged literal.

Judging figurative language is a difficult task (Zayed et al., 2019; Piccirilli and Schulte im Walde, 2021; Zhou et al., 2021), and we therefore observe disagreements across annotators on both languages. For example, the French VOs *plonger économie, jeter doute* (lit. *dive in economy, throw doubt*) and *comprendre signification, investir fond* (lit. *understand meaning, invest fund*) are unanimously judged figurative and literal, respectively. However, the VOs *trouver voie/moyen/issue/excuse* (lit. *find path/way (out)/excuse*) or *évoquer idée/question* (lit. *evoke idea/question*) were source of disagreement. Despite the difficulty of such a task, we obtain however fair-to-moderate inter-annotator agreement (IAA), with an average $\kappa = 0.41$ for both *en2fr* and *en2de*. Table 4 reports all κ scores between all annotators for both languages. We discuss in Section 5 some aspects on collecting human judgements regarding figurativeness which might have consequences for the analysis.

In the bottom part of Table 3, we also report the judgements regarding figurativeness of the translations with respect to the metaphoricity of the source phrases. The assumption that source metaphorical VOs are more likely to be translated figuratively and that literal source VOs are more likely to be translated with literal equivalences is confirmed for French. In fact, 63% of the translations from source metaphorical phrases are judged figurative, e.g., "float idea": *lancer idée* (lit. *throw idea*) rather than its literal paraphrase *suggérer idée* (lit. *suggest*

idea), and more than half of the translations from source literal phrases are judged literal (53%), e.g., "address question": *considérer question* (lit. *consider question*) rather than *s'attaquer à question* (lit. *attack question*), the figurative paraphrase. We observe different results for German translations. Even though literal translations from literal source phrases are largely favored (59%), e.g., "address crisis": *etw. gegen Krise tun* (lit. *do sth against crisis*), rather than *Krise bekämpfen* (lit. *fight crisis*), this correlation is not noted for source metaphorical phrases being translated figuratively. As a matter of fact, half of the translations of metaphorical source phrases are judged literal, e.g., "find excuse": *Entschuldigung haben* (lit. *have excuse*), and not *als Entschuldigung nehmen* (lit. *take excuse*), its figurative alternative.

5 Discussion

VOs Frequency From our original 297 metaphorical vs. literal VO pairs, there were more literal VOs which were extracted from the source part of our parallel corpus (75 vs. 58). For some VO pairs we find clear preferences for one option over the other, i.e., either the metaphorical or the literal VO of a pair is clearly more frequent. For instance, the metaphorical VOs *tackle crisis/challenge* are nine times more frequent than their literal paraphrases *address crisis/challenge*. One might think that this phenomenon has to do with the verb only, e.g., *tackle* is always favored over *address*. This is however not the case as *address* is favored over *tackle* when combined with other objects (e.g., *problem, concern*). For computational tasks such as text generation or machine translation, this finding gives support to the necessity to consider a verb along with its object: when considering paraphrases, one cannot rely on the frequency of the verb only, as more nuance might be brought by whichever argument is used along with that verb.

This finding should be taken with a grain of salt because we are lacking data points to gener-

228

alize across VO pairs. For instance, the verb pair *grasp/understand* only occurs with the object *point* in our data, and while the metaphorical usage *grasp* is favored over its literal counterpart, we do not know whether we would observe similar behaviors with other objects (e.g., *meaning, concern*, etc.) of that verb pair.

VOs Syntactic Variation   Looking at syntactic variations can shed light on (i) whether there are some clearly lexicalized VOs, but also (ii) whether there exist discrepancies between the paraphrase VO pairs, where one variant is more lexicalized than its paraphrase. Unlike idioms, metaphors are considered (syntactically) more flexible expressions that retain their metaphorical meaning if they undergo syntactic variations (Fazly et al., 2009; Kövecses, 2010). For example, *It's raining cats and dogs* cannot be replaced by *It's raining a cat and a dog* without losing its idiomatic interpretation. We however expect to observe a metaphorical VO such as *tackle question* in different morpho-syntactic forms, as in *tackling the questions* or *the question was tackled*, where the (metaphorical) meaning remains intact.

We have observed that both metaphorical and literal VOs that appear more than once in the data present up to eight different syntactic variations (three on average). None of the VOs therefore presents clear signs of syntactic fixedness, but there also exists no clear discrepancy in syntactic flexibility between metaphorical and literal VOs. In other words, this finding suggests that amongst our paraphrased pairs, there is complete consistency in terms of (non-)lexicalization.

Variability in Translation   As we discussed in the two previous paragraphs, we observe quite some diversity in the use of metaphorical vs. literal VO pairs in the source language, both in terms of frequency and syntactic variations. Indeed, we saw that some VOs present a preference for either the metaphorical or literal variant (see Figure 2), and that VOs display many inflections. Is this diversity encountered in the SL also observed in the French and German translations, respectively? We observe many syntactic variations in translation for both *en2fr* and *en2de*, indicating that the syntactic structure of the source VO is therefore not necessarily respected in the translation process. For example, the verb-object construction *break agreement* is translated as a noun-preposition-noun construction into French (*rupture de accord*, lit. *breaking of*

agreement). Regarding variability, we have seen that each source VO, regardless of its metaphoricity, is on average aligned to four translations, in both language pairs. Only very few VOs occurring more than once in the parallel corpus correspond to a one-to-one translation. Not only does this confirm that there are many ways to transfer one concept from one language to another, but also that humans tend to be very creative in the way they produce language. In other words, we generally find (large) variability in translation per SL concept (see Section 4.2).

We also notice slightly less variability in translation per literal VO than per metaphorical VO, i.e., one-to-many translations are more frequent for metaphorical VOs than for literal VOs. For example, the metaphorical concept of *boost* in *boost economy* does not have a *sensu stricto* metaphorical translation in French; we observed seven different translations for the 15 instances of the source metaphorical VO. However, we found only one (sensu stricto) French translation of the literal paraphrase *improve economy*. It therefore seems that if a metaphor in the SL does not have an equivalent in the TL, translators seem to show more creativity in their translation process. This perspective is especially interesting to keep in mind for natural language processing downstream tasks such as MT. Unlike the translation of idioms, which is either right or wrong (Volk, 1998; Huet and Langlais, 2013; Salton et al., 2014), efforts should be focused on building MT systems which are able to be more nuanced with respect to the use of metaphors.

Lost in Translation   To which extent is metaphoricity/literalness transferred from source to target in the translation process, and does the language pair matter? We observe a clear divergence in behavior between the two TLs, according to expert judgements. In French, metaphorical vs. literal uses in the SL tend to be preserved in translation, i.e., source metaphorical VOs tend to be translated into figurative phrases, and source literal VOs tend to be translated into literal phrases. This is however not the case for translations into German: overall, they have been judged more literal, even if the source text contained a metaphorical VO. Further investigation is needed in order to find the reasons behind this behavior.

Indeed, judging metaphoricity is a difficult task (Zayed et al., 2019; Zhou et al., 2021; Piccirilli and Schulte im Walde, 2021, 2022) for which many

aspects need to be taken into account and where the community has yet to find an optimal way to collect human judgements, e.g., number of annotators, binary vs. scale decision, but also the level of conventionality of metaphorical vs. literal phrases.

6 Conclusion

We presented VOLIMET, the first parallel corpus of English–German and English–French paraphrased metaphorical and literal verb-object pairs. Besides offering a novel lexicon of 297 metaphorical vs. literal VO paraphrase pairs, VOLIMET also provides their cross-lingual contexts at the sentence level.

We conducted substantial human work to provide gold standard alignments of source VOs to all their corresponding translations. We performed quantitative and qualitative analyses from both a monolingual and cross-lingual perspective. Monolingually, we showed that for some VO pairs, there exists a clear preference for either the metaphorical or the literal variant. It is however crucial to consider the verb and its object as a unit, as we observed apparent differences in behaviors when the verb is considered with or without its object. Cross-lingually, our findings revealed substantial variability in translations, i.e., one-to-many mappings between source VOs and their target translations. Finally, we investigated the extent to which metaphoricity/literalness gets preserved in the translation process. We found different behaviors between our two target languages, where French translations show equal use of metaphorical and literal language, while German tends to favor literal translations by a large margin.

Ethical Considerations

In the context of our annotation tasks, we collected judgements from human participants. For this, the participants were provided an Informed Consent Letter with the name and the contact of the investigators; the title, purpose and procedure of the study; risks and benefits for participating in the study; confirmation of confidential anonymous data handling; and confirmation that participation in the study is paid (12€/hour). The Informed Consent Letter was signed before the participants took part in the study.

Limitations

The creation of the VOLIMET parallel corpus and the research conducted represent significant advancements in understanding monolingual and cross-lingual metaphorical and literal language use and subsequently handling metaphors in machine translation. However, some limitations should be acknowledged. First, the corpus focuses on English–German and English–French translations and therefore does not fully capture the diversity of languages and translation challenges in other language combinations. Additionally, even though we provided clear instructions and examples of metaphorical vs. literal language, the human judgments collected for figurativeness and literalness in translations remain potentially subjective and may not represent the full spectrum of possible interpretations. Finally, the corpus' size and coverage as well as the number of verb-object pairs we used, might not encompass all possible metaphorical constructs and translation variations, requiring further expansion and exploration. These limitations highlight the need for ongoing research and the development of more comprehensive resources to enhance metaphor-aware machine translation systems.

References

Yuri Bizzoni and Shalom Lappin. 2018. Predicting Human Metaphor Paraphrase Judgments with Deep Neural Networks. In *Proceedings of the Workshop on Figurative Language Processing*, pages 45–55, New Orleans, Louisiana. Association for Computational Linguistics.

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, Los Angeles, California. Association for Computational Linguistics.

Francesca M.M. Citron and Adele E. Goldberg. 2014. Metaphorical Sentences Are More Emotionally Engaging Than Their Literal Coun-

terparts. *Journal of cognitive neuroscience*, 26(11):2585–2595.

Verna Dankers, Karan Malhotra, Gaurav Kudva, Volodymyr Medentsiy, and Ekaterina Shutova. 2020. Being Neighbourly: Neural Metaphor Identification in Discourse. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 227–234, Online. Association for Computational Linguistics.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.

Marzieh Fadaee and Christof Monz. 2018. Examining the Tip of the Iceberg: A Data Set for Idiom Translation. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, Miyazaki, Japan.

Afsaneh Fazly, Paul Cook, and Suzanne Stevenson. 2009. Unsupervised Type and Token Identification of Idiomatic Expressions. *Computational Linguistics*, 35(1):61–103.

Pablo Gamallo, Susana Sotelo, José Ramom Pichel, and Mikel Artetxe. 2019. Contextualized Translations of Phrasal Verbs with Distributional Compositional Semantics and Monolingual Corpora. *Computational Linguistics*, 45(3):395–421.

Stéphane Huet and Philippe Langlais. 2013. Translation of Idiomatic Expressions Across Different Languages: A Study of the Effectiveness of TransSearch. In A. Neustein and J. A. Markowitz, editors, *Where Humans Meet Machines: Innovative Solutions to Knotty Natural Language Problems*, pages 185–209. New York: Springer.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the 10th Machine Translation Summit*, pages 79–86, Phuket, Thailand.

Zoltan Kövecses. 2010. *Metaphor: A Practical Introduction*, 2nd edition. Oxford University Press, New York.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago.

Rui Mao, Chenghua Lin, and Frank Guerin. 2018. Word Embedding and WordNet Based Metaphor Identification and Interpretation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 1222–1231, Melbourne, Australia. Association for Computational Linguistics.

James H. Martin. 2008. A Corpus-based Analysis of Context Effects on Metaphor Comprehension. In A. Stefanowitsch and S. Th. Gries, editors, *Corpus-Based Approaches to Metaphor and Metonymy*, pages 214–236. De Gruyter Mouton.

Saif Mohammad, Ekaterina Shutova, and Peter Turney. 2016. Metaphor as a Medium for Emotion: An Empirical Study. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 23–33, Berlin, Germany. Association for Computational Linguistics.

Jesse Mu, Helen Yannakoudakis, and Ekaterina Shutova. 2019. Learning Outside the Box: Discourse-level Features Improve Metaphor Identification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 596–601, Minneapolis, Minnesota. Association for Computational Linguistics.

Prisca Piccirilli and Sabine Schulte im Walde. 2021. Synonymous Pairs of Metaphorical and Literal Expressions in Context: An Empirical Study and Dataset *to tackle* or *to address the question*. In *Proceedings of the Workshop DiscAnn*, Tübingen, Germany.

Prisca Piccirilli and Sabine Schulte im Walde. 2022. Features of Perceived Metaphoricity on the Discourse Level: Abstractness and Emotionality. In *Proceedings of the 13th Conference on Language Resources and Evaluation*, pages 5261–5273, Marseille, France. European Language Resources Association.

Ella Rabinovich, Shuly Wintner, and Ofek Luis Lewinsohn. 2018. A Parallel Corpus of Translationese. In *Proceedings of the 17th International Conference on Computational Linguistics*

and Intelligent Text Processing, pages 140–155, Konya, Turkey. Springer International Publishing.

Giancarlo Salton, Robert Ross, and John Kelleher. 2014. Evaluation of a Substitution Method for Idiom Transformation in Statistical Machine Translation. In *Proceedings of the 10th Workshop on Multiword Expressions*, pages 38–42, Gothenburg, Sweden. Association for Computational Linguistics.

Roland Schäfer. 2015. Processing and Querying Large Web Corpora with the COW14 Architecture. Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora, pages 28–34, Mannheim, Germany. Institut für Deutsche Sprache.

Roland Schäfer and Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 486–493, Istanbul, Turkey. European Language Resources Association.

Christina Schäffner. 2004. Metaphor and Translation: Some Implications of a Cognitive Approach. *Journal of Pragmatics*, 36(7):1253–1269.

Ekaterina Shutova. 2010. Automatic Metaphor Interpretation as a Paraphrasing Task. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies:*, pages 1029–1037, Los Angeles, California. Association for Computational Linguistics.

Ekaterina Shutova and Simone Teufel. 2010. Metaphor Corpus Annotated for Source - Target Domain Mappings. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, Valletta, Malta. European Language Resources Association.

Anatol Stefanowitsch. 2008. *Words and Their Metaphors: A Corpus-Based Approach*, pages 63–105. De Gruyter Mouton.

Kevin Stowe, Prasetya Utama, and Iryna Gurevych. 2022. IMPLI: Investigating NLI Models' Performance on Figurative Language. In *Proceedings*

of the 60th Annual Meeting of the Association for Computational Linguistics, pages 5375–5388, Dublin, Ireland. Association for Computational Linguistics.

Paul H. Thibodeau and Lera Boroditsky. 2011. Metaphors We Think With: The Role of Metaphor in Reasoning. *PLOS one*, 6(2).

Xiaoyu Tong, Ekaterina Shutova, and Martha Lewis. 2021. Recent Advances in Neural Metaphor Processing: A Linguistic, Cognitive and Social Perspective. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4673–4686, Online. Association for Computational Linguistics.

Raymond van den Broek. 1981. The Limits of Translatability Exemplified by Metaphor Translation. *Poetics Today*, 1(4):73–87.

Martin Volk. 1998. The Automatic Translation of Idioms. Machine Translation vs. Translation Memory Systems. Technical report, University of Zürich.

Yorick Wilks. 1978. Making Preferences More Active. *Artificial Intelligence*, 11(3):197–223.

Omnia Zayed, John P. McCrae, and Paul Buitelaar. 2019. Crowd-Sourcing A High-Quality Dataset for Metaphor Identification in Tweets. In *Proceedings of the 2nd Conference on Language, Data, and Knowledge*, pages 1–17, Leipzig, Germany.

Jianing Zhou, Hongyu Gong, and Suma Bhat. 2021. PIE: A Parallel Idiomatic Expression Corpus for Idiomatic Sentence Generation and Paraphrasing. In *Proceedings of the 17th Workshop on Multiword Expressions*, pages 33–48, Online. Association for Computational Linguistics.

A   Verb-Object Pairs

See Table 6.

B   VOs: One-only Translations

See Table 7.

Original met/lit VO-pairs	Extended sets of objects
absorb/assimilate knowledge	information, idea, culture, material, lesson, fact, content, experience, concept, thought
absorb/pay costs	fee, bill, tax, debt, interest, expense
abuse/misuse alcohol	drug, substance, medication, product
attack/address problem	issue, need, question, challenge, point, situation, topic, change, cause, matter
boost/improve economy	service, system, situation, process, education, work, business, result, number
break/end agreement	cycle, relationship, contract, marriage, process, pattern
breathe/instill life	sense, confidence, value, spirit, love, hope, idea, passion
buy/believe story	word, lie
cast/cause doubt	issue, fear
catch/get disease	idea, chance, information, result, message, point, call, problem, opportunity
close/end investigation	season, deal, case, debate, operation, story
close/finaliz(s)e deal	case, plan, arrangement, agreement, project
cloud/impair memory	ability, judgement, judgment, mind, vision, thinking, perception, understanding
colo(u)r/affect judgement	decision, choice, perception, experience, interpretation
deflate/reduce economy	cost, price, value, supply, wage, market, currency
devour/read book	article, story, page, novel, information, chapter, news
digest/comprehend information	meaning, material, fact, text, concept, idea, word, content, situation, message
disown/reject past	idea, policy, responsibility
drop/reduce price	cost, rate, temperature
drown/forget trouble	pain, problem, feeling
dull/decrease appetite	pain, sense, noise, feeling
find/make excuse	way, connection
float/suggest idea	theory, concept
flood/saturate market	X
follow/practis(c)e profession	religion, activity
frame/pose question	problem, challenge, issue, debate, concern, argument, idea, hypothesis
fuel/stimulate debate	growth, economy, interest, discussion, demand, activity, imagination, creativity, conversation
grasp/understand meaning	concept, issue, point, problem, situation, reason, language, idea, risk, question
juggle/manage job	project, work, life, career, school
kill/cancel proposal	project, bill, program, process, agreement, deal
leak/disclose report	information, document, story
mount/organiz(s)e production	event, campaign, conference, exhibition, demonstration, protest
poison/corrupt mind	system, process, soul, relationship
pour/invest money	fortune
push/sell drug	X
recapture/recall feeling	memory, moment, experience
shake/damage confidence	foundation
shape/determine result	life, outcome, success, strategy
shipwreck/ruin career	X
sow/cause doubt	death, confusion, chaos, conflict, panic, fear, violence, uncertainty, terror, hatred
stir/cause excitement	confusion, reaction, feeling, emotion
suck/attract worker	talent
tackle/address question	issue, problem, concern, challenge, situation, point, crisis, matter, inequality, task
taste/experience freedom	pain, life, joy
throw/make remark	comment
twist/misinterpret word	fact, meaning, comment, situation, information, message
wear/have smile	X

Table 6: Original metaphorical/literal VO-pairs and their sets of extended arguments. X means that no further objects were found according to our criteria (see Section 3.1 for a description of our extended pairs' selection).

	English	→	German	English	→	French
Met.				find excuse		trouver excuse
				flood market		inonder marché
				shake confidence		ébranler confiance
Lit.	organise production		Produktion organisieren	address challenge		relever défi
	read book		Buch lesen	end agreement		mettre terme à accord
				organise conference		organiser conférence
				organise production		organiser production
				pose problem		poser problème
				suggest idea		suggérer idée
				understand problem		comprendre problème
				understand reason		comprendre raison

Table 7: English metaphorical and literal VOs for which only one German/French translation was suggested.

C   Annotation Study: Guidelines

The purpose of this human annotation study was to evaluate the figurativeness of German and French translations of metaphorical and literal English phrases. We wanted to quantify the diversity we encounter in translation, for answering the question: are metaphorical vs. literal VOs translated as metaphors vs. literal phrases, respectively?

Using Google Forms, we compiled the 351 unique French and 575 German translated phrases, and asked three native French and German speakers and expert linguists, respectively, to judge whether the phrases were *figurative* or *literal*. For each phrase, we also provided one sentence containing that phrase, in case more context was needed for the binary decision. We estimated the task to take 3–5 hours, and we paid the annotators the (German) legal minimum wage of 12€/hour.

Description of the research study   In this project, we are interested in annotating whether French phrases are figurative or literal.

Purpose of the research study   The gold standard annotations will be used as training data for modeling the detection of figurative language.

What is figurative language?   As opposed to **literal** language, whose interpretation does not deviate from the word's defined and most frequent senses, the meaning of a **figurative** phrase is not simply composed of the common meanings of its components: its surface form and its underlying semantics do not directly correspond to each other. This is for example very clear when a phrase is an **idiom**: *"It's raining cats and dogs"*. This can be a bit more subtle when dealing with other forms of figurative language, such as **metaphors**, when one concept is viewed in terms of the properties of another: *"Let's kill the process"*, where the computational process is viewed as a living being. A **figurative** word/phrase can be recognized if it represents a violation of **selectional preference** in a given context: e.g., the verb "drink" normally requires a grammatical subject of type ANIMATE and a grammatical object of type LIQUID, as in (1-a). As a result, "drink" taking a "car" as a subject in (1-b) is an anomaly, indicative a figurative use of the verb.

(1)   a.   "She *drinks* tea"
      b.   "My car *drinks* gasoline" (Wilks, 1978)

Your task   You will evaluate whether phrases in French/German are figurative or literal.

- You will be given a list of French phrases. For each phrase, you will judge whether it is figurative or literal. Note that there is no ambiguity, i.e., each phrase has only one interpretation (figurative or literal).

- The phrases might not be as clear-cut as in the example (1). Do your best to make a judgement, based on the intuition you get from the explanation given above. There is no "right" or "wrong" answer!

- You can make use of whatever external resource you think might be helpful, e.g., dictionaries, etc.

- The phrase context (minimum two words) should be enough to emit a judgement. However, for each phrase, we provide one sentence containing the phrase, in case it helps you make a final decision.

- Do not leave any blank: always provide a judgement, i.e., Figurative or Literal

- We provide an example in Figure 3 (Note that this is a random annotation).

Phrases	Figurative or Literal?	Sentences
contracter maladie	Figurative ▼	Si l' un des animaux contracte la maladie , pourquoi faut -il abattre tout le troupeau ?
poursuivre remarque	Figurative ▼	Pouvez -vous répondre à ma question avant que je ne poursuive mes remarques ?
remédier à situation	Literal ▼	Espérons que ce rapport , ainsi que la directive sur l' emploi , permettront de remédier à cette situation .

Figure 3: Example of the annotation task set up for judging figurativeness of VO translations.