

# VOLIMET: A Parallel Corpus of Literal and Metaphorical Verb-Object Pairs for English—German and English—French

Prisca Piccirilli, Alexander Fraser, Sabine Schulte im Walde

{prisca.piccirilli,schulte}@ims.uni-stuttgart.de



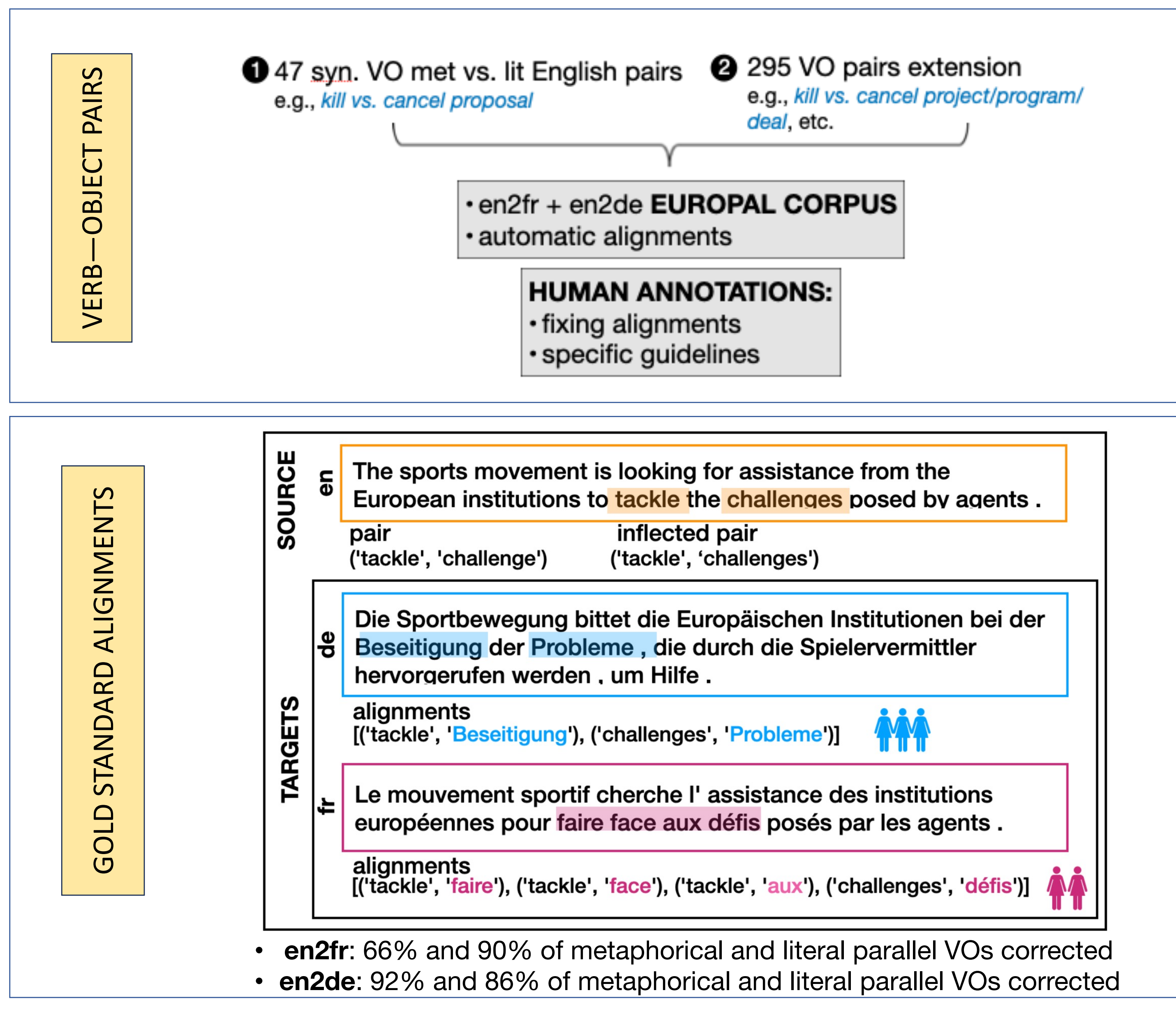
## Metaphorical and Literal Language in Translation

- ❖ Metaphors are a **hard nut to crack**: even more so **beyond monolingual** settings.
- ❖ **Contrast** between **metaphorical** and **literal** language in **translation** and its potential **effect on translatability** and variability in language production and generation.
- ❖ Translation **modes**.

- ❖ **VOLIMET**, a corpus of **2,916 parallel sentences** containing **gold standard en2de** and **en2fr alignments** of metaphorical verb-object pairs and their literal paraphrases.  
*e.g., tackle/address question.*
- ❖ The **parallel** nature of our corpus enables us to:
  - ❖ Explore **monolingual patterns** for **metaphorical vs. literal** uses in **English**.
  - ❖ 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.

	Source	Target	
<b>metaphor</b>	float idea	X	(sensu strictu) <b>metaphor</b>
<b>literal</b>	suggest idea	suggérer idée	<b>literal</b>
<b>metaphor</b>	float idea	• lancer idée lit. throw idea	(equivalence) <b>metaphor</b>
<b>literal</b>	suggest idea	• avancer idée lit. move forward idea	<b>literal</b>
<b>metaphor</b>	float idea	• émettre idée lit. emit idea	<b>literal</b>
		• évoquer idée lit. mention idea	

## Creating VOLIMET



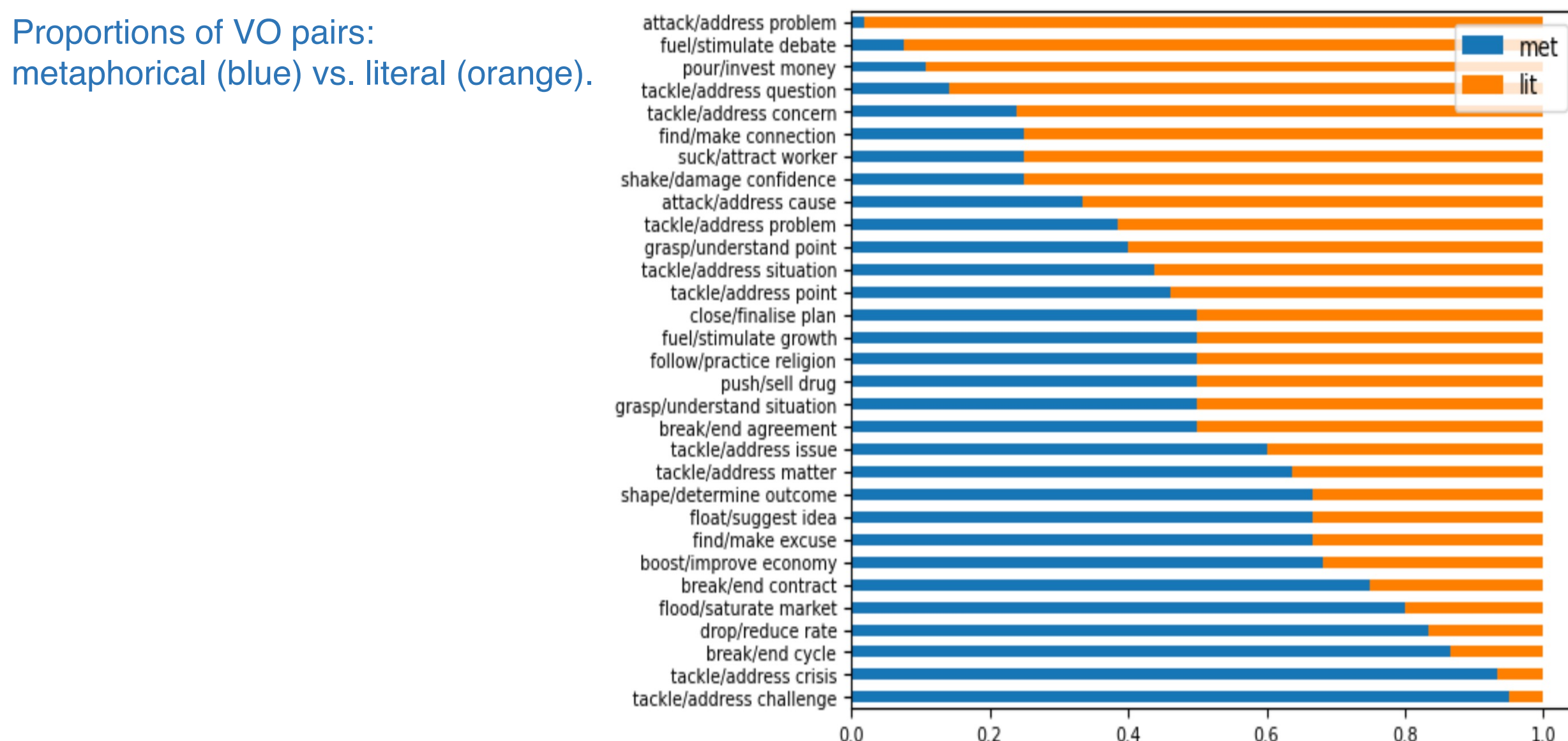
## Monolingual Analyses

	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	—	—

Statistics on extracted monolingual English data:

- number of instances containing metaphorical and literal VOs,
- number of extracted VOs,
- number of inflected variants

**Imbalances in the frequencies of VOs, e.g., 2 instances of *shape outcome*, but 169 instances of *tackle problem*.**



**Out of the 58 metaphorical and 88 literal retrieved VOs, 31 of them are actual paraphrased pairs whose frequencies can be compared.**

## Crosslingual Analyses

	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)	–

Statistics on parallel datasets containing literal vs. metaphorical VO pairs:

- number of parallel sentences,
- VO pair inflections,
- all and unique (=type) translations,
- number of VO pairs,
- whether the respective VO translations were judged figurative or literal by humans.

### SIZE AND FREQUENCY

- The obtained data is overall **balanced** for **en2de**. However, the **en2fr** parallel dataset containing source metaphorical VOs is **twice as large** as the one containing source literal VOs.

### SYNTACTIC VARIATION

- Average of **11 parallel sentences for each VO**. But variations **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*.

### VARIABILITY IN TRANSLATION

- Avg. **280 unique translations across language pairs**, we find only 19 **en2fr** and 27 **en2de** translations that are translations of several metaphorical VOs,  
*e.g., tackle challenge* and *tackle issue* → *répondre à question*
- Variation of unique translations across VOs**: e.g., 302 **en2de** parallel sentences containing the VO *address problem* → 79 different (unique) translations. The number of instances per VO is highly correlated with the number of translations (avg. 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**.

### LOST IN TRANSLATION

- Source metaphorical VOs more likely to be translated figuratively and that literal source VOs more likely to be translated with literal equivalences?

- French
- German Even though literal translations from literal source phrases are favored (59%), this correlation is not noted for source metaphorical phrases being translated figuratively.

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

Cohen's kappa scores

## Takeaways

- **VOLIMET**, the first parallel corpus of **English—German** and **English—French paraphrased metaphorical and literal verb-object pairs**.
- **Gold standard alignments** of source VOs to all their corresponding translations.
- **Mono-lingually**, there exists a **clear preference** for either the metaphorical or the literal variant for some VOs.
- **Cross-lingually**, our findings revealed substantial **variability in translations**, i.e., one-to-many mappings between source VOs and their target translations.
- French translations show equal use of metaphorical and literal language, while German tends to favor literal translations by a large margin.



EXPLORE  
VOLIMET