Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4402–4409
Marseille, 11–16 May 2020

A Multi-word Expression Dataset for Swedish

Murathan Kurfalı, Robert Östling, Johan Sjons, Mats Wirén[†]
Department of Linguistics, Stockholm University
Universitetsvägen 10C, SE-106 91 Stockholm
{murathan.kurfali, robert, johan.sjons, mats.wiren}@ling.su.se

Abstract

We present a new set of 96 Swedish multi-word expressions annotated with degree of (non-)compositionality. In contrast to most previous compositionality datasets we also consider syntactically complex constructions and publish a formal specification of each expression. This allows evaluation of computational models beyond word bigrams, which have so far been the norm. Finally, we use the annotations to evaluate a system for automatic compositionality estimation based on distributional semantics. Our analysis of the disagreements between human annotators and the distributional model reveal interesting questions related to the perception of compositionality, and should be informative to future work in the area.

Keywords: multi-word expressions, compositionality, distributional semantics

1. Introduction

A major challenge to the interpretation of natural language for humans and computers alike is non-compositionality. For instance, the compositional meaning of the verb-object pair *break the ice* could be paraphrased as "clearing the ice from the sea so that ships could pass". However, the same expression also has an established sense, "to overcome initial stiffness or reserve in a social setting" (Ammer, 2013). In this example, the literal (compositional) sense is very distant from the idiomatic (non-compositional) sense. In other cases, such as "search engine," only relatively minor semantic extensions of the inherent senses of 'search' and 'engine' are required to match the established meaning of "search engine." Being able to estimate the level of compositionality is important for several applications, including second language learning where idiomaticity is a major obstacle. However, the existing practical data is very limited in size and only available for several high resource languages. In the current study, we present the first resource for multi-word expressions (MWEs) in Swedish annotated for compositionality by humans[1].

2. Background

Several previous studies have performed computational investigations of compositionality, and in some cases created extensive annotated resources of multi-word expressions. The main focus in these studies has been on simple word bi-gram constructions (adjective–noun and noun–noun compounds as well as verb–particle pairs). Early computational work on compositionality relied on binary classifications by the authors, as in the semi-manual method of Korkontzelos and Manandhar (2009) to classify MWEs from WordNet (Fellbaum, 1998) as either compositional or non-compositional. Reddy et al. (2011) obtained compositionality scores for 90 English two-word noun compounds and their constituent words, using the Amazon Mechanical Turk (AMT) crowd-sourcing platform. Ramisch et al. (2016) extended this approach to a multilingual setting, covering 180 noun–noun and adjective–noun compounds in each of English, French and Portuguese. Similar data is also available for nouns in German (Roller et al., 2013), a language which makes extensive use of morphological noun compounding (where separate nouns are concatenated to a single noun). In terms of the range of the types covered, (Biemann and Giesbrecht, 2011) is the closest dataset to ours containing three types of constructions: adjective-noun, verb-subject and verb-object. Finally, a somewhat different approach to studying compositionality is to obtain paraphrases of each expression (Hendrickx et al., 2013).

3. Data Collection

3.1. Compositionality in Swedish

Much of the compositionality datasets available consist of either multi-word compounds in languages which lack morphological compounding (Reddy et al., 2011; Ramisch et al., 2016) or single-word compounds in languages such as German which do use morphological compounding (Roller et al., 2013). Swedish, like German, uses morphological compounds for many of the concepts present in existing datasets.

Since we are interested in applying our data to study MWEs, we have chosen not to use morphological compounds. Instead we study a syntactically much broader range of constructions, including nominal, prepositional and verbal MWEs (Baldwin and Kim, 2010). An example of the latter is "föra [någon] bakom ljuset" (literally: *bring [someone] behind the light*, figuratively *to deceive [someone]*). Syntactically this consists of a verb *föra* ('bring') with an attached prepositional phrase *bakom ljuset* ('behind the light') and an nominal object which may in turn be a complex noun phrase. There is a considerable amount of variation in form attested: the verb can be negated, modified by adverbs, or passivized. Other parts are fixed, such

[†] Author order is alphabetical. Author contributions are as follows. Conception of study: MK, JS; survey design and data collection: MK, RÖ, JS; implementation of distributional model: MK; formal specification of expressions: RÖ; drafting manuscript: MK, RÖ, JS, MW; final approval: MK, RÖ, JS, MW.

[1] The dataset can be accessed at https://github.com/MurathanKurfali/swedish-mwe-dataset

as the noun which is nearly always determined and in the singular. This results in a large number of possible forms, posing a challenge to traditional methods for extraction of idiomatic expressions. A formal specification for each of the expressions in our data can be found in Section 3.4.

3.2. Compound Set

Multi-word expressions were selected from the SALDO lexicon of Swedish words and multi-word expressions (Borin et al., 2013). A list of candidates was collected by the authors, taking care to avoid uncommon expressions that would run the risk of receiving less reliable annotations, as well as attempting to obtain a balance between different syntactic constructions. The final list consists of 96 MWEs, which are listed in Table 4.

The MWEs range from 2 to 4 lexically specified words, although as discussed above the actual number of consecutive words spanned by each MWE may vary significantly depending on syntactic form. A wide range of syntactic constructions are represented: prepositional phrases (often attached to a specific verb, noun or verb–object pair), adjective–noun pairs and verb–object pairs. These furthermore differ in how strictly specified the constituent words are with respect to inflectional forms.

3.3. Annotation Setup

Data collection was carried out through an online survey. Annotators were recruited through various informal channels, and required to have a native-level proficiency of Swedish. No metadata on respondents has been collected, including on the level of linguistic schooling. Given the distribution of the survey, we estimate that approximately a third of the respondents have some academic background in linguistics. The instructions were aimed at non-linguist readers, and avoided technical terms. The word *figurativeness* (Swedish: *bildlighet*) was used to describe the quality to be annotated. Three examples were given and briefly analyzed in the instructions, representing (in the judgement of the authors) high, medium and low levels of figurativeness, respectively. These examples are not among the 96 expressions in our dataset. It was made clear in the instructions, for both the examples and the actual survey, that there were no correct or incorrect answers, and that we only wanted the participant's individual judgement.

During the data collection, each participant was presented with a random subset of the whole set of 96 MWEs. Each subset consisted of 24 MWEs. The order of the MWEs within subsets was randomized between subjects, so as to avoid potential systematic biases from having scored the previous MWEs. For each MWE, participants were asked to annotate (a) the distance between the literal and the figurative meaning of the MWE as a whole and (b) how figurative each content word (nouns, verbs and adjectives) was:[2]

- State the distance between the literal and the figurative interpretation of *break the ice* with a score between 0 and 5, where 0 means there is no difference (i.e., there is only a literal interpretation of the expression), and where 5 means that the literal meaning and the figurative meaning do not correspond at all.
- On a scale from 0–5, how figurative is the word *break* in the expression *break the ice*?
- On a scale from 0–5, how figurative is the word *ice* in the expression *break the ice*?

In all cases, there was also an additional alternative "I do not know the meaning of this expression" beside the 6-point Likert scale.

The annotation set-up described above is largely based on Reddy et al. (2011) and Biemann and Giesbrecht (2011), with some modifications. Most importantly, in trying to capture the degree of compositionality of each expression, we use the following means:

- We ask explicitly for the degree of *difference* between the literal and figurative interpretations.
- Instead of presenting annotators with the target MWEs highlighted in example sentences, our MWEs were presented in isolation, without usage contexts, in order not to bias the annotators in their judgements of literal versus figurative with a (small-sample) corpus frequency of each possible interpretation.

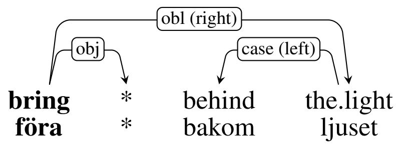
3.4. Syntactic Patterns

In order to facilitate computational analysis of our MWEs we specify each of them using the query language of the Turku NLP dependency search tool.[3] This allows anyone with access to a corpus of Swedish annotated with Universal Dependencies (McDonald et al., 2013) to extract instances matching the *form* or our MWEs.

For instance, the phrase "föra [någon] bakom ljuset" (to bring [someone] behind the light/to deceive [someone]) is coded as follows:

```
L=föra >obj _ >obl@R
     ("ljuset" >case@L "bakom")
```

This represents the following dependency structure, where **boldface** indicates that the slot can be filled by any instance of the given lemma, and asterisk (*) indicates that the slot can be filled by *any* subtree.



Note also that the word order is partially specified, with the oblique argument always following the verb, while the direct object may be located on either side of the verb. This syntactic template corresponds to the specification of the form of a construction in construction grammar, while leaving the meaning of the construction not formally specified. We do give a rudimentary example of inferring the semantics of our set of MWEs in Section 5., where we use a distributional model to compare the whole MWE to the sum of its parts.

[2]English examples are used here for illustration, but we only collected actual data for Swedish.

[3]http://bionlp-www.utu.fi/

Annotators	
Total number	72
Left after filtering	69
Annotations	
Number of MWEs	96
Content words in MWEs	184
Mean annotators per MWE	17.58
MWEs with $\sigma > 1.5$	9
Mean σ	1.09

Table 1: Annotation statistics

4. Dataset

A total number of 72 participants filled in the online survey. That is, each MWE is annotated 17 times on average (ranging between 11 and 25) (Table 1). To ensure the reliability of collected judgments, we applied two filters: (i) we calculate Spearman's ρ between annotators who filled the same subset of expressions. Any annotator whose mean correlation with other annotators is negative is removed. The purpose of this is to filter out annotators who inverted the scale or performed other gross errors. Two annotators were excluded by this criterion. (ii) For each annotator, we check how many standard deviations away on average s/he is from the rest for the expressions s/he annotated. We applied a threshold of 1.5 which resulted in excluding only one annotator.

For each MWE, we report the mean score along with the standard deviation (σ) in Table 4. Following the previous work (Reddy et al., 2011; Ramisch et al., 2016), we also report number of multi-word expressions with a $\sigma > 1.5$ which has been regarded as a test for annotation consistency. There are only 8 such expressions, suggesting a high inter-annotator agreement compared to previous studies.

Furthermore, we also check the relation between the multi-word expressions and their components. We calculate the correlation between the mean score of each MWE and that of their components. The results (Pearson's r= 0.92; Spearman's ρ= 0.93) indicate that there is a strong correlation between MWEs' compositionality and how literal their components are perceived.

5. Computational Model

We also report the results of a baseline model, accompanying our dataset.

5.1. Background

Compositionality prediction is the task of measuring to what extent the meaning of a given expression is constructed from its parts. The most popular line of research in this area focuses on employing distributional semantics models (DSMs) following the intuition that a MWE is likely to be compositional if it occurs in the same contexts as its components.

(Salehi et al., 2015) is the first work to utilize word embeddings showing that they can accurately model compositionality without requiring any labeled data. Cordeiro et al. (2016) conducts a systematic review of different DSM models by showing the effect of hyperparameters on the results. Cordeiro et al. (2019) extends that review by taking other languages into account where the results indicate that the CBOW model (as implemented in word2vec) and DSM based on positive pointwise mutual information (PPMI) achieves the best performance.

5.2. Model

Following the previous literature (Salehi et al., 2015; Cordeiro et al., 2016), we model the compositionality of a given multi-word expression as the cosine distance[4] between the expression and the center of its components in the vector space. We use two different DSMs: word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014), as they employ different techniques to learn the meanings of the words. Word2vec learns representations through prediction from context, whereas GloVe obtains representations by reducing co-occurrence matrices to lower dimensions. Hence, for each MWE with a vector representation of e_{MWE}, we predict its compositionality via

$$\mathrm{score} = d(e_{\mathrm{MWE}}, \sum_{w \in \mathrm{MWE}} e_w)$$

where $d(\cdot)$ is the cosine distance function, v_w is the embedding of the word w, and only words w that are content words are included in the sum.

5.3. Implementation details

We use a corpus of Swedish blog texts (Östling and Wirén, 2013), consisting of about 6 billion tokens. This is lemmatized using Stagger (Östling, 2013), and each consecutive sequence of words that match a MWE is collapsed into a special token representing that MWE.

We trained both the word2vec and GloVe models with their default configurations, the only exception being the embedding size which was always set to 300. All the expressions were present in the corpus, so we were able to generate an embedding for each MWE.

5.4. Results

We calculate the correlation between our metric of non-literalness and the human judgments for each MWE. Table 2 presents our results, which are discussed further and compared to previous work in Section 6.. As shown in previous work (Cordeiro et al., 2016), word2vec achieves a better performance than GloVe for the current dataset as well.

Since the frequencies of the MWEs roughly follow a Zipfian distribution (see Figure 1), we also tested whether embedding noise due to data sparsity had an effect on prediction accuracy for low-frequency items. However, there is no correlation between the frequency of a MWE and the performance of the computational model (Spearman's $r = -0.19$, non-significant difference from 0 correlation, $p > 0.05$) which indicates that higher frequency does *not* imply higher agreement with human ratings.

[4]We use distance instead of similarity as higher scores imply less compositionality in our dataset.

Data set	Pearson's r	Spearman's ρ
Reddy	0.634	-
Ramisch (Eng)	0.581	-
DiSCo$_{\mathrm{adj}}$	0.427	-
EVPC	0.489	-
GNC	0.361	-
Current study (CBOW)	0.384	0.388
Current study (GloVe)	0.314	0.308

Table 2: Our results compared to those for the other datasets. All results are obtained using word2vec-cbow as the DSM, where *Reddy, Ramisch,* DiSCo$_{\mathrm{adj}}$ are taken from Nandakumar et al. (2019) and *EVPC, GNC* are from Salehi et al. (2015).

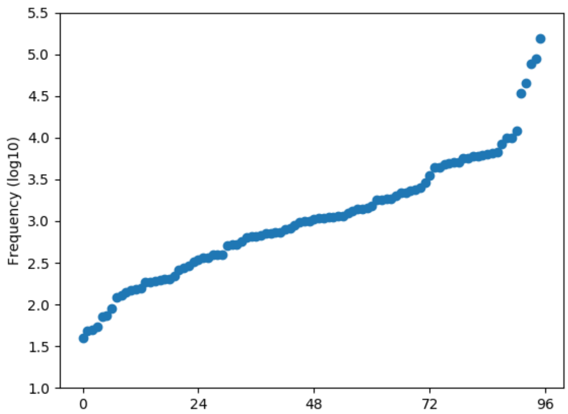


Figure 1: Frequency distributions of the MWEs in the blog corpus. The most frequent MWE is *gå vidare* "move on" occuring 156592 times and the least frequent one is *gnissla med tänderna* "grind one's teeth" with 40 instances.

6. Discussion

Table 2 shows the results on different datasets obtained by the same computational model. The correlations are relatively low, indicating that the model agrees poorly with human notions of compositionality.

Our results in Section 5.4. indicate that this is not simply an issue of data sparsity. Our next question is whether this difference is due to the differing syntactic complexity among our MWEs, and between our MWEs and those in previous work (which generally are shorter and simpler). A list of the strongest agreements and disagreements can be found in Table 3, which shows that the strongest disagreements (bottom of table) are in fact very simple expressions. The full table, omitted here for space reasons, shows no strong relation between the syntactic complexity of expressions (here counted as the number of dependency relations in the description) and the disagreement level (Spearman's $r = -0.02$, non-significant difference from 0 correlation, $p > 0.05$).

We hypothesize that the DSM can underestimate non-compositionality relative to humans when one or several of the constituent words are more often used in a sense derived from the expression itself, than in their core sense. This is a likely cause for expressions such as *ulv i fårakläder* (literally *wolf in sheep's clothing*, same figurative meaning as in English), where the constituent content words are rare except in a sense close to this expression.

MWE	Human (rank)	Model (rank)
snyta sig	0.62 (1)	0.26 (1)
se mellan fingrarna	4.35 (79)	0.86 (79)
[något] att hänga i julgranen	4.64 (90)	0.95 (91)
skörda frukterna	3.50 (53)	0.64 (54)
göra pengar	2.31 (18)	0.46 (17)
objuden gäst	1.46 (4)	0.34 (3)
föra [någon] bakom ljuset	4.06 (69)	0.78 (70)
...	...	...
mitt i prick	3.00 (37)	0.95 (90)
ulv i fårakläder	4.26 (75)	0.50 (22)
uppföra sig som folk	1.67 (6)	0.67 (60)
skaka hand	1.29 (3)	0.66 (59)
i backspegeln	3.74 (61)	0.31 (2)
ond cirkel	4.00 (68)	0.38 (4)
tak över huvudet	1.65 (5)	1.02 (95)

Table 3: Multi-word expressions ranked by agreement between human raters and the DSM prediction. The seven expressions with highest agreement (top) and seven with lowest agreement (bottom) are shown.

It is important here to note again that our distributional model excludes instances from our set of MWEs when computing word-level embeddings, so that an instance of *ulv i fårakläder* (*wolf in sheep's clothing*) will not affect the embedding for *fårakläder*. However, nearly all other uses of *fårakläder* invoke the meaning of this expression (falsehood, etc.), so from the point of view of the distributional model this word is nearly identical to the full expression *ulv i fårakläder*.

For future studies, it is important to clarify to annotators which sense of the constituent words to consider when estimating the level of compositionality. Alternatively, one could follow Hendrickx et al. (2013) and collect paraphrases. Presumably *ulv i fårakläder* (*wolf in sheep's clothing*) would be paraphrased using words such as *false* or *treacherous*, which do have a high degree of distributional similarity to *fårakläder* (*sheep's clothing*).

The model also overestimates non-compositionality in some cases, such as "bakom galler" (*behind bars*, with the same figurative meaning as in English, i.e. imprisoned). In this case the literal meaning is closely connected to the figurative meaning, in a way that humans but not a naive word-based distributional semantic model can discover. However, while these expressions are to some extent transparent to humans (or at least native speakers consider them so), they are still highly idiomatic in the sense that they can not easily be produced given the compositional rules of a language. As such, they pose a challenge to second-language learners and natural language generation systems alike.

7. Conclusion and Future Work

In the current study, we present the first Swedish multi-word expression dataset manually annotated for degree of compositionality. Consistent with previous studies, the results suggest that there is high correlation between how figurative humans consider a MWE to be, and how figurative they consider each individual word in that expression to be. Unlike most of the existing datasets, the current resource covers MWEs with a great variety of syntactic constructions, posing new challenges to the existing systems as suggested by our baseline scores. We employ the most widely used computational model for predicting compositionality, and show that its correlation with human-assigned scores is low to moderate. This agrees with previous work, which also found moderate correlation between model predictions and human annotations in a number of languages and expression types.

We hypothesize that capturing all the syntactic variations of a multi-word expression can lead to better representations. Following this intuition, as a future study, we plan to integrate the syntactic patterns we have prepared into the representation learning model.

8. References

Ammer, C. (2013). *The Dictionary of Clichés: A Word Lover's Guide to 4,000 Overused Phrases and Almost-Pleasing Platitudes*. Skyhorse.

Baldwin, T. and Kim, S. N. (2010). Multiword Expressions. In Nitin Indurkhya, editor, *Handbook of Natural Language Processing*, pages 267–292. CRC Press, New York, second edition.

Biemann, C. and Giesbrecht, E. (2011). Distributional semantics and compositionality 2011: Shared task description and results. In *Proceedings of the workshop on distributional semantics and compositionality*, pages 21–28. Association for Computational Linguistics.

Borin, L., Forsberg, M., and Lönngren, L. (2013). SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4):1191–1211.

Cordeiro, S., Ramisch, C., Idiart, M., and Villavicencio, A. (2016). Predicting the compositionality of nominal compounds: Giving word embeddings a hard time. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1986–1997.

Cordeiro, S., Villavicencio, A., Idiart, M., and Ramisch, C. (2019). Unsupervised compositionality prediction of nominal compounds. *Computational Linguistics*, 45(1):1–57.

Christiane Fellbaum, editor. (1998). *WordNet: An Electronic Lexical Database*. MIT Press. `http://wordnet.princeton.edu/`.

Hendrickx, I., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Szpakowicz, S., and Veale, T. (2013). SemEval-2013 task 4: Free paraphrases of noun compounds. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 138–143, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Korkontzelos, I. and Manandhar, S. (2009). Detecting compositionality in multi-word expressions. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 65–68. Association for Computational Linguistics.

McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., Bedini, C., Bertomeu Castelló, N., and Lee, J. (2013). Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria, August. Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nandakumar, N., Baldwin, T., and Salehi, B. (2019). How well do embedding models capture non-compositionality? a view from multiword expressions. In *Proceedings of the 3rd Workshop on Evaluating Vector Space Representations for NLP*, pages 27–34.

Östling, R. and Wirén, M. (2013). Compounding in a Swedish Blog Corpus. In Laura Álvarez López, et al., editors, *Computer mediated discourse across languages*, Acta Universitatis Stockholmiensis, Stockholm Studies in Modern Philology, New Series 16, pages 45–63. Stockholm University, Stockholm.

Östling, R. (2013). Stagger: An open-source part of speech tagger for Swedish. *North European Journal of Language Technology*, 3:1–18.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Ramisch, C., Cordeiro, S., Zilio, L., Idiart, M., and Villavicencio, A. (2016). How naked is the naked truth? a multilingual lexicon of nominal compound compositionality. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 156–161.

Reddy, S., McCarthy, D., and Manandhar, S. (2011). An empirical study on compositionality in compound nouns. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 210–218.

Roller, S., Schulte im Walde, S., and Scheible, S. (2013). The (un)expected effects of applying standard cleansing models to human ratings on compositionality. In *Proceedings of the 9th Workshop on Multiword Expressions*, pages 32–41, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Salehi, B., Cook, P., and Baldwin, T. (2015). A word embedding approach to predicting the compositionality of multiword expressions. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 977–983.

MWE	Literal translation	Translation	Score
torr i munnen	dry in the mouth	nervous	0.50 ± 0.65
snyta sig	blow one's nose	bog off	0.62 ± 1.41
rynka pannan	frown [at something]	same	0.94 ± 1.03
skaka hand	shake hands	same	1.29 ± 0.75
objuden gäst	uninvited guest	unwelcome guest	1.45 ± 1.08
tak över huvudet	roof over the head	not homeless	1.65 ± 1.28
uppföra sig som folk	behave like people	behave properly	1.67 ± 1.15
föröka sig	to procreate	same	1.67 ± 1.53
sträcka sig	stretch oneself	go as far as	1.89 ± 1.45
bakom galler	behind bars	same	1.90 ± 1.41
gnissla med tänderna	gnash one's teeth	same	1.91 ± 1.24
fullt hus	full house	packed/sold-out	2.00 ± 0.94
knipa käft	pinch mouth	keep quiet	2.00 ± 1.30
hålla tyst	keep quiet	keeping a secret	2.00 ± 1.45
sova sked	sleep spoon	spooning	2.05 ± 1.36
dagens sanning	truth of the day	[it is] the truth	2.13 ± 1.36
hålla tätt	keep tight	not leaking	2.15 ± 1.23
röd av ilska	red of anger	very angry	2.21 ± 1.47
göra pengar	make money	same	2.31 ± 1.31
betala för kalaset	pay for the party	foot the bill	2.41 ± 1.03
öppet vatten	open waters	same	2.56 ± 1.83
blotta sig	to expose oneself	same	2.59 ± 1.29
skatta sig lycklig	estimate oneself happy	consider oneself lucky	2.64 ± 1.84
kasta på sophögen	throw on the garbage pile	discard	2.67 ± 1.11
rapp i käften	quick in the mouth	quick-witted	2.67 ± 1.25
leva livet	live the life	enjoy oneself	2.67 ± 1.29
gå åt skilda håll	go in separate directions	part ways	2.68 ± 1.45
uppe med tuppen	up with the rooster	up early	2.75 ± 1.44
med vänster hand	with the left hand	half-heartedly	2.75 ± 1.79
hålla [sitt] ord	keep one's word	same	2.79 ± 1.58
rasa samman	collapse together	collapse, fail	2.84 ± 1.14
mannen på gatan	the man on the street	same	2.88 ± 1.18
tala i gåtor	speak in riddles	same	2.93 ± 1.24
gå i kras	go in crack	go to pieces	2.94 ± 1.71
sätta punkt	put a period	same	2.95 ± 1.64
gå vidare	go further	move on	3.00 ± 1.06
spilla tid	waste/spill time	waste time	3.00 ± 1.06
mitt i prick	right on dot	bull's eye/spot on	3.00 ± 1.32
hålla tummarna	hold the thumbs	crossing one's fingers	3.00 ± 1.59
hissa segel	raise sails	leave/prepare to leave	3.00 ± 1.97
hänga med huvudet	hanging with one's head	feeling down	3.05 ± 1.05
gå i fällan	go into the trap	fall into a trap	3.11 ± 1.17
på högvarv	in high gear	intensely	3.11 ± 1.29

Continued on next page

Continued from previous page

MWE	Literal translation	Translation	Score
enligt konstens alla regler	according to all the rules of the art	by the book	3.13 ± 1.31
gå sin egen väg	go one's own way	same	3.18 ± 1.46
[ha] händerna fulla	[have] one's hands full	same	3.21 ± 1.47
öm punkt	evil circle	vicious circle	3.25 ± 0.92
[ha] [något] på tungan	on one's tongue	on the tip of one's tongue	3.31 ± 1.21
[ha] [någon] på tråden	[have someone] on the thread	being in touch with someone	3.33 ± 1.25
kasta en blick	throw a glance	same	3.36 ± 1.17
lugna gatan	calm street	calm/cool	3.39 ± 1.11
röra på påkarna	move the legs	get moving	3.41 ± 1.19
slänga-ur sig	heave out of oneself	saying throughtlessly	3.42 ± 0.86
skörda frukterna	harvest the fruit	reaping the rewards	3.50 ± 0.87
på banan	on the track	in the game	3.60 ± 0.66
väga orden	weigh the/one's words	same	3.61 ± 1.06
kasta vatten	throw water	urinate	3.62 ± 1.21
med handen på hjärtat	with the hand on the heart	if truth be told	3.65 ± 1.24
kasta handsken	throw the glove	throw down the gauntlet	3.69 ± 1.10
spotta i glaset	spit in the glass	dislike alcohol	3.71 ± 0.88
visa tänderna	show one's teeth	same	3.73 ± 1.42
i backspegeln	in the rearview mirrow	same	3.74 ± 1.29
hämta sig	recover [oneself]	same	3.80 ± 1.11
brinna av iver	burning of eagerness	very eager	3.85 ± 0.85
blotta strupen	bare the throat	expose oneself	3.85 ± 0.85
ta i nackskinnet	take in the neck skin	take by the scruff of the neck	3.86 ± 0.91
höja [någon] till skyarna	elevate [someone] to the skies	praise [someone] to the skies	3.91 ± 0.90
ond cirkel	sore point	same	4.00 ± 0.85
ge [någon] sparken	give someone the kick	fire someone	4.00 ± 1.12
föra [någon] bakom ljuset	bring [someone] behind the light	deceive someone	4.06 ± 1.14
lägga på hyllan	put on the shelf	to shelve/abandon	4.11 ± 0.64
i krokarna	in nearabouts	[being] around	4.20 ± 0.98
fara åt helvete	go to hell	same	4.21 ± 1.06
skjuta sig i foten	shoot oneself in the foot	same	4.24 ± 1.27
fjärilar i magen	butterflies in the stomach	nervous	4.25 ± 1.04
ulv i fårakläder	wolf in sheep's clothing	same	4.26 ± 1.19
tappa tråden	lose the thread	same	4.27 ± 0.68
hänga på en tråd	hanging by a thread	same	4.33 ± 0.94
tidens tand	the tooth of time	the ravages of time	4.33 ± 0.94
se mellan fingrarna	look between the fingers	turn a blind eye to	4.35 ± 0.48
femte hjulet	fifth wheel	useless/redundant	4.42 ± 0.67
tappad bakom en vagn	dropped behind a wagon	stupid	4.44 ± 0.83
förlora ansiktet	lose face	same	4.45 ± 0.66
kräla i stoftet	crawl in the dust	show one's inferiority	4.45 ± 0.67
kort stubin	short fuse	same	4.46 ± 0.75
det fina i kråksången	the good in the crow's song	the good part/the beauty of it	4.50 ± 0.65
gå in i väggen	go into the wall	become mentally exhausted	4.50 ± 0.69

Continued on next page

Continued from previous page

MWE	Literal translation	Translation	Score
vända kappan efter vinden	turn the coat after the wind	turn one's coat	4.55 ± 0.50
hålla sig på mattan	keeping oneself on the rug	keeping oneself in the game/toe the line	4.56 ± 0.50
raka rör	straight pipes	straight talk	4.60 ± 0.49
[något] att hänga i julgranen	[sth] to hang in the christmas tree	[sth] to be proud of	4.64 ± 0.81
eld i baken	fire in the buttocks	in a hurry	4.69 ± 0.46
på lyset	on the light	drunk	4.73 ± 0.44
bita i det sura äpplet	bite the sour apple	reluctantly do something	4.78 ± 0.53
vända sig i sin grav	turn in one's grave	same	4.79 ± 0.41
torr bakom öronen	dry behind one's ears	same	4.79 ± 0.56

Table 4:Human rating scores (mean $\pm$ standard deviation) for the non-compositionality (figurativeness) of MWEs in our dataset.