

Contextual Choice between Synonymous Pairs of Metaphorical and Literal Expressions: An Empirical Study and Novel Dataset to *tackle* or to *address the question*

Prisca Piccirilli and Sabine Schulte im Walde

Institute for Natural Language Processing, University of Stuttgart, Germany
{[prisca.piccirilli](mailto:prisca.piccirilli@ims.uni-stuttgart.de), [schulte](mailto:schulte@ims.uni-stuttgart.de)}@ims.uni-stuttgart.de

Abstract

Research on metaphorical language detection and interpretation has produced a large number of resources mainly focusing on metaphoric vs. literal uses of specific expressions, and on metaphor paraphrases. As to our knowledge, however, no existing NLP resource provides a basis for understanding the choice between a synonymous pair of a literal and a metaphorical expression. E.g., why would one favor the use of *grasp a term* over *understand a term* in a given context, and does the preceding context prime for one or the other usage? We address these questions and provide an empirical study and a novel resource: Based on 50 pairs of English synonymous literal/metaphorical verb-object and subject-verb expressions in discourse, we asked participants in crowdsourcing experiments to (1) rate the degree of metaphoricality of a discourse, and (2) choose the expression that fits best. Our resource contains a total of 1,000 discourses and is ready to be exploited for computational research on discourse conditions for metaphorical vs. literal expression choices.

1 Introduction

Metaphors represent a "necessary, not just nice" element of everyday thought and communication (Ortony, 1975; Lakoff and Johnson, 1980; van den Broek, 1981; Schäffner, 2004), and frequently manifest themselves in general-domain text corpora (Gedigian et al., 2006; Shutova and Teufel, 2010). Accordingly, metaphors pose a real challenge across NLP applications, and research on metaphorical language detection and interpretation has produced a large number of resources. Up to now, however, there is no empirical study providing a basis for understanding the choice between a synonymous pair of a literal and a metaphorical expression, when they can be used interchangeably in a given context. Consider the following discourse: "For her, writing is an effective tool to express your

viewpoints [...] To write is already to choose, thus, writing should be done along with a critical mind and a caring soul. [...] Reading lets her travel to far-off imagined places and situations." This discourse might be followed by "She also learns a lot from (i) *devouring*/(ii) *reading books*, especially from the socio-political and historical ones.", where both (i) and (ii) seem equally acceptable.

The underlying choice leads to the following research questions: Why would one favor the metaphorical expression *devour a book* over its literal alternative *read a book*, or vice versa? Is the choice driven by the preceding context? If so, to which extent? These are necessary questions to tackle in order to build a robust NLP system for predicting which choice fits best in a given discourse. Extending the context-induced hypothesis (Kövecses, 2009) to metaphorical vs. literal usage, contextual salience would expect a metaphorical discourse preceding a metaphorical expression, and a literal context preceding a literal expression.

The current paper addresses the above questions by collecting and analyzing judgements on 1,000 instances of 50 pairs of English synonymous literal vs. metaphorical verb-object and subject-verb expressions in corpus-extracted discourses. In Task 1, asking participants to rate the degree of metaphoricality of the discourses sheds light on whether the preceding context plays a role in the choice of metaphorical vs. literal expressions within that discourse. Task 2, in which annotators provide a binary decision that favors one usage over the other, provides insight on the metaphorical vs. literal usage in context. To our knowledge, our work constitutes the first empirical study on conscious discourse-embedded choices about synonymous pairs of metaphoric vs. literal expressions. Our novel dataset constitutes a solid starting point for computational research on salient discourse conditions for contextual metaphorical vs. literal usage.

2 Related Work

Theoretical Background Different metaphor theories were broadly discussed in linguistics and philosophy, first as an attempt to understand what metaphors are. In parallel, researchers looked at what drives people to use metaphors (Glucksberg, 1989; Kövecses, 2010; Ortony, 1975) as well as “what metaphor actually does” (Hampe, 2017). The cognitive linguistics view of metaphors in Conceptual Metaphor Theory (Lakoff and Johnson, 1980) describes how metaphors are frequently used everyday and by everybody, and moreover in an unintentional way.

A corpus study by Stefanowitsch (2006) provides evidence for this view. He shows that metaphors are used not only as a stylistic choice but also as a cognitive function, since people seem to use them to explicate things or reasonings. However, further studies show clear signs that metaphors can be of stylistic choice, e.g. metaphorical language has a stronger emotional impact than literal language (Mohammad et al., 2016; Köper and Schulte im Walde, 2018). This statement gives support to the idea that there exists a difference in choice between metaphorical versus literal expressions.

In their psycholinguistics study, Thibodeau and Boroditsky (2011) also illustrate that using metaphors over their literal alternatives may influence the way humans conceptualize an act. While on the one hand it seems like people do feel a difference when using one version or the other, on the other hand it also seems that it affects the way people react. It is therefore necessary to find a way for a computational system to capture this difference.

Existing Resources Stefanowitsch (2006) provides a corpus-based study using carefully collected and curated data. He explores whether the use of metaphors is a stylistic choice or a cognitive function, and relies on sentences where both the metaphorical expression and a literal alternative may be used (e.g. *in the heart of* versus *in the center of*). His examples are close to what we aim for in our dataset, but his study is based solely on a handful of metaphorical expressions.

The NLP tasks of figurative language detection and interpretation have led to the creation of several datasets. Mohammad et al. (2016) composed 171 sentences where a verb is used metaphorically, e.g. *abuse* in “Her husband often abuses alcohol”. For

each sentence, the authors of the paper chose a literal synonym of the target verb, such as *drink* in the above example sentence.

Shutova (2010) aimed for metaphor interpretation and collected sentences containing metaphorical verbs from the British National Corpus, e.g. *grasp* in “Anyone who has introduced speech act theory to students will know that these technical terms [...] are not at all easy to grasp.” She asked annotators to provide an alternative verb with a literal meaning. The dataset consists of a list of metaphorically-used verb-object and subject-verb expressions, with one or more literal verb alternatives. For the verb-object expression *grasp term*, the verb *grasp* was given the literal alternatives *understand* and *comprehend*, for example.

Similarly, the model developed by Bizzoni and Lappin (2018) automatically ranks the best four paraphrases for each metaphorical sentence. The final dataset consists of 200 metaphorical sentences, each with their four automatically generated and ranked paraphrases.

The setup of the latter three datasets is what we were looking for; however, all present only a one-sentence context, which in our opinion is not sufficient when addressing the importance of preceding discourse. Moreover, the dataset from Bizzoni and Lappin (2018) automatically generated the literal alternatives, so they would require additional careful human judgements. Even though we were inspired from all these useful resources, we have not been able to find an existing dataset that can be fully used for our goals.

3 Experimental Setup

3.1 Compiling pairs of expressions

We collect 50 pairs of expressions from Shutova (2010) and Mohammad et al. (2016), 36 of which are verb-object (VO) expressions, and 14 of which are subject-verb (SV) expressions. Our corpus thus consists of 50 expressions where the verb is used metaphorically, and 50 expressions where the verb is a synonymous literal alternative, such as *tackle/address question* for a VO expression and *tension mount/increase* for a SV expression. As the original datasets were created for different purposes, we perform slight changes in some cases. For instance, we exchange *catch contagion* by the more common version *catch disease*. We provide an overview of all pairs in the Appendix B.

-
- (a) It is true indeed that not a sparrow drops unnoticed by the Mind of THE ALL - that even the hairs on our head are numbered, as the scriptures have said. There is nothing outside of Law; nothing that happens contrary to it. And yet, do not make the mistake of supposing that man is but a blind automaton - far from that. The Hermetic Teachings are that man may use Law to overcome laws, and that the higher will always prevail against the lower, until at last he has reached the stage in which he seeks refuge in the LAW itself, and laughs the phenomenal laws to scorn. Are you able to **grasp** the inner **meaning** of this?
-
- (b) This wasn't just a play on words, rather it was a demand that they should 'maintain a consistency between their words and their actions'. But I agree, that still does not absolve them from the need to speak truth to power. In our times when people spend so much time with TV and the internet, do they have the interest and time to read poetry? Many people believe that it is difficult to read poetry. Can everyone **understand** the **meaning** of a good poem, or is a skill necessary?
-

Table 1: Examples of discourses for the synonymous pair *grasp/understand meaning*. The metaphorical expression *grasp meaning* is used in (a), its literal paraphrase *understand meaning* is used in (b), and both are applicable in both contexts.

3.2 Collecting discourses

We automatically extract all sentences from the ukWaC (Baroni et al., 2009) containing inflected forms of our compiled expressions, with a maximum of 25 characters in between the verb and its argument in VO/SV. We select 20 instances for each pair of expressions, with 10 instances each for the metaphorical/literal versions. Our dataset thus contains a total of 1,000 discourses. As we are testing the extent to which context plays a role in favoring one expression over the other, we extract four to five sentences preceding the sentence containing the target expression, followed by the actual sentence with the metaphorical/literal expression. The discourses contain 31–216 words, with an average of 98 words. Table 1 shows examples of discourses for a pair of expressions.

3.3 Crowdsourcing experiments

As we are interested in (i) the influence of context in the choice of a target expression and (ii) human preferences for metaphor vs. literal expressions, the annotation process is directed in two tasks.

Task 1: The first task tests for the **degree of metaphoricity vs. literalness** of the expression-preceding discourse, in order to answer the question "Does the discourse influence the choice of a metaphorical/literal expression?" To obtain a minimum of 10 ratings per instance, we present the discourses up to the word preceding the target expression to 15 workers on Amazon Mechanical Turk (AMT)¹ and ask them to indicate on a scale from 1 to 6 where they judge the overall discourse on the range between mostly literal–mostly metaphorical.

Task 2: The second task tests **which expression (metaphorical vs. literal) is favored in a given discourse**, in order to answer the question "Does one favor the use of a metaphorical vs. a literal expression given a specific (metaphorical/literal) preceding discourse?" As in Task 1, we show the

discourses to 15 AMT workers, however now including the target sentence but with a blanked spot for the target expression, and ask them to choose which expression fits better (binary choice).

For both tasks, we limit the location of the workers to English-speaking countries, and specify in the instructions that the tasks are only for English native speakers. A total of 183 workers annotated the 1,000 Task-1 instances, on average providing ratings for 81 instances. We disregard 73 workers who completed less than 20 instances, which results in a total of 14,514 judgements by 110 workers on rating the degree of metaphoricity of the discourses. Each instance is rated by at least 11 workers. For Task 2, 238 workers completed 63 instances on average. Similarly to Task 1, we only keep the 136 workers who completed at least 20 instances, which results in a final dataset with 14,378 judgements. Appendix A provides a detailed explanation and examples of the setup of the AMT experiments.

4 Results and data analyses

Task 1: Table 2 shows the workers' ratings on the degree of metaphoricity of the expression-preceding discourses, across all 14,514 judgements, next to the resulting medians for our 1,000 instances. We can see a clear preference of the workers for the middle rather than the extreme categories (1 for mostly literal and 6 for mostly metaphorical), with a slight preference for metaphoricity (also see top part in Table 3, using 3.5 as threshold for literal vs. metaphorical categorisation). For the medians this results in a strong focus on the range 3–4.

Scale	#Ratings	#Median
1	1,905	3
2	2,147	35
3	2,689	340
4	3,496	536
5	3,101	86
6	1,176	0
Total	14,514	1,000

Table 2: Number of ratings and medians across scale.

¹<https://www.mturk.com/>

metaphoricity of discourse (annotated)	
metaphorical	622 (62.2%)
literal	378 (37.8%)
metaphoricity of expression (annotated)	
metaphorical	425 (45.6%)
literal	506 (54.4%)
metaphoricity of discourse (annotated–original)	
metaphorical – metaphorical	315 (31.5%)
literal – literal	193 (19.3%)
metaphorical – literal	307 (30.7%)
literal – metaphorical	185 (18.5%)
metaphoricity of discourse–metaphoricity of expression	
metaphorical – metaphorical	260 (27.9%)
literal – literal	187 (20.1%)
metaphorical – literal	319 (34.3%)
literal – metaphorical	165 (17.7%)

Table 3: Summary and comparison of annotations for Tasks 1 and 2 and the original corpus-based discourses+expressions. *Top part*: proportions of annotated metaphorical vs. literal discourses + proportions of annotated metaphorical vs. literal expressions. *Middle part*: metaphoricity of discourses in comparison to metaphoricity of original corpus expression. *Bottom part*: metaphoricity of annotated discourses in relation to metaphoricity of annotated expressions. Threshold for literal/metaphorical categories: median of 3.5.

According to the contextual salience hypothesis, we expect that metaphorically-rated discourses are more likely to be followed by a metaphorical expression, and ditto for literal discourses and expressions. In the middle part of Table 3 we looked at the expressions that were originally collected from the corpus, and compared their categorizations to the discourse ratings. Judging from the discourses where a metaphorical expression was used (*–metaphorical), one may induce that the context-salient hypothesis is valid, since raters mostly judged the respective discourses as metaphorical: 31.5% vs. 18.5%. However, for discourses where a literal expression was used (*–literal), the context-salient hypothesis fails, as raters favored a metaphorical context preceding a literal usage: 30.7% vs. 19.3%.

Task 2: Across 931 binary judgements with an absolute majority (67 discourses are a tie), we note a slight preference for literal expressions (54.4% vs. 45.6%). Only when looking at the individual 50 pairs (see Appendix) we find a more diverse picture. Cases where the literal usages were preferred (e.g., *devour/read book*, *suck/attract worker*, *attack/solve problem*) may be explained by the rather strong emotional effect of the metaphorical expressions, cf. [Mohammad et al. \(2016\)](#), if they are not coherent with the context. A preference for the metaphorical expressions, as in *breathe/instill life*, *painting capture/represent*, *fate lie/be*, may be explained by the high conventionality of these metaphors.

Combining Tasks 1 and 2: As demonstrated above, the metaphorically- vs. literally-rated contexts were not necessarily in accordance with the original choice of expression in the corpus data. In the bottom part of Table 3 we bring Task-1 and Task-2 results together and relate the binary metaphoricity judgements of the target expressions to the judgements of the respective preceding discourses. As before, we observe tendencies against the context-salient hypothesis: while on the one hand a metaphorically-rated discourse seems to prime for the use of a metaphorical expression (27.9%), it similarly primes for the use of a literal expression (34.3%); on the other hand, only 20.1% of the literal expressions are preceded by literally-judged discourses. Figure 1 shows the average proportions of metaphorical vs. literal expressions across the median ratings for our 1,000 instances. While we observe a slight downward trend for choosing literal expressions – and, in parallel, a slight upward trend for choosing metaphorical phrases – with increasing medians (i.e., when the discourse is rated more metaphorical), we also note that literal expressions are favored over metaphorical ones across all medians (i.e., irrespective of the metaphoricity of the discourse.)

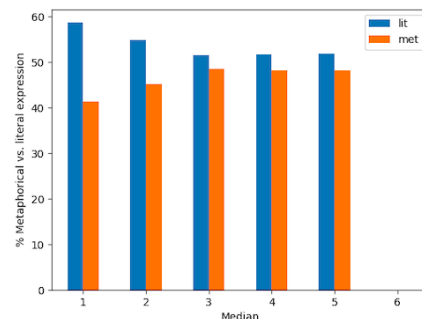


Figure 1: Proportions of metaphorical vs. literal expressions across median ratings for all 1,000 discourse instances.

5 Conclusion

This work offers a new approach and dataset to study metaphor vs. literal language usage in relation to discourse embedding. Our collection counters the theoretical context-salient hypothesis that metaphorical vs. literal usage is expected to be primed by metaphorical vs. literal preceding contexts, respectively. Even more so, it provides a valuable starting point for computational explorations on further discourse conditions for metaphorical vs. literal choices, such as lexical semantic relatedness ([Birke and Sarkar, 2006](#); [Sporleder and Li, 2009](#); [Do Dinh, 2013](#)) and contextual abstractness ([Turney et al., 2011](#); [Tsvetkov et al., 2014](#)).

References

- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43(3):209–226.
- Julia Birke and Anoop Sarkar. 2006. A Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language. In *Proceedings of the 11th Conference of the European Chapter of the ACL*, pages 329–336, Trento, Italy.
- 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.
- Erik-Lan Do Dinh. 2013. Automatic Identification of Novel Metaphoric Expressions. Master’s thesis, Technische Universität Darmstadt.
- Matt Gedigian, John Bryant, Srinu Narayanan, and Branimir Cicic. 2006. Catching Metaphors. In *Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, pages 41–48, New York City, NY.
- Sam Glucksberg. 1989. [Metaphors in conversation: How are they understood? Why are they used?](#) *Metaphor and Symbolic Activity*, 4(3):125–143.
- Beate Hampe, editor. 2017. *Metaphor: Embodied Cognition and Discourse*. Cambridge University Press, Cambridge, UK.
- Maximilian Köper and Sabine Schulte im Walde. 2018. Analogies in Complex Verb Meaning Shifts: The Effect of Affect in Semantic Similarity Models. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 150–156, New Orleans, LA, USA.
- Zoltan Kövecses. 2009. The Effect of Context on the Use of Metaphor in Discourse. *Iberica*, 17:11–24.
- 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.
- Saif M. Mohammad, Ekaterina Shutova, and Peter D. Turney. 2016. Metaphor as a Medium for Emotion: An Empirical Study. In *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics*, pages 23–33, Berlin, Germany.
- Andrew Ortony. 1975. [Why metaphors are necessary and not just nice](#). *Educational Theory*, 25(1):45–53.
- Ellie Pavlick, Johan Bos, Malvina Nissim, Charley Beller, Benjamin Van Durme, and Chris Callison-Burch. 2015. Adding Semantics to Data-Driven Paraphrasing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1512–1522, Beijing, China.
- Christina Schäffner. 2004. Metaphor and Translation: Some Implications of a Cognitive Approach. *Journal of Pragmatics*, 36:1253–1269.
- Ekaterina Shutova. 2010. Automatic Metaphor Interpretation as a Paraphrasing Task. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the ACL*, pages 1029–1037, Los Angeles, CA, USA.
- Ekaterina Shutova and Simone Teufel. 2010. [Metaphor corpus annotated for source - target domain mappings](#). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Valletta, Malta. European Language Resources Association.
- Caroline Sporleder and Linlin Li. 2009. Unsupervised Recognition of Literal and Non-Literal Use of Idiomatic Expressions. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, pages 754–762, Athens, Greece.
- Anatol Stefanowitsch. 2006. Words and their Metaphors: A Corpus-based Approach. In Anatol Stefanowitsch and Stefan Th. Gries, editors, *Corpus-Based Approaches to Metaphor and Metonymy*, pages 17–35. de Gruyter, Berlin.
- Paul H. Thibodeau and Lera Boroditsky. 2011. Metaphors We Think With: The Role of Metaphor in Reasoning. *PLoS ONE*, 6(2).
- Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014. Metaphor Detection with Cross-Lingual Model Transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 248–258, Baltimore, MD, USA.
- Peter D. Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 680–690, Edinburgh, UK.
- Raymond van den Broek. 1981. The Limits of Translatability Exemplified by Metaphor Translation. *Poetics Today*, 1(4):73–87.

Appendices

A Setup of crowdsourcing experiments

In this section, we provide a detailed explanation of the AMT tasks setup.

We randomly shuffled the 1,000 discourses composing our dataset, and created 50 batches of 20 instances. AMT workers were asked to complete all instances of a batch, and were allowed to complete as many batches as they wished. We discarded workers who completed less than 20 instances.

We provide below an example of what the workers were shown when completing each task.

A.1 Task 1

Workers were asked to rate the degree of metaphoricality they overall perceived, when reading the discourse preceding the target expression. On purpose, we did not give them a definition of what a *metaphor* or a *metaphorical discourse* is, in order to not bias them. Previous cognitive research has shown that metaphors and metaphorical concepts are used without even being aware of using them (Lakoff and Johnson, 1980), and this is what this study has attempted to look at. Below is an example of a discourse that AMT workers were asked to complete:

How metaphorical / literal is this discourse according to you? Please read the following text carefully and rate the degree of literalness or metaphoricality of the discourse from 1 to 6, where 1 means that the discourse is mostly literal and 6 means that the discourse is mostly metaphorical.

The fact that there's a lunar eclipse that day heightens that need. Indeed, it could be that your focus in the next fortnight (until the solar eclipse on 22nd) will be very strong indeed. True, this could be triggered by unexpected events on Monday which underline the need for change. And true, you've had several sidewinders thrust your way in recent years - and these haven't left you racing for more education. Now though, you may want to [...]

mostly literal 1 2 3 4 5 6 mostly metaphorical

A.2 Task 2

Workers were asked to choose which expression they believe fits best in the given discourse.

Please fill in the blank. Pick the option that fits best in this discourse according to you:

Why is Saddam Hussein pushing ahead with weapons of mass destruction if at some point he is not going to use them? It's certainly got to be a factor in all of this. Unlike anthrax, the bacteria used in last year's unsolved mail attacks, the highly contagious smallpox virus can be passed from person to person. The virus causes ugly pustules to form both on the skin and inside the mouth and throat. About a third of unvaccinated people who _____ would die.

*Which expression fits best in the blank?
1) caught the disease 2) got the disease*

A.3 Inter-annotator agreement and standard deviation

As we obtained over 14,000 judgements from 110 workers (Task 1), we calculated IAA in the same way done by Pavlick et al. (2015), i.e., computing IAA as the average of each rater's correlation with the average of all other workers. We reach Spearman's $\rho = 0.26$, which might seem to be rather low but this is IAA for 110 workers, rather than 7 (as in Pavlick et al. (2015)). Figure 2 gives an idea of annotation reliability as it represents the dispersion from the individual data values to the mean score of each instance. Overall, it seems that raters tend to agree on extreme cases, i.e., agreement is high on rating mostly literal (lower left corner) and mostly metaphorical (upper right corner) discourses. Agreement varies across the in-between average values but stays rather reasonable.

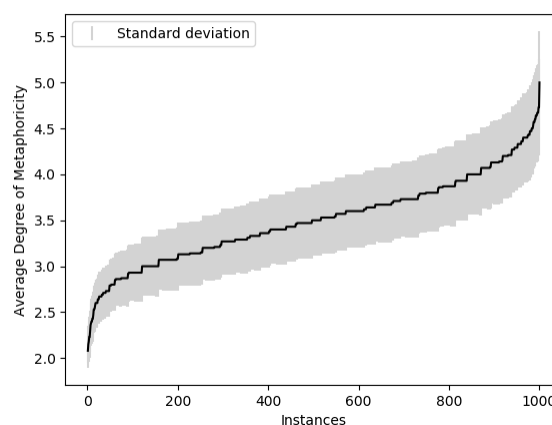


Figure 2: Sorted average values (in black) from all workers across the 1,000 instances. The grey cloud represents the standard deviation values for each average value.

B Summary Tasks 1 and 2

Metaphorical (met)/Literal expressions (lit)	#metexp (%)	(a)#met context metexp	(b)#lit context metexp	#litexp (%)	(c)#met context litexp	(d)#lit context litexp
subject-verb pairs (SV):						
example illustrate/show	173 (59.45%)	78	66	118 (40.55%)	72	73
fire devour/destroy	127 (44.10%)	90	57	161 (55.90%)	71	72
factor shape/affect	124 (43.66%)	66	84	160 (58.51%)	74	72
painting capture/represent	188 (65.73%)	86	62	98 (34.27%)	78	69
tension mount/increase	146 (51.41%)	83	66	138 (48.59%)	78	70
mess reflect/show	156 (53.98%)	80	67	133 (46.02%)	82	65
moon peep/appear	111 (38.14%)	105	41	180 (61.86%)	86	57
fate lie/be	201 (69.07%)	83	64	90 (30.93%)	75	70
colour harmonise/match	132 (47.31%)	86	59	147 (52.69%)	74	72
story grab/intrigue	125 (44.80%)	91	53	154 (55.20%)	86	58
distinction blur/disappear	143 (50.00%)	72	75	143 (50.00%)	76	67
view reflect/represent	151 (52.80%)	76	68	135 (47.20%)	68	76
result emerge/appear	170 (59.44%)	59	87	116 (40.56%)	75	70
war uproot/displace	113 (38.97%)	78	67	177 (61.03%)	72	72
verb-object pairs (VO):						
mount/organise production	113 (38.57%)	68	78	180 (61.43%)	74	72
recapture/recall feeling	138 (48.08%)	81	64	149 (51.92%)	89	54
grasp/understand meaning	129 (44.48%)	88	55	161 (55.52%)	84	62
drown/forget trouble	147 (50.17%)	96	45	146 (49.83%)	90	56
catch/get disease	162 (55.67%)	83	65	129 (44.33%)	71	74
breathe/instill life	193 (66.55%)	78	67	97 (33.45%)	78	67
flood/saturate market	146 (51.05%)	70	71	140 (48.95%)	74	67
push/urge someone	107 (37.54%)	74	73	178 (62.46%)	73	69
stir/cause excitement	169 (59.09%)	76	68	117 (40.91%)	69	70
cast/cause doubt	191 (65.41%)	66	82	101 (34.59%)	81	67
leak/disclose report	128 (44.60%)	76	73	159 (55.40%)	70	76
devour/read book	98 (34.15%)	81	67	189 (65.85%)	89	56
suck/attract worker	83 (29.23%)	71	76	201 (70.77%)	78	70
dull/decrease appetite	132 (45.52%)	77	69	158 (54.48%)	66	78
frame/pose question	119 (41.18%)	61	80	170 (58.82%)	79	67
abuse/drink alcohol	152 (53.90%)	77	63	130 (46.10%)	75	70
juggle/manage job	158 (55.05%)	79	65	129 (44.95%)	67	77
attack/solve problem	95 (32.31%)	76	71	199 (67.69%)	75	72
disown/reject past	164 (56.94%)	89	53	124 (43.06%)	83	62
pour/invest money	119 (42.20%)	78	70	163 (57.80%)	76	68
follow/practise profession	110 (38.46%)	69	76	176 (61.54%)	63	83
taste/experience freedom	103 (35.27%)	93	54	189 (64.73%)	83	62
break/end agreement	161 (55.71%)	74	69	128 (44.29%)	72	71
sow/instill doubt	138 (47.59%)	73	69	152 (52.41%)	83	64
twist/misinterpret word	151 (53.17%)	92	55	133 (46.83%)	83	60
boost/improve economy	145 (50.88%)	67	81	140 (49.12%)	77	68
throw/make remark	108 (38.03%)	83	63	176 (61.97%)	71	72
tackle/address question	113 (38.18%)	85	62	183 (61.82%)	79	67
buy/believe story	141 (48.62%)	83	62	149 (51.38%)	84	61
fuel/stimulate debate	169 (58.28%)	68	77	121 (41.72%)	76	67
float/discuss idea	120 (41.24%)	67	80	171 (58.76%)	62	78
wear/have smile	175 (61.62%)	89	58	109 (38.38%)	90	55
poison/corrupt mind	142 (48.97%)	71	73	148 (51.03%)	78	67
shape/determine result	118 (41.40%)	82	64	167 (58.60%)	92	54
colour/affect judgement	114 (39.72%)	79	68	173 (60.28%)	71	73

Table 4: Summary of results when combining Tasks 1 and 2. For each pair of metaphorical/literal expression, in order: number of times the metaphorical expression was chosen (%); number of times the preceding context of the metaphorical expression was rated as (a) metaphorical vs. (b) literal; number of times the literal expression was chosen (%); number of times the preceding context of the literal expression was rated as (c) metaphorical vs. (d) literal.