

# Paraphrasing Compound Nominalizations

**John S. Y. Lee, Ho Hung Lim**

Department of Linguistics and  
Translation

City University of Hong Kong  
Hong Kong SAR, China

jsylee@cityu.edu.hk

hhlim3@cityu.edu.hk

**Carol Webster**

College of Professional and  
Continuing Education

The Hong Kong Polytechnic University  
Hong Kong SAR, China

carol.webster@cpce-polyu.edu.hk

## Abstract

A nominalization uses a deverbal noun to describe an event associated with its underlying verb. Commonly found in academic and formal texts, nominalizations can be difficult to interpret because of ambiguous semantic relations between the deverbal noun and its arguments. Our goal is to interpret nominalizations by generating clausal paraphrases. We address compound nominalizations with both nominal and adjectival modifiers, as well as prepositional phrases. In evaluations on a number of unsupervised methods, we obtained the strongest performance by using a pre-trained contextualized language model to re-rank paraphrase candidates identified by a textual entailment model.

## 1 Introduction

Nominalizations are widely used in academic and other genres of formal texts that adopt a compact and abstract writing style. A nominalization is a noun (e.g., “treatment”) that is morphologically derived from a verb (“treat”), and that designates some aspects of the event referred to by the verb (Quirk et al., 1985). Because of the systematic correspondence between nominalization and clause structure, a noun phrase headed by a deverbal noun (e.g., “surgical *treatment* of fracture”) often has a clausal paraphrase headed by the underlying verb (“the surgery *treats* fracture”) (Cohen et al., 2008).

This paper aims to generate clausal paraphrases for nominalizations. A key requirement in this task is to map the arguments of the deverbal noun to those of the underlying verb. In the example above, the deverbal noun “treatment” has a prepositional phrase (PP) argument (“of fracture”) that corresponds to the object of the verb “treat”, while the prenominal modifier (“surgical”) corresponds to the subject. This mapping varies according to the semantic context of the nominalization. An alternative nominalization, such as “fracture treatment

with surgery”, may have its PP argument (“with surgery”) mapped to the subject, and its prenominal modifier (“fracture”) mapped to the object. Table 1 shows a range of possible mappings.

Our task is distinguished from previous research in terms of both the input and output. In the field of nominalization disambiguation, most studies have focused on deverbal nouns with one argument, typically a nominal modifier (Lapata, 2002; Nicholson and Baldwin, 2008). We extend the scope by addressing both nominal and adjectival modifiers, as well as PP argument.

In terms of the output, most previous work assigned semantic role labels to arguments of the deverbal noun; in contrast, we generate paraphrases in predicate-argument form (Table 1). This approach has the advantage of avoiding commitment to a particular semantic theory, and can facilitate direct application to various NLP downstream tasks. The paraphrases may improve performance in machine translation, for example, for sentences that do not use nominalization in the target language. In information extraction, they can potentially expand query formulation to increase recall. In text simplification, the paraphrase may be easier to understand than the original nominalization for less proficient readers.

The rest of the paper is organized as follows. We define our task in the next section, and then review previous work in Section 3. After a description of the dataset in Section 4, we present our approach in Section 5 and discuss experimental results in Section 6.

## 2 Task Definition

As shown in Table 1, the input is a nominalization with two arguments. Specifically, the deverbal noun ( $N_V$ ) is modified by the preceding noun ( $M_n$ ) or adjective ( $M_a$ ), and by a prepositional phrase with one prepositional object ( $O$ ). Although our dataset does not include nominalizations with mul-

Type	Input: nominalization				Output: clausal paraphrase
	$M_a$ or $M_n$	$N_V$	$p$	$O$	
MVO	American police	influence response	on to	global culture the rioting	America <sup>M</sup> influences <sup>V</sup> global culture <sup>O</sup> the police <sup>M</sup> responds <sup>V</sup> to the rioting <sup>O</sup>
OVM	climatic war	effects preparations	of of	air pollutants the government	air pollutants <sup>O</sup> affect <sup>V</sup> the climate <sup>M</sup> the government <sup>O</sup> prepares <sup>V</sup> for war <sup>M</sup>
VMO	bodily student	injury admissions	to into	a friend universities	injure <sup>V</sup> the body <sup>M</sup> of a friend <sup>O</sup> admit <sup>V</sup> students <sup>M</sup> into universities <sup>O</sup>
VOM	presidential business	nomination travel	of to	Harrison Greece	nominate <sup>V</sup> Harrison <sup>O</sup> as president <sup>M</sup> travel <sup>V</sup> to Greece <sup>O</sup> for business <sup>M</sup>
MOV	naval majority	assistance decision	from of	Italian powers the panel	Navies <sup>M</sup> of Italian powers <sup>O</sup> assist <sup>V</sup> a majority <sup>M</sup> of the panel <sup>O</sup> decides <sup>V</sup>

Table 1: The system input is a nominalized verb ( $N_V$ ) that is modified by either a preceding noun ( $M_n$ ) or adjective ( $M_a$ ), and by a prepositional object ( $O$ ); the system output is a clausal paraphrase involving  $V$  (the verb underlying  $N_V$ ),  $M$  (the noun corresponding to the prenominal modifier), and  $O$ .

tiple prenominal and PP modifiers<sup>1</sup>, and other modifier types such as s-genitives and possessive pronouns<sup>2</sup>, our proposed algorithm can be extended in a straightforward manner to address them.

The output is a clausal paraphrase, i.e., a predicate headed by a verb that corresponds to the deverbal noun  $N_V$ . This verb may have a subject, object, or prepositional phrase corresponding to  $M_a$ ,  $M_n$  or  $O$ . Our model assumes that every input has a paraphrase. In actual deployment, the system would need to determine whether an input can be paraphrased, a task that we leave to future work.

### 3 Previous work

Our research is most closely related to noun compound interpretation (Section 3.1) and nominal semantic role labeling (Section 3.2).

#### 3.1 Noun compound interpretation

Research on noun compound interpretation has taken two main approaches. One approach assigns an abstract label to describe the relation between the head noun and the noun modifier (Tratz and Hovy, 2010). Another, similar to ours, generates a paraphrase that links the two nouns with prepositions and verbs (Butnariu et al., 2010; Nakov and Hearst, 2013; Ponkiya et al., 2020), or in a free-form paraphrase (Hendrickx et al., 2013). As unsupervised methods has recently been found to perform well in NCI (Ponkiya et al., 2020), we

<sup>1</sup>e.g., “regional electricity consumption” or “consumption of electricity in the region”

<sup>2</sup>e.g., “the region’s electricity consumption” or “its electricity consumption”

likewise pursue this direction.

Our work can be viewed as a special case in NCI, where the head noun is required to be a deverbal noun. Given the properties of the deverbal noun, the paraphrase in this paper takes a different form, making use of the underlying verb to form a clause.

#### 3.2 Nominal semantic role labeling

Previous work on nominalization interpretation has mostly focused on nominal semantic role labeling (SRL), which assigns abstract labels (e.g., agent, patient) to arguments of nominalizations (Lapata, 2002; Padó et al., 2008; Kilicoglu et al., 2010). SRL can be performed with a classifier, trained on features such as syntactic structure and corpus frequencies of verb-argument pairs (Lapata, 2002; Pradhan et al., 2004). Compared to SRL, paraphrasing a nominalization requires the generation of the verb and its arguments. Some downstream NLP tasks, such as semantic parsing into meaning representations (Samuel and Straka, 2020), can directly make use of SRL output. To others, such as machine translation and text simplification, a paraphrase of the nominalization can potentially be more immediately applicable.

Previous studies have only addressed nominalizations with one argument, typically a nominal modifier (e.g., “fracture treatment”). However, adjective modifiers (“surgical treatment”) are also prevalent and have been analyzed by linguists along with nominal modifiers under the term “complex nominal” (Levy, 1978). While there has been previous work on predicting the attribute of an adjective (Hartung et al., 2017) and the use of pertainyms

Type	$M_a$	$M_n$
MVO	<b>92</b>	44
VOM	75	<b>57</b>
OVM	50	35
VMO	57	36
MOV	1	2

Table 2: Distribution of paraphrase types (see examples in Table 1) in our dataset (Section 4), which includes nominalizations with nominal modifiers ( $M_n$ ) and with adjectival modifiers ( $M_a$ ). The majority baseline uses MVO for  $M_a$  inputs and VOM for  $M_n$  inputs.

in question answering (Greenwood, 2004), neither has been evaluated on paraphrase generation in the general domain.

## 4 Dataset

We collected 369 sentences from English Wikipedia that match the input pattern defined in Section 2, i.e., a deverbal noun ( $N_V$ ) modified by the preceding noun ( $M_n$ ) or adjective ( $M_a$ ), and by a prepositional phrase (PP). The  $M_a$  must have a pertainym in WordNet;  $N_V$  in the Academic Word List (Coxhead, 2016) with high frequency are given priority, to reflect the widespread use of nominalizations in academic text.

Two annotators, a native speaker and a near-native speaker of English, composed a paraphrase for each nominalization. A professor of linguistics who is a native speaker of English reviewed the paraphrases, either keeping both or selecting one of them. As shown in Table 2, the final dataset contains a total of 449 paraphrases.<sup>3</sup>

The annotators were instructed to use the simple present active for  $V$ , since tense and aspect cannot always be inferred from the rest of the sentence. They were also asked, as far as possible, to paraphrase  $M_a$  with an etymologically related word, and to preserve the lemma of  $M_n$  and  $O$ . If necessary, an additional word may be inserted for clarity or fluency of the paraphrase.<sup>4</sup>

## 5 Approach

The system takes as input the original sentence and the strings  $N_V$ ,  $p$ ,  $O$ , and  $M_a$  or  $M_n$ , (Table 1). These strings were provided, rather than automatically extracted, so that errors in automatic parsing would not confound experimental results.

<sup>3</sup>Accessible at <https://github.com/NominalizationParaphrase>

<sup>4</sup>E.g., paraphrasing “nuclear attack on Japan” as “attack Japan with nuclear weapons”.

### 5.1 Paraphrase generation

We first generate all possible word choices for the verb and its arguments that will appear in candidate paraphrases:

**Verb ( $V$ )** Candidates include all verbs that are derivationally related to  $N_V$  in the WordNet database. If there is none, we use all verbs in the entry of  $N_V$  in CatVar (Habash and Dorr, 2003).

**Arguments ( $M$ ,  $O$ )** The argument  $M$  is constructed from either  $M_n$  or  $M_a$ . For the former, candidates include the singular and plural forms of  $M_n$ . For the latter, candidates include the singular and plural forms of its pertainyms in WordNet. The argument  $O$  remains unchanged from the input.

We place all permutations of candidates for  $V$ ,  $M$ , and  $O$  in the word orders of all paraphrase types shown in Table 1, as well as their passive voice equivalents.<sup>5</sup> The arguments  $M$  and  $O$  can be preceded with a determiner, and also with a preposition except when at the sentence-initial position. We used the T5 model (Raffel et al., 2020) to generate up to two words — a preposition followed by a determiner; a preposition alone; a determiner alone; or the empty string.

### 5.2 Paraphrase selection

We then select the best candidate paraphrase with the following approaches:

**Sentence similarity** selects the candidate paraphrase whose sentence embedding is most similar to that of the original nominalization, according to cosine similarity. We obtained sentence embeddings with the pre-trained stsb-roberta-large model<sup>6</sup> from Sentence-BERT (Reimers and Gurevych, 2019).

**Language model (LM)** selects the candidate paraphrase that yields the highest language model score. We evaluated the log-probability score based on GPT-2 (117M), and the pseudo-log-likelihood score based on DistilBERT (Salazar et al., 2020).<sup>7</sup>

<sup>5</sup>The passive paraphrases are “ $O$   $V_{prt}$  by  $M$ ” (for MVO), “ $M$   $V_{prt}$  by  $O$ ” (for OVM), “ $O$   $V_{prt}$   $p$   $M$ ” (for VOM) and “ $M$   $V_{prt}$   $p$   $O$ ” (for VMO), where  $V_{prt}$  represents ⟨be⟩ followed by the past participle of  $V$ , and  $p$  represents any preposition except “by”.

<sup>6</sup><https://huggingface.co/sentence-transformers/stsb-roberta-large>

<sup>7</sup>Both from <https://github.com/awsml/mlm-scoring>

**Majority + LM** selects the candidate paraphrase from the majority paraphrase type — i.e., MVO for an  $M_a$  input and VOM for an  $M_n$  input (Table 2) — that yields the highest LM score.

**Semantic parser + LM** first obtains the Abstract Meaning Representation of the input sentence with PERIN (Samuel and Straka, 2020) and predicts the paraphrase type from the arg0 and arg1 values of the node aligned to the deverbal noun  $N_V$ . It predicts MVO if arg0 is aligned to  $M_a$  or  $M_n$ , and OVM if arg0 is aligned to  $O$ . Otherwise, it predicts VMO if arg1 is  $M_a$  or  $M_n$ , and VOM if arg1 is  $O$ . It then selects the candidate paraphrase from the predicted type that yields the highest LM score.

**Textual Entailment** takes the original sentence as the premise, and the candidate paraphrase as the hypothesis, and predicts whether the facts in the former imply those in the latter. We used the textual entailment model from AllenNLP that is fine-tuned on SNLI with RoBERTa (Liu et al., 2019).<sup>8</sup> The candidate paraphrase with the highest score is selected.

**Textual Entailment + LM** selects the candidate paraphrase with the highest language model score among the three highest-scoring candidate paraphrases identified with the Textual Entailment model.

## 6 Experiment

### 6.1 Evaluation metrics

In a clausal paraphrase, the determiner and number of the noun ( $M$ ) that corresponds to the prenominal modifier ( $M_a$  or  $M_n$ ) can be ambiguous and open to different interpretations. In our evaluation, we removed all determiners and lemmatized all words in both the gold and the predicted paraphrase, and then compared them on two metrics:

**Paraphrase accuracy** The system is considered correct if the lemmatized form of the predicted paraphrase and gold paraphrase are identical.

**Word order accuracy** Same as above, except that prepositions are not taken into consideration. This metric essentially measures the system’s ability to predict the verb and arguments and to put them into the correct word order.

Further, we consider two experimental settings:

**Gold arguments** The system has access to the gold  $V$ ,  $M$  and  $O$ . In other words, it needs only to determine the paraphrase type and prepositions.

**Automatic** The system automatically generates  $V$ ,  $M$  and  $O$  from the input.

### 6.2 Results

**Paraphrase accuracy.** As shown in Table 3, the **Textual Entailment + LM** model, in conjunction with GPT-2, outperformed all other models with respect to this metric in both settings. When given the gold arguments, it achieved 51.22% accuracy, a statistically significant improvement over using **Textual Entailment** alone (38.21%) and **DistilBERT** alone (43.36%).<sup>9</sup> As expected, its performance degraded in the fully automatic setting (38.48%), but it continued to outperform all other models with statistical significance, including **Textual Entailment** and **DistilBERT**.<sup>10</sup>

These results suggest that the language model and textual entailment model can complement each other. While the former optimizes the likelihood of a candidate paraphrase on its own, the former also takes into account word choices in the premise (i.e., the nominalization), including the preposition<sup>11</sup>, when estimating semantic equivalence.

**Word order accuracy.** The **Textual Entailment + LM** model gave the strongest performance in both settings, but with DistilBERT rather than with GPT-2 on this metric. In the “gold arguments” setting, its accuracy is highest at 65.58%, followed by **DistilBERT** (62.87%) and **Textual Entailment** (60.70%), although the improvement is not statistically significant.<sup>12</sup> In the automatic setting, it yields 51.76% accuracy, again outperforming both **Textual Entailment** and **DistilBERT**.<sup>13</sup> Further, our model, which outperformed **Semantic parser + LM** in all settings, has potential to improve semantic role labeling.

<sup>9</sup>At  $p = 3 \times 10^{-5}$  and  $p = 0.008$ , respectively, by McNemar’s Test.

<sup>10</sup>At  $p = 3 \times 10^{-4}$  and  $p = 0.02$ , respectively, by McNemar’s Test.

<sup>11</sup>e.g., the choice of preposition in “environmental effect *of*” vs. “environmental effect *on*” indicates whether “environment” should serve as subject or object in the paraphrase.

<sup>12</sup>At  $p = 0.12$  and  $p = 0.34$ , respectively, by McNemar’s Test.

<sup>13</sup>Statistically significant against Textual Entailment at  $p = 0.006$ , but not significant against DistilBERT at  $p = 0.33$

<sup>8</sup><https://demo.allennlp.org/textual-entailment/roberta-snli>

Metric → ↓ Model	Word order accuracy		Paraphrase accuracy	
	Gold args	Automatic	Gold args	Automatic
Sentence similarity	30.62%	23.04%	12.20%	8.94%
Majority + LM	34.15%	28.73%	23.58%	19.51%
GPT-2	53.93%	36.59%	43.36%	30.35%
Semantic parser + LM	56.64%	45.26%	34.15%	26.29%
Textual entailment	60.70%	43.36%	38.21%	27.64%
DistilBERT	62.87%	49.32%	43.36%	32.52%
Textual entailment + LM (DistilBERT)	<b>65.58%</b>	<b>51.76%</b>	45.26%	36.31%
Textual entailment + LM (GPT-2)	62.60%	44.99%	<b>51.22%</b>	<b>38.48%</b>

Table 3: Paraphrase generation performance of the various models in terms of word order accuracy and paraphrase accuracy

## 7 Conclusion

We have presented a study on generating clausal paraphrases for compound nominalizations. Extending previous work, our dataset contains nominalizations with two arguments, namely a prepositional phrase argument as well as a nominal or adjectival modifier. We evaluated a number of unsupervised methods for paraphrase generation. Experimental results show that using a textual entailment model, followed by re-ranking with a language model score, yields the best performance. The proposed method can contribute to downstream NLP tasks that require natural language understanding of texts in which nominalizations are frequently found.

## Acknowledgments

This project was supported by an HKSAR UGC Teaching & Learning Grant (Meeting the Challenge of Teaching and Learning Language in the University: Enhancing Linguistic Competence and Performance in English and Chinese, 2016-19 Triennium).

## References

- Cristina Butnariu, Su Nam Kim, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz, and Tony Veale. 2010. SemEval-2010 Task 9: The interpretation of noun compounds using paraphrasing verbs and prepositions. In *Proc. NAACL HLT Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, page 100–105.
- K. B. Cohen, M. Palmer, and L. Hunter. 2008. Nominalization and alternations in biomedical language. *PLoS ONE*, 3(9):e3158.
- Averil Coxhead. 2016. A New Academic Word List. *TESOL Quarterly*, 34(2):213–238.
- Mark A. Greenwood. 2004. Using Pertainyms to Improve Passage Retrieval for Questions Requesting Information About a Location. In *Proc. SIGIR Workshop on Information Retrieval for Question Answering (IR4QA)*, page 17–22, Sheffield, UK.
- Nizar Habash and Bonnie Dorr. 2003. A Categorical Variation Database for English. In *Proc. NAACL*.
- Matthias Hartung, Fabian Kaupmann, Soufian Jebbara, and Philipp Cimiano. 2017. Learning Compositionality Functions on Word Embeddings for Modelling Attribute Meaning in Adjective-Noun Phrases. In *Proc. 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, page 54–64.
- Iris Hendrickx, Preslav Nakov, Stan Szpakowicz, Zornitsa Kozareva, Diarmuid Ó Séaghdha, and Tony Veale. 2013. SemEval-2013 task 4: Free paraphrases of noun compounds. In *Proc. 7th International Workshop on Semantic Evaluation (SemEval)*, page 138–143.
- Halil Kilicoglu, Marcelo Fiszman, Graciela Rosembat, Sean Marimpietri, and Thomas C. Rindfleisch. 2010. Arguments of Nominals in Semantic Interpretation of Biomedical Text. In *Proc. Workshop on Biomedical Natural Language Processing*.
- Maria Lapata. 2002. The Disambiguation of Nominalizations. *Computational Linguistics*, 28(3):357–388.
- Judith N. Levy. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press, London.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Preslav I. Nakov and Marti A. Hearst. 2013. Semantic Interpretation of Noun Compounds Using Verbal and Other Paraphrases. *ACM Transactions on Speech and Language Processing*, 10(3).



- Jeremy Nicholson and Timothy Baldwin. 2008. Interpreting Compound Nominalisations. In *Proc. LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*.
- Sebastian Padó, Marco Pennacchiotti, and Caroline Sporleder. 2008. Semantic role assignment for event nominalisations by leveraging verbal data. In *Proc. COLING*.
- Girishkumar Ponkiya, Rudra Murthy, Pushpak Bhattacharyya, and Girish Palshikar. 2020. Looking inside noun compounds: Unsupervised prepositional and free paraphrasing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, page 4313–4323.
- Sameer Pradhan, Honglin Sun, Wayne Ward, James H. Martin, and Dan Jurafsky. 2004. Parsing Arguments of Nominalizations in English and Chinese. In *Proc. HLT/NAACL*.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21:1–67.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proc. EMNLP-IJCNLP*.
- Julian Salazar, Davis Liang, Toan Q. Nguyen, and Katrin Kirchhoff. 2020. Masked Language Model Scoring. In *Proc. ACL*.
- David Samuel and Milan Straka. 2020. ÚFAL at MRP 2020: Permutation-invariant Semantic Parsing in PERIN. In *Proc. CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, page 53–64.
- Stephen Tratz and Eduard Hovy. 2010. A Taxonomy, Dataset, and Classifier for Automatic Noun Compound Interpretation. In *Proc. 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, page 678–687.