Unsupervised Paraphrasability Prediction for 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

Commonly found in academic and formal texts, a nominalization uses a deverbal noun to describe an event associated with its corresponding verb. Nominalizations can be difficult to interpret because of ambiguous semantic relations between the deverbal noun and its arguments. Automatic generation of clausal paraphrases for nominalizations can help disambiguate their meaning. However, previous work has not identified cases where it is awkward or impossible to paraphrase a compound nominalization. This paper investigates unsupervised prediction of paraphrasability, which determines whether the prenominal modifier of a nominalization can be re-written as a noun or adverb in a clausal paraphrase. We adopt the approach of overgenerating candidate paraphrases followed by candidate ranking with a neural language model. In experiments on an English dataset, we show that features from an Abstract Meaning Representation graph lead to statistically significant improvement in both paraphrasability prediction and paraphrase generation.

1 Introduction

A nominalization is a noun (e.g., "response") that is morphologically derived from a verb ("respond"), and that designates some aspects of the event referred to by the verb (Quirk et al., 1985). In a compound nominalization, this deverbal noun may have both prenominal and postnominal modifiers. The prenominal modifier can be a noun (e.g., "*police* response to the rioting") or an adjective ("*bodily* injury to a friend"), while postnominal modifiers are prepositional phrases ("presidential nomination *of Harrison*").

Academic and other formal texts utilize nominalization extensively to produce a compact and abstract writing style. The meaning of compound nominalizations can however be difficult to interpret because of ambiguous semantic relations between the deverbal noun and its modifiers. In particular, the prenominal modifier can play multiple semantic roles in the corresponding predicate or clausal paraphrase: as a subject (e.g., "the *police* response" $\rightarrow$ "the *police* responds"); as an object ("*bodily* injury" $\rightarrow$ "injure the *body*"); as an oblique ("*presidential* nomination" $\rightarrow$ "nominate *as president*"; as an adverb ("*symbolic* admission" $\rightarrow$ "admit *symbolically*"; or none of the above ("*stellar* performance" $\not\rightarrow$ "a *star* performs").

The paraphrasability of the prenominal modifier — whether it describes an entity, the manner of an action, or neither — therefore has direct impact on NLP tasks that require interpretation of compound nominalizations. This ambiguity affects accuracy in relation extraction, which is important for information retrieval and question answering (Greenwood, 2004; Klein et al., 2020). A machine translation system must also be able to render the deverbal noun and its prenominal modifier properly when there is no equivalent nominalization in the target language. Further, paraphrasability prediction could benefit nominal semantic role labeling, which needs to identify the role played by the prenominal modifier (Lapata, 2002; Padó et al., 2008; Kilicoglu et al., 2010). Finally, it is critical for nominalization paraphrasing. When a clausal paraphrase is not available for the input nominalization, approaches that do not consider paraphrasability may produce an invalid or misleading output (Lee et al., 2021).

This study focuses on English, the dominant language for academic texts. It aims to make two contributions. First, we enlarge an existing dataset to cover three paraphrasability categories for prenominal modifiers in a compound nominalization (paraphrased as noun, as adverb or non-paraphrasable). Second, we extend an algorithm to take paraphrasability into account, and show that features from Abstract Meaning Representation graphs improve performance in both paraphrasabil-

Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3254 - 3263
July 10-15, 2022

ity prediction and paraphrase generation.

The rest of the paper is organized as follows. After defining our task (Section 2), we summarize previous research (Section 3). We then describe our dataset (Section 4) and present our approach (Sections 5-6). Then, we discuss experimental results (Section 7-8) and conclude (Section 9).

2 Paraphrasability of nominalizations

2.1 Motivation

Our goal is to paraphrase a nominalization into a "clausal paraphrase", which we define as a clause headed by a verb whose syntactic arguments (e.g., subject, object, and prepositional object) are transformed from the nominal arguments in the input nominalization. We focus on compound nominalizations in which the head noun has a prenominal modifier and a prepositional object, following the syntactic form targeted by the only publicly available dataset for nominalization paraphrasing (Lee et al., 2021). Some example inputs and outputs are shown in Table 1.

A clausal paraphrase would not be possible for a compound nominalization if there is no suitable verb equivalent for its head noun. Large-scale language resources (Meyers et al., 1998) already exist to help determine whether such a verb exists, and the task has been tackled in the context of QA semantic role labeling for nominalization (Klein et al., 2020). Less attention has been paid to another factor, namely, whether the prenominal modifier can be expressed as a subject, object or prepositional object in the paraphrase. We are not aware of previous data-driven research on this task, which is the focus of this paper. We will not consider the prepositional object in the input nominalization, since it can be incorporated into a clausal paraphrase in most cases.

2.2 Task definition

The term "paraphrasability" has been used for the degree of semantic equivalence between syntactic variants of predicate phrases (Fujita and Sato, 2008). We will use this term to refer to the three categories of paraphrasing behavior of prenominal modifiers in compound nominalizations:

Noun The prenominal modifier is a noun, or is an adjective that *pertains* to a noun, that can serve as the subject, object or prepositional object in a clausal paraphrase. In other words, either the prenominal modifier itself (e.g., "police") or its *pertainym* ("president" for "presidential") literally refers to the entity that participates in the event denoted by the deverbal noun ("*police* response", "*presidential* nomination").

Adverb The prenominal modifier is an adjective that can appear in the clausal paraphrase in its adverbial form (e.g., "*frontal* opposition" $\rightarrow$ "oppose *frontally*"), but not as pertainym ("*frontal* opposition" $\not\rightarrow$ "the *front* opposes").

Nil The prenominal modifier cannot be paraphrased with either method above (e.g., "*stellar* performance" $\not\rightarrow$ "*a star* performs"; "*brain* drain" $\not\rightarrow$ "drain a *brain*").

As shown in Table 1, the input is a nominalization that consists of a deverbal noun (derived from the verb V); its prenominal modifier (bolded); and a prepositional phrase. The output of the **Paraphrasability Prediction** task is the part-of-speech label of the word to which the prenominal modifier is paraphrased (bolded). The label can be `Noun`, `Adverb`, or `Nil` when it is not paraphrasable.

The output of the **Paraphrase Generation** task is a clausal paraphrase of the input. It incorporates the verb V, the prepositional object from the input (marked with O); and either a noun (marked with M) or an adverb (marked with B) corresponding to the prenominal modifier. The gold paraphrase of the `Nil` type input is defined as *null*. The only way to render such an input as a clause is with a support verb or light verb (e.g., "stellar performance" $\rightarrow$ "*give* a stellar performance"). Since the paraphrase retains the original nominalization, it does not serve our goal of unpacking its meaning.

3 Previous work

3.1 Noun literality prediction

There has been extensive research on compositionality analysis on noun compounds (Reddy et al., 2011), adjective-noun combinations and other types of multiword expressions (MWEs) (Biemann and Giesbrecht, 2011; Ramisch et al., 2016; Cordeiro et al., 2019; Jana et al., 2019). Compositionality refers to the extent to which the meaning of the MWE can be expressed in terms of the meaning of its constituents. It therefore has considerable overlap with literality prediction, which would identify, for example, the noun "rat" in "*rat* race" as non-literal (Reddy et al., 2011).

Input Nominalization	Paraphrase Generation	Paraphrasability Prediction
American influence on global culture **police** response to the rioting	**America**^{M} influences^V global culture^O the **police**^{M} responds^V to the rioting^O	`Noun`
climatic effects of air pollutants **war** preparations of the government	air pollutants^O affect^V the **climate**^{M} the government^O prepares^V for **war**^{M}	
bodily injury to a friend **student** admissions into universities	injure^V the **body**^{M} of a friend^O admit^V **students**^{M} into universities^O	
presidential nomination of Harrison **business** travel to Greece	nominate^V Harrison^O as **president**^{M} travel^V to Greece^O for **business**^{M}	
naval assistance from Italian powers **majority** decision of the panel	**navies**^{M} of Italian powers^O assist^V a **majority**^{M} of the panel^O decides^V	
frontal opposition of the employers **dynamic** allocation of licenses	The employers^O oppose^V **frontally**^{B} allocate^V licenses^O **dynamically**^{B}	`Adverb`
stellar performance of the rookies **brain** drain of scientists	*null* *null*	`Nil`

Table 1: Input and output of the **Paraphrasability Prediction** and **Paraphrase Generation** task (Section 2.2)

Our task is closely related to literality prediction since compound nominalizations are a subset of noun-noun compounds; in particular, a prenominal modifier that is "literal" would likely be of paraphrasability type `Noun` (Section 2.2). We will therefore evaluate the performance of a state-of-the-art noun literality prediction model (Shwartz and Dagan, 2019) in our experiment.

Our task is nonetheless distinct from literality prediction since it focuses on paraphrasability rather than literalness. For example, even when a prenominal modifier is used metaphorically and is non-literal (e.g. "*circular* argument"), it would be labeled `Noun` in terms of paraphrasability if it can appear in a clausal paraphrase ("argue in a *circle*").

3.2 Noun compound interpretation

A noun compound can be disambiguated with a free-form paraphrase (Hendrickx et al., 2013), or with verbs and prepositions linking the two nouns, e.g., "apple pie" is a "*pie* with *apples*" (Butnariu et al., 2010; Nakov and Hearst, 2013). Unsupervised approaches have been found to be effective for noun compound interpretation. Paraphrase templates with slots for prepositions and predicates, for example, can be filled using pre-trained masked language models (Ponkiya et al., 2020). We will likewise investigate unsupervised approaches in this work. Even though compound nominalizations are a subset of noun-noun compounds, our task is different since paraphrases in noun compound interpretation do not transform the head noun into a verb.

3.3 Paraphrasing nominalizations

Research on nominalization interpretation has mostly focused on nominal semantic role labeling, which assigns abstract labels (e.g., agent, patient, manner) to arguments of nominalizations (Lapata, 2002; Nicholson and Baldwin, 2008; Padó et al., 2008; Kilicoglu et al., 2010). Given the systematic correspondences between nominalization and clause structure, there have also been efforts to paraphrase nominalizations as clauses. Algorithms have been proposed for automatic acquisition of paraphrase templates, which can cover nominalization inputs (Shinyama et al., 2002). The paraphrasing task has also been indirectly addressed in a model for question and answer generation from nominalizations (Klein et al., 2020).

The most closely related work to this paper was reported in Lee et al. (2021). Their proposed model first overgenerates paraphrase candidates, and then uses textual entailment to identify the optimal candidate. However, since all nominalizations in their dataset have paraphrases, their algorithm makes no judgment on paraphrasability. We extend their dataset and investigate paraphrasability prediction to fill in this research gap.

4 Dataset

The only publicly available dataset of clausal paraphrases, developed by Lee et al. (2021), provides 450 paraphrases for English nominalizations. All

Paraphrasability Label	Adjectival modifier	Nominal modifier
Noun	281	169
Adverb	55	n/a
Nil	91	38

Table 2: Breakdown of our dataset according to paraphrasability label and part-of-speech of the prenominal modifier

instances are of the paraphrasability label Noun (Section 2.2). To facilitate our study, we enlarged this dataset with inputs of the Adverb and Nil paraphrasability labels.

4.1 Data source

In the interest of consistency with the existing dataset, we focus on nominalizations with the same syntactic pattern. Specifically, we collected sentences from English Wikipedia that contain a noun phrase headed by a deverbal noun with one prenominal modifier and one postnominal modifier.

As shown in Table 1, the postnominal modifier is a prepositional phrase with prepositional object O. The prenominal modifier can be a noun or an adjective. To create a challenging dataset, the adjective must have a pertainym, or can itself also serve as noun (e.g., "light"), such that multiple paraphrasability labels are plausible.

4.2 Annotation

Two annotators, a native speaker and a near-native speaker of English, independently classified the nominalizations into one of three paraphrasability labels (Section 2.2). For those labeled as Adverb, the annotator composed a paraphrase with an adverb that is derivationally related to the adjective. For those labeled as Nil, no paraphrase was required. Examples for each label are provided in Table 1.

A professor of linguistics who is a native speaker of English reviewed the annotation, either keeping both or selecting one of them. A total of 184 non-Noun instances were collected and added to the dataset, resulting in an expanded dataset with 634 paraphrases (Table 2).[1]

5 Paraphrase candidates

Our approach is to first overgenerate paraphrase candidates for each input, and then identify the optimal candidate. This section presents the candidate types, and the next section describes the candidate selection algorithm.

Table 3 shows the paraphrase candidates for the input "*frontal* opposition of the employers". The prenominal modifier, "frontal", is transformed into various parts-of-speech and placed at different positions in the candidates. Each candidate is associated with one of the three paraphrasability types:

5.1 Noun

The Noun type candidates are the five paraphrases defined in Lee et al. (2021). The prenominal modifier is paraphrased as a noun in the subject (MVO, MOV), object (OVM, VMO), and oblique (VOM) positions.

5.2 Adverb

There are four paraphrase candidates for the Adverb type. The prenominal modifier must be an adjective. It is paraphrased as an adverb that pertains to itself, according to WordNet (Fellbaum, 2010); or an adverb that is derivationally related to itself, according to CatVar (Habash and Dorr, 2003). The adverb (B) is placed either before the verb (BVO, OBV) or at the end of the clause (VOB, OVB).

5.3 Nil

There are, by definition, no obvious paraphrase candidates to represent inputs of the Nil type. We implemented the following alternatives:

Identity The nominalization input itself.

Light verb The paraphrase retains the nominalization as the object of a light verb or support verb (Grefenstette and Teufel, 1995). One paraphrase candidate prepends the light verb; e.g., "home run against Arizona" → "*hit* a home run against Arizona". The other candidate uses the prepositional object (O) as subject; e.g., "stellar performance of the rookies" → "the rookies *give* a stellar performance".

Predicative adjective Limited to prenominal modifiers that are adjectives, this paraphrase uses the adjective predicatively to form a clause. The paraphrase is designed, on the one hand, to be acceptable for Nil type inputs, e.g., "*stellar* performance of the rookies" → "the performance of the rookies is *stellar*"; and on the other hand, to be unacceptable for

[1] Accessible at https://github.com/NominalizationParaphrase

Paraphrasability Label	Paraphrase Candidate	Word Order
`Noun`	the *front* opposes the employers	MVO
	the employers oppose the *front*	OVM
	oppose the *front* of the employers	VMO
	oppose the employers at the *front*	VOM
	front employers oppose	MOV
`Adverb`	oppose the employers *frontally*	VOB
	frontally oppose the employers	BVO
	the employers oppose *frontally*	OVB
	the employers *frontally* oppose	OBV
`Nil`	*frontal* opposition of the employers	Identity
	Opposition of the employers is *frontal*	Predicative Adjective
	have *frontal* opposition of the employers	Light Verb
	the employers face *frontal* opposition	Light Verb

Table 3: Automatically generated paraphrase candidates (Section 5) for the input "frontal opposition of the employers", which place the prenominal modifier (M or B), verb (V) and prepositional object (O) in the input nominalization in different word orders

nominal adjectives, which cannot be used as a predicate (Coates, 1971), e.g., "*presidential* nomination of Harrison" $\not\rightarrow$ "the nomination of Harrison is *presidential*".

We used a masked language model, BERT (Devlin et al., 2019), to generate the most likely determiners, prepositions and light verbs for the paraphrase candidates above.

6 Approach

As discussed in Section 4.1, the input is a sentence that contains a nominalization, headed by a deverbal noun that has a prenominal modifier and a prepositional phrase (Table 1). The prenominal modifier may be a noun or an adjective.

We first overgenerate paraphrases to construct a candidate pool (Section 6.1), and then filter the pool by considering paraphrasability (Section 6.2). For the Paraphrase Generation task, the output is the best candidate selected by the textual entailment and language models (Section 6.3). For the Parapharasability Prediction task, the output is the paraphrasability label associated with the selected candidate (Table 3).

6.1 Candidate pool construction

This step constructs a pool of paraphrase candidates. We evaluated the following methods:

All Include all paraphrase candidates in Table 3.

Gold Include only those paraphrase candidates associated with the gold paraphrasability label.

Majority baseline Include only those paraphrase candidates associated with the majority paraphrasability label, which is `Noun` in our dataset. This baseline replicates the algorithm proposed by Lee et al. (2021), which considers only the MVO, OVM, VMO, VOM, and MOV paraphrases in Table 3.

Word frequency baseline Include the `Noun` type (`Adverb` type) paraphrase candidates only when the noun (the adverb) corresponding to the prenominal modifier has high frequency. The frequency threshold is optimized on our dataset based on frequency statistics in the Google Web 1T N-gram Corpus (Brants and Franz, 2006).

6.2 Candidate pool filtering

This step filters the candidate pool constructed by the **All** model. We evaluated two methods that consider paraphrasability through semantic parsing and literality prediction, respectively.

6.2.1 Filtering with AMR

Abstract Meaning Representation (AMR) abstracts away from the syntactic realization of a sentence and expresses its meaning with a directed acyclic graph, where nodes represent events and concepts, and edges represent relationships between

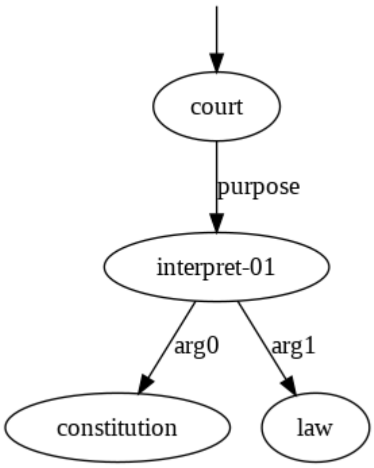

Figure 1: AMR graph of the sentence "... a court for the *constitutional* interpretation of law", which is predicted as `Noun` by the All+AMR model (Section 6.2.1)

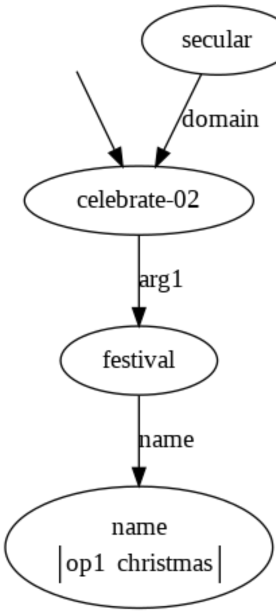

Figure 2: AMR graph of the sentence "... the *secular* celebration of Christmas", which is predicted as non-`Noun` by the All+AMR model (Section 6.2.1)

the nodes (Banarescu et al., 2013). In an AMR graph, deverbal nouns are annotated as verbs, and adjectives pertaining to nouns are annotated in their nominal form whenever possible.

We use PERIN (Samuel and Straka, 2020) to construct an AMR graph for the input sentence, and align the nodes with the words in the input. We will refer to the node aligned to the deverbal noun as the "deverbal noun node"; and the node aligned to the prenominal modifier as the "prenominal modifier node". In Figure 1, the "constitution" node is the prenominal modifier node (aligned to "constitutional" in the input); and the "interpret-01" node is the deverbal noun node (aligned to "interpretation").

The **All+AMR** model predicts `Noun` as paraphrasability label and removes all non-`Noun` type candidates from the pool if the prenominal modifier node:

- is an argument of the deverbal noun node; or
- is the domain of the deverbal noun node, and is annotated as a noun.

Otherwise, it predicts paraphrasability to be non-`Noun` and removes all `Noun` type candidates from the pool.

For example, the model predicts `Noun` as paraphrasability label for the sentence in Figure 1, since the prenominal modifier node ("constitution") serves as arg0 to the deverbal noun node ("interpret-01"). The model rejects `Noun` as paraphrasability label for the sentence in Figure 2. Even though the prenominal modifier node ("secular") is the domain of the deverbal noun node ("celebrate-02"), it is annotated with the original adjective rather than its nominal form.

6.2.2 Filtering with noun literality prediction

Noun literality prediction is closely related to paraphrasability prediction (Section 3.1). We use LexComp (Shwartz and Dagan, 2019), a state-of-the-art model in noun literality prediction, which is trained on datasets from Reddy et al. (2011) and Tratz et al. (2010) using contextualized word embeddings.[2]

Given the prenominal modifier and the deverbal noun in the input, the **All+LexComp** model predicts `Noun` as paraphrasability label if LexComp predicts "literal", and removes all non-`Noun` type candidates. Otherwise, it predicts `Nil` and keeps only the `Nil` type paraphrases. This model does not perform filtering on an input with an adjectival modifier.

6.3 Candidate selection

A textual entailment model (TE), enhanced with re-ranking by language model scores, was found to yield the strongest performance in paraphrase generation for compound nominalizations (Lee et al., 2021). Taking the nominalization input as the

[2] https://github.com/vered1986/lexcomp

premise and a paraphrase candidate as the hypothesis, the TE model predicts whether the facts in the former imply those in the latter. Among the three candidate paraphrases that yield the highest TE scores, the candidate with the highest language model (LM) score is selected.

We replicate this algorithm and apply it on the filtered candidate pool. We use the AllenNLP textual entailment model[3], and the log-probability score based on GPT-2 (117M) as the LM score (Salazar et al., 2020).[4]

7 Experimental set-up and metrics

The entire dataset was used for evaluation since our approach is unsupervised. We used SpaCy (Honnibal and Johnson, 2015) for POS tagging to determine the POS of the prenominal modifier.

Paraphrasability prediction. We report precision, recall and F_1. Precision is defined as the number of actual Noun instances, out of all instances predicted by the system as Noun. Recall is defined as the proportion of gold Noun instances that are correctly identified by the system as Noun.

Paraphrase generation. We report "paraphrase accuracy" and "word order accuracy" as defined in Lee et al. (2021). For the former, the determiners are removed from all paraphrases. The system is considered correct if the lemmatized form of all words in the predicted paraphrase are identical with those in the gold paraphrase. The latter is defined likewise, except that prepositions are not taken into consideration. It thus essentially measures the system's ability to predict the verb and arguments and to put them into the correct word order. The word orders VOB/BVO and OVB/OBV are considered interchangeable.

8 Results

Table 4 shows system performance on the paraphrasability prediction (Section 8.1) and its effect on paraphrase generation (Section 8.2).

8.1 Paraphrasability prediction

Given the preponderance of the Noun label in our dataset, the Majority baseline produced a strong performance at 0.673 precision and perfect recall. It outperforms the Word Frequency baseline, which has slightly higher precision (0.686) but lower recall (0.911), both in terms of F_1 and accuracy.

Using all paraphrase candidates resulted in an improvement in binary classification accuracy (0.711 vs. 0.673) over the Majority baseline, demonstrating the effectiveness of the Adverb and Nil paraphrases (Section 5.2-5.3). In terms of three-way classification, however, it offered no improvement over the Majority baseline (0.671 vs. 0.673). This indicates that while the candidate selection method (Section 6.3) can correctly detect some Noun type candidates as inappropriate, it is less competent in judging between Adverb vs. Nil paraphrases.

The All+LexComp model raised the accuracy by only 0.4% in comparison to the All model. This result suggests that noun literality prediction is only slightly helpful as a proxy for paraphrasability.

The All+AMR model achieved the highest F_1 (0.852) by raising both the precision and recall of the All model. The improvement is statistically significant in terms of both binary classification (0.782)[5] and three-way classification (0.744)[6]. These results show that AMR is useful for predicting paraphrasability, which may be due to the more fine-grained semantic information in the AMR graphs that could not be inferred by the LM and TE models in the candidate selection step. The improvement of the All+AMR model over the All+LexComp model is also significant[7], likely because the semantic features in the AMR graphs are more relevant to paraphrasability than literality.

Table 5 shows the paraphrasability labels predicted by the All+AMR model. While it was able to identify most of the Noun inputs, it did so for only half of the Adverb ones. The most challenging turned out to be the Nil inputs, which the model succeeded in detecting less than one-third of the time.

8.2 Paraphrase generation

Despite its higher accuracy in paraphrasability prediction, the Majority baseline (0.264 paraphrase accuracy) performed worse than the Word Frequency baseline (0.275) in paraphrase generation. This likely reflects the greater challenge in identifying the correct Noun type paraphrases than the Advice and Nil types.

For the other models, performance in para-

[3]https://demo.allennlp.org/textual-entailment/roberta-snli
[4]https://github.com/awslabs/mlm-scoring

[5]$p = 0.000556$ according to McNemar's Test with continuity correction; the same test is used henceforth
[6]At $p = 0.000172$
[7]$p = 0.00169$ on binary classification and $p = 0.000667$ on three-way classification

Metric → ↓ Model	Paraphrasability Prediction					Paraphrase Generation	
	Noun vs. non-Noun				3-way	Word order	Paraphrase
	P	R	F_1	Accuracy	Accuracy	Accuracy	Accuracy
Word Frequency	0.686	0.911	0.783	0.660	0.645	0.384	0.275
Majority	0.673	**1.000**	0.804	0.673	0.673	0.376	0.264
All	0.743	0.873	0.802	0.711	0.671	0.416	0.318
All+LexComp	0.752	0.859	0.802	0.715	0.675	0.424	0.333
All+AMR	**0.784**	0.932	**0.852**	**0.782**	**0.744**	**0.462**	**0.347**
Gold	1	1	1	1	1	0.704	0.591

Table 4: Performance on paraphrasability prediction and paraphrase generation

Predicted → ↓ Gold	Noun	Adverb	Nil
Noun	**345**	13	12
Adverb	17	**24**	8
Nil	**78**	13	40

Table 5: Contingency table for the paraphrasability prediction of the All+AMR model

phrasability prediction is largely correlated to paraphrase generation. The All model improved over the Majority baseline both in terms of word order (0.416) and paraphrase accuracy (0.318). Consider the input "... the apocalyptic destruction of the town and the cult". The All model correctly declined to paraphrase (*null* output) on the basis of the high score secured by the light-verb paraphrase "the town and the cult suffer an apocalyptic destruction". In contrast, the Majority baseline produced the inappropriate paraphrase "the apocalypses destroy the town and the cult".

The All+AMR model again offered the best performance, at 0.462 word order accuracy and 0.347 paraphrase accuracy.[8] For the sentence in Figure 1, the All model generated the predicative adjective paraphrase "interpretation of law is constitutional" due to the high LM score, even though the word "constitutional" yields a different meaning in this paraphrase. The All+AMR model was able to reject this paraphrase since the word "constitution" was inferred to play the subject role. Conversely, in Figure 2, the model was able to reject paraphrases involving the noun "secularism", since the AMR parser annotated with the original adjective "secular". The considerable performance gap from the Gold model (paraphrase accuracy 0.591), however, indicates there is still much room for improvement in interpreting nominalizations.

[8]The improvement in word order accuracy is statistically significant over the All and All+LexComp models at $p = 0.0211$ and $p = 0.0482$, respectively. The improvement in paraphrase accuracy is not significant, however, at $p = 0.0970$ and $p = 0.428$ against the All and All+LexComp models, respectively.

9 Conclusion

A clausal paraphrase can help disambiguate a nominalization semantically, especially when the prenominal modifier is difficult to interpret. This paper has presented the first study on determining the paraphrasability of the prenominal modifier in a compound nominalization. We have expanded an existing dataset to cover cases when the prenominal modifier can appear as a noun in the paraphrase, as an adverb, or not at all.

Our experiments suggest that overgeneration of paraphrase candidates, followed by ranking with a textual entailment model and language model, can yield competitive results. Further, AMR-based features lead to statistically significant improvement in performance.

A limitation of our study is the restricted syntactic form of the input nominalizations. To facilitate more comprehensive evaluation, future research should consider expanding the dataset further to cover a wider range of nominalizations, and richer variations in their clausal paraphrases.

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

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proc. 7th Linguistic Annotation Workshop and Interoperability with Discourse*.

Chris Biemann and Eugenie Giesbrecht. 2011. Distributional Semantics and Compositionality 2011: Shared Task Description and Results. In *Proc. Workshop on Distributional Semantics and Compositionality (DiSCo)*.

Thorsten Brants and Alex Franz. 2006. The Google Web 1T 5-gram Corpus Version 1.1. In *LDC2006T13*.

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.

Jennifer Coates. 1971. Denominal Adjectives: A Study in Syntactic Relationships between Modifier and Head. *Lingua*, 27:160–169.

Silvio Cordeiro, Aline Villavicencio, Marco Idiart, and Carlos Ramisch. 2019. Unsupervised compositionality prediction of nominal compounds. *Computational Linguistics*, 45(1):1–57.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding. In *Proc. North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT)*.

Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243, Springer, Dordrecht.

Atsushi Fujita and Satoshi Sato. 2008. A Probabilistic Model for Measuring Grammaticality and Similarity of Automatically Generated Paraphrases of Predicate Phrases. In *Proc. 22nd International Conference on Computational Linguistics (COLING)*.

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.

Gregory Grefenstette and Simone Teufel. 1995. Corpus-based Method for Automatic Identification of Support Verbs for Nominalizations. In *Proc. 7th Conference on European chapter of the Association for Computational Linguistics (EACL)*.

Nizar Habash and Bonnie Dorr. 2003. A Categorial Variation Database for English. In *Proc. North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL)*.

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.

Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Abhik Jana, Dima Puzyrev, Alexander Panchenko, Pawan Goyal, Chris Biemann, and Animesh Mukherjee. 2019. On the Compositionality Prediction of Noun Phrases using Poincaré Embeddings. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Halil Kilicoglu, Marcelo Fiszman, Graciela Rosemblat, Sean Marimpietri, and Thomas C. Rindflesch. 2010. Arguments of Nominals in Semantic Interpretation of Biomedical Text. In *Proc. Workshop on Biomedical Natural Language Processing*.

Ayal Klein, Jonathan Mamou, Valentina Pyatkin, Daniela Stepanov, Hangfeng He, Dan Roth, Luke Zettlemoyer, and Ido Dagan. 2020. QANom: Question-Answer driven SRL for Nominalizations. In *Proc. 28th International Conference on Computational Linguistics (COLING)*.

Maria Lapata. 2002. The Disambiguation of Nominalizations. *Computational Linguistics*, 28(3):357–388.

John Lee, Ho Hung Lim, and Carol Webster. 2021. Paraphrasing Compound Nominalizations. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Adam Meyers, Catherine Macleod, Roman Yangarber, Ralph Grishman, Leslie Barrett, and Ruth Reeves. 1998. Using NOMLEX to Produce Nominalization Patterns for Information Extraction. In *Proc. Computational Treatment of Nominals*.

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. 22nd International Conference on Computational Linguistics (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.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.

Carlos Ramisch, Silvio Cordeiro, Leonardo Zilio, Marco Idiart, Aline Villavicencio, and Rodrigo Wilkens. 2016. How naked is the naked truth? a multilingual lexicon of nominal compound compositionality. In *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*.

Siva Reddy, Diana McCarthy, and Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218.

Julian Salazar, Davis Liang, Toan Q. Nguyen, and Katrin Kirchhoff. 2020. Masked Language Model Scoring. In *Proc. Annual Meeting of the Association for Computational Linguistics (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.

Yusuke Shinyama, Satoshi Sekine, and Kiyoshi Sudo. 2002. Automatic Paraphrase Acquisition from News Articles. In *Proc. Second international conference on Human Language Technology Research (HLT)*.

Vered Shwartz and Ido Dagan. 2019. Still a Pain in the Neck: Evaluating Text Representations on Lexical Composition. *Transactions of the Association for Computational Linguistics*, 7:403–419.

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.