Noun Compound Interpretation Using Paraphrasing Verbs: Feasibility Study

Preslav Nakov**

Linguistic Modeling Department,
Institute for Parallel Processing,
Bulgarian Academy of Sciences
25A, Acad. G. Bonchev St., 1113 Sofia, Bulgaria
and
Department of Mathematics and Informatics,
Sofia University,
5, James Bourchier Blvd., 1164 Sofia, Bulgaria
nakov@lml.bas.bg
http://nakov.eu

Abstract. The paper addresses an important challenge for the automatic processing of English written text: understanding noun compounds' semantics. Following Downing (1977) [1], we define noun compounds as sequences of nouns acting as a single noun, e.g., *bee honey*, *apple cake*, *stem cell*, etc. In our view, they are best characterised by the set of all possible paraphrasing verbs that can connect the target nouns, with associated weights, e.g., *malaria mosquito* can be represented as follows: *carry* (23), *spread* (16), *cause* (12), *transmit* (9), etc. These verbs are directly usable as paraphrases, and using multiple of them simultaneously yields an appealing fine-grained semantic representation.
In the present paper, we describe the process of constructing such representations for 250 noun-noun compounds previously proposed in the linguistic literature by Levi (1978) [2]. In particular, using human subjects recruited through Amazon Mechanical Turk Web Service, we create a valuable manually-annotated resource for noun compound interpretation, which we make publicly available with the hope to inspire further research in paraphrase-based noun compound interpretation. We further perform a number of experiments, including a comparison to automatically generated weight vectors, in order to assess the dataset quality and the feasibility of the idea of using paraphrasing verbs to characterise noun compounds' semantics; the results are quite promising.

Key words: Noun Compounds, Lexical Semantics, Paraphrasing.

** Part of this research was performed while the author was a PhD student at the EECS department, Computer Science division, University of California at Berkeley.

1 Introduction

An important challenge for the automatic analysis of English written text is posed by noun compounds – sequences of nouns acting as a single noun[1], e.g., *colon cancer tumor suppressor protein* – which are abundant in English: Baldwin&Tanaka'04 [3] calculated that noun compounds comprise 3.9% and 2.6% of all tokens in the *Reuters corpus* and the *British National Corpus*[2], respectively.

Understanding noun compounds' syntax and semantics is difficult but important for many natural language applications (NLP) including but not limited to question answering, machine translation, information retrieval, and information extraction. For example, a question answering system might need to determine whether '*protein acting as a tumor suppressor*' is a good paraphrase for *tumor suppressor protein*, and an information extraction system might need to decide whether *neck vein thrombosis* and *neck thrombosis* could possibly co-refer when used in the same document. Similarly, a machine translation system facing the unknown noun compound *WTO Geneva headquarters* might benefit from being able to paraphrase it as *Geneva headquarters of the WTO* or as *WTO headquarters located in Geneva*. Given a query like *migraine treatment*, an information retrieval system could use suitable paraphrasing verbs like *relieve* and *prevent* for page ranking and query refinement.

Throughout the rest of the paper, we hold the view that noun compounds' semantics is best characterised by the set of all possible paraphrasing verbs that can connect the target nouns, with associated weights, e.g., *malaria mosquito* can be represented as follows: *carry* (23), *spread* (16), *cause* (12), *transmit* (9), etc. Such verbs are directly usable as paraphrases, and using multiple of them simultaneously yields an appealing fine-grained semantic representation.

The remainder of the paper is organised as follows: Section 2 provides a short overview of the different representations of noun compounds' semantics previously proposed in the literature. Section 3 gives details on the process of creating a lexicon of human-proposed paraphrasing verbs for 250 noun-noun compounds. Section 4 describes the experiments we performed in order to assess the lexicon's quality and the feasibility of using paraphrasing verbs to characterise noun compounds' semantics. Section 5 contains a discussion on the applicability of the approach. Section 6 concludes and suggests possible directions for future work.

2 Related Work

The dominant view in theoretical linguistics is that noun compound semantics can be expressed by a small set of abstract relations. For example, in the theory of Levi [2], complex nominals – a general concept grouping together the partially overlapping classes of nominal compounds (e.g., *peanut butter*), nominalisations

[1] This is Downing's definition of noun compounds [1], which we adopt throughout the rest of the paper.

[2] There are 256K distinct noun compounds out of the 939K distinct wordforms in the 100M-word *British National Corpus*.

RDP	Example	Subj/obj	Traditional Name
$\mathtt{CAUSE}_1$	*tear gas*	object	causative
$\mathtt{CAUSE}_2$	*drug deaths*	subject	causative
$\mathtt{HAVE}_1$	*apple cake*	object	possessive/dative
$\mathtt{HAVE}_2$	*lemon peel*	subject	possessive/dative
$\mathtt{MAKE}_1$	*silkworm*	object	productive/composit.
$\mathtt{MAKE}_2$	*snowball*	subject	productive/composit.
USE	*steam iron*	object	instrumental
BE	*soldier ant*	object	essive/appositional
IN	*field mouse*	object	locative
FOR	*horse doctor*	object	purposive/benefactive
FROM	*olive oil*	object	source/ablative
ABOUT	*price war*	object	topic

Table 1. Levi's recoverably deletable predicates (RDPs). Column 3 shows the modifier's function in the corresponding paraphrasing relative clause: when the modifier is the subject of that clause, the RDP is marked with the index 2.

	Subjective	**Objective**	**Multi-modifier**
Act	*parental refusal*	*dream analysis*	*city land acquisition*
Product	*clerical errors*	*musical critique*	*student course ratings*
Agent	—	*city planner*	—
Patient	*student inventions*	—	—

Table 2. Levi's nominalisation types with examples.

(e.g., *dream analysis*), and nonpredicate noun phrases (e.g., *electric shock*) – can be derived by the following two processes:

1. **Predicate Deletion.** It can delete the 12 abstract recoverably deletable predicates (RDPs) shown in Table 1, e.g., *pie made of apples* → *apple pie*. In the resulting nominals, the modifier is typically the object of the predicate; when it is the subject, the predicate is marked with the index 2;
2. **Predicate Nominalisation.** It produces nominals whose head is a nominalised verb, and whose modifier is derived from either the subject or the object of the underlying predicate, e.g., *the President refused general MacArthur's request* → *presidential refusal*. Multi-modifier nominalisations retaining both the subject and the object as modifiers are possible as well. Therefore, there are three types of nominalisations depending on the modifier, which are combined with the following four types of nominalisations the head can represent: *act*, *product*, *agent* and *patient*. See Table 2 for examples.

In the alternative linguistic theory of Warren [4], noun compounds are organised into a four-level hierarchy, where the top level is occupied by the following six major semantic relations: `Possession`, `Location`, `Purpose`, `Activity-Actor`, `Resemblance`, and `Constitute`. `Constitute` is further sub-divided into finer-grained level-2 relations: `Source-Result`, `Result-Source` or `Copula`. Furthermore, `Copula` is sub-divided into the level-3 relations `Adjective-Like_Modifier`, `Subsumptive`, and `Attributive`. Finally, `Attributive` is divided into the level-4 relations `Animate_Head` (e.g., *girl friend*) and `Inanimate_Head` (e.g., *house boat*).

A similar view is dominant in computational linguistics. For example, Nastase&Szpakowicz [5] proposed a two-level hierarchy consisting of thirty fine-grained relations, grouped into the following five coarse-grained ones (the corresponding fine-grained relations are shown in parentheses): `CAUSALITY` (`cause`, `effect`, `detraction`, `purpose`), `PARTICIPANT` (`agent`, `beneficiary`, `instrument`, `object_property`, `object`, `part`, `possessor`, `property`, `product`, `source`, `whole`, `stative`), `QUALITY` (`container`, `content`, `equative`, `material`, `measure`, `topic`, `type`), `SPATIAL` (`direction`, `location_at`, `location_from`, `location`), and `TEMPORALITY` (`frequency`, `time_at`, `time_through`). For example, *exam anxiety* is classified as `effect` and therefore also as `CAUSALITY`.

Similarly, Girju&al. [6] propose a set of 21 abstract relations (`POSSESSION`, `ATTRIBUTE-HOLDER`, `AGENT`, `TEMPORAL`, `PART-WHOLE`, `IS-A`, `CAUSE`, `MAKE/PRODUCE`, `INSTRUMENT`, `LOCATION/SPACE`, `PURPOSE`, `SOURCE`, `TOPIC`, `MANNER`, `MEANS`, `THEME`, `ACCOMPANIMENT`, `EXPERIENCER`, `RECIPIENT`, `MEASURE`, and `RESULT`) and Rosario & Hearst [7] use 18 abstract domain-specific biomedical relations (e.g., `Defect`, `Material`, `Person_Afflicted`).

An alternative view is held by Lauer [8], who defines the problem of noun compound interpretation as predicting which among the following eight prepositions best paraphrases the target noun compound: `of`, `for`, `in`, `at`, `on`, `from`, `with`, and `about`. For example, *olive oil* is *oil <u>from</u> olives.*

Lauer's approach is attractive since it is simple and yields prepositions representing paraphrases directly usable in NLP applications. However, it is also problematic since mapping between prepositions and abstract relations is hard [6], e.g., `in`, `on`, and `at`, all can refer to both `LOCATION` and `TIME`.

Using abstract relations like `CAUSE` is problematic as well. First, it is unclear which relation inventory is the best one. Second, being both abstract and limited, such relations capture only part of the semantics, e.g., classifying *malaria mosquito* as `CAUSE` obscures the fact that mosquitos do not directly cause malaria, but just transmit it. Third, in many cases, multiple relations are possible, e.g., in Levi's theory, *sand dune* is interpretable as both `HAVE` and `BE`.

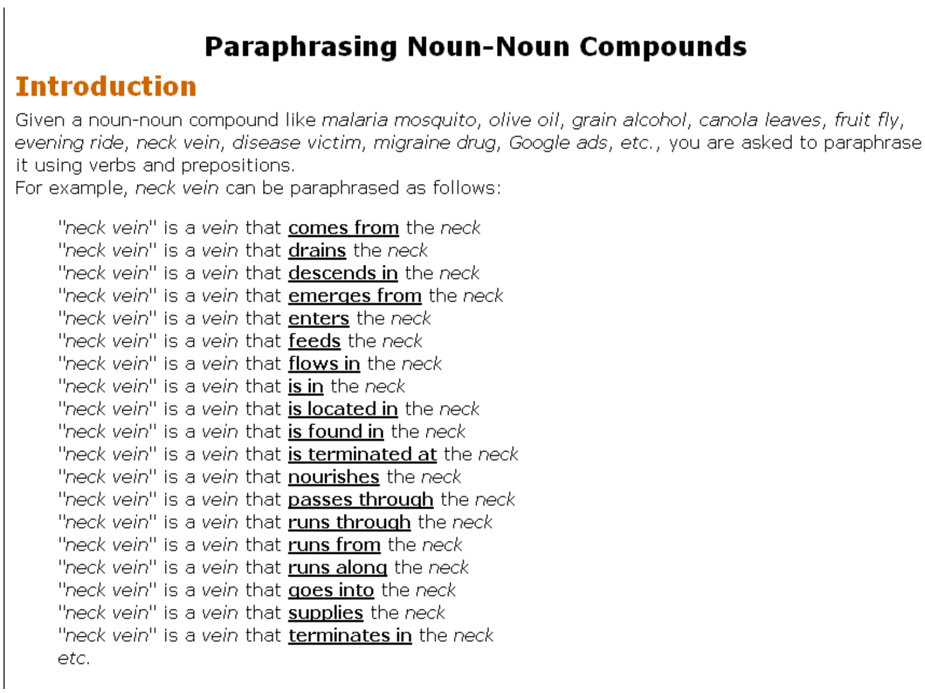
Some of these issues are addressed by Finin [9], who proposes to use a specific verb, e.g., *salt water* is interpreted as *dissolved in.* In a number of publications [10–12], we introduced and advocated an extension of this idea, where noun compounds are characterised by the set of all possible paraphrasing verbs, with associated weights, e.g., *malaria mosquito* can be *carry (23)*, *spread (16)*, *cause (12)*, *transmit (9)*, *etc.* These verbs are fine-grained, directly usable as paraphrases, and using multiple of them for a given noun compound approximates its semantics better.

Following this line of research, below we describe the process of building a lexicon of human-proposed paraphrasing verbs, and a number of experiments in assessing both the lexicon's quality and the feasibility of the idea of using paraphrasing verbs to characterise noun compounds' semantics.

3 Creating a Lexicon of Paraphrasing Verbs

Below we describe the process of creating a new lexicon for noun compound interpretation in terms of multi-sets of paraphrasing verbs. We used the Amazon Mechanical Turk Web Service[3] to recruit human subjects to annotate 250 noun-noun compounds previously proposed in the linguistic literature.

We defined a special noun-noun compound paraphrasing task, which, given a noun-noun compound, asks human subjects to propose verbs, possibly followed by prepositions, that could be used in a paraphrase involving *that*. For example, *nourish*, *run along* and *come from* are good paraphrasing verbs for *neck vein* since they can be used in paraphrases like '*a vein that nourishes the neck*', '*a vein that runs along the neck*' or '*a vein that comes from the neck*'. In an attempt to make the task as clear as possible and to ensure high quality of the results, we provided detailed instructions, we stated explicit restrictions, and we gave several example paraphrases. We instructed the participants to propose at least three paraphrasing verbs per noun-noun compound, if possible. The instructions we provided and the actual interface the human subjects were seeing are shown in Figures 1 and 2.



Paraphrasing Noun-Noun Compounds

Introduction

Given a noun-noun compound like *malaria mosquito*, *olive oil*, *grain alcohol*, *canola leaves*, *fruit fly*, *evening ride*, *neck vein*, *disease victim*, *migraine drug*, *Google ads*, *etc.*, you are asked to paraphrase it using verbs and prepositions.
For example, *neck vein* can be paraphrased as follows:

"*neck vein*" is a *vein* that comes from the *neck*
"*neck vein*" is a *vein* that drains the *neck*
"*neck vein*" is a *vein* that descends in the *neck*
"*neck vein*" is a *vein* that emerges from the *neck*
"*neck vein*" is a *vein* that enters the *neck*
"*neck vein*" is a *vein* that feeds the *neck*
"*neck vein*" is a *vein* that flows in the *neck*
"*neck vein*" is a *vein* that is in the *neck*
"*neck vein*" is a *vein* that is located in the *neck*
"*neck vein*" is a *vein* that is found in the *neck*
"*neck vein*" is a *vein* that is terminated at the *neck*
"*neck vein*" is a *vein* that nourishes the *neck*
"*neck vein*" is a *vein* that passes through the *neck*
"*neck vein*" is a *vein* that runs through the *neck*
"*neck vein*" is a *vein* that runs from the *neck*
"*neck vein*" is a *vein* that runs along the *neck*
"*neck vein*" is a *vein* that goes into the *neck*
"*neck vein*" is a *vein* that supplies the *neck*
"*neck vein*" is a *vein* that terminates in the *neck*
etc.

Fig. 1. Paraphrasing in Mechanical Turk: task introduction.

[3] `http://www.mturk.com`

We used *Amazon Mechanical Turk* Web service, which represents a cheap and easy way to recruit subjects for various tasks that require human intelligence. The service provides an API allowing a computer programme to ask a human to perform a task and returns the results. *Amazon* calls the process *Artificial Artificial Intelligence*. The idea behind the latter term and behind the origin of the service's name come from the *Mechanical Turk*, a life-sized wooden chess-playing mannequin the Hungarian nobleman Wolfgang von Kempelen constructed in 1769, which was able to defeat skilled opponents including Benjamin Franklin and Napoleon Bonaparte. The audience believed the automaton was making decisions using Artificial Intelligence, but the secret was a chess master hidden inside. Now *Amazon* provides a similar service to computer applications.

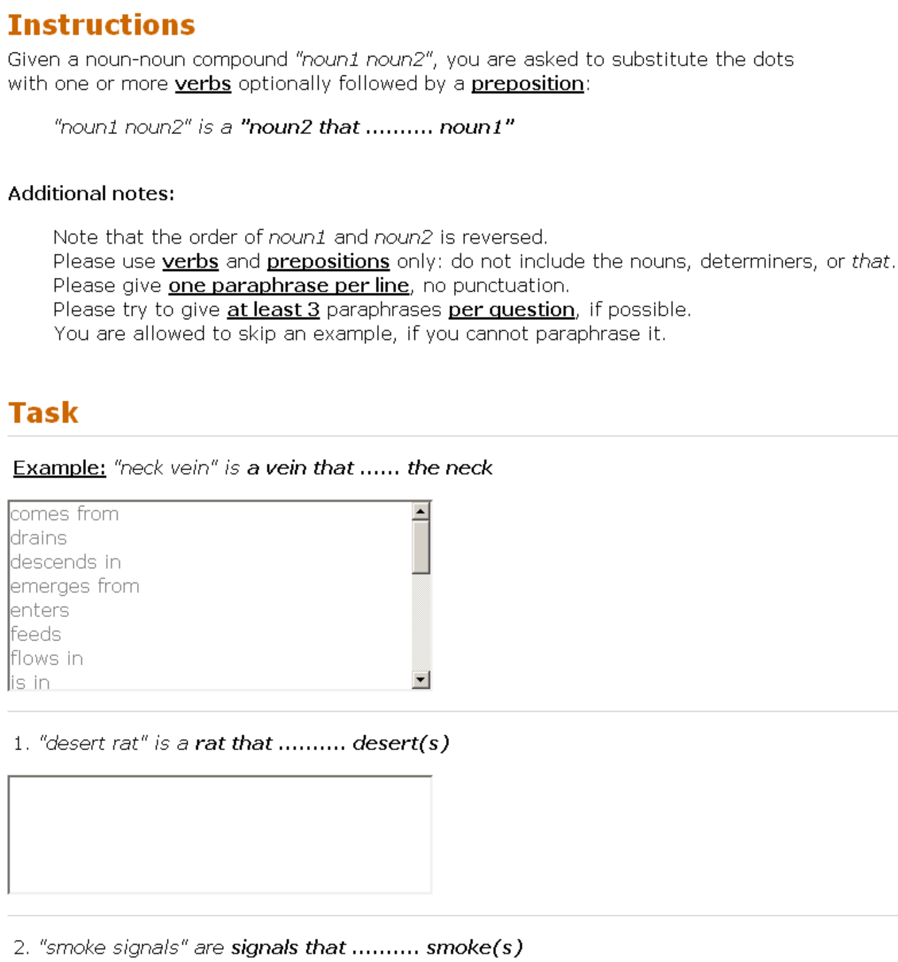


Fig. 2. Paraphrasing in Mechanical Turk: instructions, example, sample questions.

We used the 387 complex nominals Levi studied in her theory, listed in the appendix of [2]. We had to exclude the examples with an adjectival modifier, which are allowed in that theory, but do not represent noun compounds under our definition as was mentioned above. In addition, the following compounds were written concatenated and we decided to exclude them as well: *whistleberries*, *gunboat*, *silkworm*, *cellblock*, *snowball*, *meatballs*, *windmill*, *needlework*, *textbook*, *doghouse*, and *mothballs*. Some other examples contained a modifier that is a concatenation of two nouns, e.g., *wastebasket category*, *hairpin turn*, *headache pills*, *basketball season*, *testtube baby*; we decided to retain these examples. A similar example (which we chose to retain as well) is *beehive hairdo*, where both the modifier and the head are concatenations. As a result, we ended up with 250 good noun-noun compounds out of the original 387 complex nominals.

We randomly distributed these 250 noun-noun compounds (below, we will be referring to them as the *Levi-250 dataset*) into groups of 5, which yielded 50 Mechanical Turk tasks known as HITs (*Amazon* Human Intelligence Tasks), and we requested 25 different human subjects (*Amazon* workers) per HIT. We had to reject some of the submissions, which were empty or were not following the instructions, in which cases we requested additional workers in order to guarantee at least 25 good submissions per HIT. Each human subject was allowed to work on any number of HITs (between 1 and 50), but was not permitted to do the same HIT twice, which is controlled by the *Amazon Mechanical Turk* Web Service. A total of 174 different human subjects worked on the 50 HITs, producing 19,018 different verbs. After removing the empty and the bad submissions, and after normalising the verbs, we ended up with a total of 17,821 verbs, i.e., 71.28 verbs per noun-noun compound on average, not necessarily distinct.

Since many workers did not strictly follow the instructions, we performed some automatic cleaning of the results, followed by a manual check and correction, when it was necessary. First, some workers included the target nouns, the complementiser *that*, or determiners like *a* and *the*, in addition to the paraphrasing verb, in which cases we removed this extra material. For example, *star shape* was paraphrased as *shape that looks like a star* or as *looks like a* instead of just *looks like*. Second, the instructions required that a paraphrase be a sequence of one or more verb forms possibly followed by a preposition (complex prepositions like *because of* were allowed), but in many cases the proposed paraphrases contained words belonging to other parts of speech, e.g., nouns (*is in the shape of*, *has responsibilities of*, *has the role of*, *makes people have*, *is part of*, *makes use of*) or predicative adjectives (*are local to*, *is full of*); we filtered out all such paraphrases. In case a paraphrase contained an adverb, e.g., *occur only in*, *will eventually bring*, we removed the adverb and kept the paraphrase. Third, we normalised the verbal paraphrases by removing the leading modals (e.g., *can cause* becomes *cause*), perfect tense *have* and *had* (e.g., *have joined* becomes *joined*), or continuous tense *be* (e.g., *is donating* becomes *donates*). We converted complex verbal construction of the form '*<raising verb> to be*' (e.g., *appear to be*, *seems to be*, *turns to be*, *happens to be*, *is expected to be*) to just *be*. We further removed present participles introduced by *by*, e.g., *are caused* *by peeling* becomes *are caused*. Furthermore, we filtered out any paraphrase that involved *to* as part of the infinitive of a verb different from *be*, e.g., *is willing to donate* or *is painted to appear like* are not allowed. We also added *be* when it was missing in passive constructions, e.g., *made from* became *be made from*. Finally, we lemmatised the conjugated verb forms using *WordNet*, e.g., *comes from* becomes *come from*, and *is produced from* becomes *be produced from*. We also fixed some occasional spelling errors that we noticed, e.g., *bolongs to*, *happens becasue of*, *is mmade from*.

The resulting lexicon of human-proposed paraphrasing verbs with corresponding frequencies, and some other lexicons, e.g., a lexicon of the first verbs proposed by each worker only, and a lexicon of paraphrasing verbs automatically extracted from the Web as described in [12], are released under the *Creative Commons License*[4], and can be downloaded from the *Multiword Expressions Website*: `http://multiword.sf.net`. See [13] for additional details.

4 Experiments and Evaluation

We performed a number of experiments in order to assess both the quality of the created lexicon and the feasibility of the idea of using paraphrasing verbs to characterise noun compounds' semantics.

For each noun-noun compound from the *Levi-250 dataset*, we constructed two frequency vectors $\overrightarrow{h}$ (human) and $\overrightarrow{p}$ (programme). The former is composed of the above-described human-proposed verbs (after lemmatisation) and their corresponding frequencies, and the latter contains verbs and frequencies that were automatically extracted from the Web, as described in [12]. We then calculated the cosine correlation coefficient between $\overrightarrow{h}$ and $\overrightarrow{p}$ as follows:

$$cos(\overrightarrow{h}, \overrightarrow{p}) = \frac{\sum_{i=1}^{n} h_i p_i}{\sqrt{\sum_{i=1}^{n} h_i^2}\sqrt{\sum_{i=1}^{n} p_i^2}} \quad (1)$$

Table 3 shows human- and programme-proposed vectors for sample noun-noun compounds together with the corresponding cosine. The average cosine correlation (in %s) for all 250 noun-noun compounds is shown in Table 4. Since the workers were instructed to provide at least three paraphrasing verbs per noun-noun compound, and they tried to comply, some bad verbs were generated as a result. In such cases, the very first verb proposed by a worker for a given noun-noun compound is likely to be the best one. We tested this hypothesis by calculating the cosine using these first verbs only. As the last two columns of the table show, using all verbs produces consistently better cosine correlation, which suggests that there are many additional good human-generated verbs among those that follow the first one. However, the difference is 1-2% only and is not statistically significant.

[4] `http://creativecommons.org`

0.96 "blood donor" NOMINALIZATION:AGENT

Human: give(30), donate(16), supply(8), provide(6), share(2), contribute(1), volunteer(1), offer(1), choose(1), hand over(1), . . .

Progr.: give(653), donate(395), receive(74), sell(41), provide(39), supply(17), be(13), match(11), contribute(10), offer(9), . . .

0.93 "city wall" $HAVE_2$

Human: surround(24), protect(10), enclose(8), encircle(7), encompass(3), be in(3), contain(2), snake around(1), border(1), go around(1), . . .

Progr.: surround(708), encircle(203), protect(191), divide(176), enclose(72), separate(49), ring(41), be(34), encompass(25), defend(25), . . .

0.91 "disease germ" $CAUSE_1$

Human: cause(20), spread(5), carry(4), create(4), produce(3), generate(3), start(2), promote(2), lead to(2), result in(2), . . .

Progr.: cause(919), produce(63), spread(37), carry(20), propagate(9), create(7), transmit(7), be(7), bring(5), give(4), . . .

0.89 "flu virus" $CAUSE_1$

Human: cause(19), spread(4), give(4), result in(3), create(3), infect with(3), contain(3), be(2), carry(2), induce(1), . . .

Progr.: cause(906), produce(21), give(20), differentiate(17), be(16), have(13), include(11), spread(7), mimic(7), trigger(6), . . .

0.89 "gas stove" USE

Human: use(20), run on(9), burn(8), cook with(6), utilize(4), emit(3), be heated by(2), need(2), consume(2), work with(2), . . .

Progr.: use(98), run on(36), burn(33), be(25), be heated by(10), work with(7), be used with(7), leak(6), need(6), consume(6), . . .

0.89 "collie dog" BE

Human: be(12), look like(8), resemble(2), come from(2), belong to(2), be related to(2), be called(2), be classified as(2), be made from(1), be named(1), . . .

Progr.: be(24), look like(14), resemble(8), be border(5), feature(3), come from(2), tend(2), be bearded(1), include(1), betoken(1), . . .

0.87 "music box" $MAKE_1$

Human: play(19), make(12), produce(10), emit(5), create(4), contain(4), provide(2), generate(2), give off(2), include(1), . . .

Progr.: play(104), make(34), produce(18), have(16), provide(14), be(13), contain(9), access(8), say(7), store(6), . . .

0.87 "cooking utensils" FOR

Human: be used for(17), be used in(9), facilitate(4), help(3), aid(3), be required for(2), be used during(2), be found in(2), be utilized in(2), involve(2), . . .

Progr.: be used for(43), be used in(11), make(6), be suited for(5), replace(3), be used during(2), facilitate(2), turn(2), keep(2), be for(1), . . .

Table 3. Human- and programme-proposed vectors, and cosines for sample noun-noun compounds. The common verbs for each vector pair are underlined.

Min # of Web Verbs	Number of Compounds	Correlation with Humans Using All Verbs	First Verb Only
0	250	31.8%	30.6%
1	236	33.7%	32.4%
3	216	35.4%	34.1%
5	203	36.9%	35.6%
10	175	37.3%	35.5%

Table 4. Average cosine correlation (in %s) between human- and programme-generated verbs for the *Levi-250 dataset.* Shown are the results for different limits on the minimum number of programme-generated Web verbs. The last column shows the cosine when only the first verb proposed by each worker is used.

A limitation of the Web-based verb-generating method is that it could not provide paraphrasing verbs for 14 of the noun-noun compounds, in which cases the cosine was zero. If the calculation was performed for the remaining 236 compounds only, the cosine increased by 2%. Table 4 shows the results when the cosine calculations are limited to compounds with at least 1, 3, 5 or 10 different verbs. We can see that the correlation increases with the minimum number of required verbs, which means that the extracted verbs are generally good, and part of the low cosines are due to an insufficient number of extracted verbs. Overall, all cosines in Table 4 are in the 30-37%, which corresponds to a medium correlation [14].

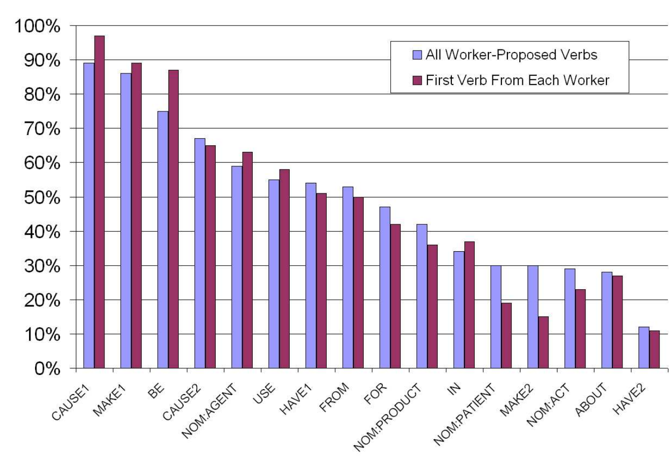


Fig. 3. Cosine correlation (in %s) between the human- and the programme-generated verbs from the *Levi-250 dataset* aggregated by relation: using all human-proposed verbs vs. only the first verb from each worker.

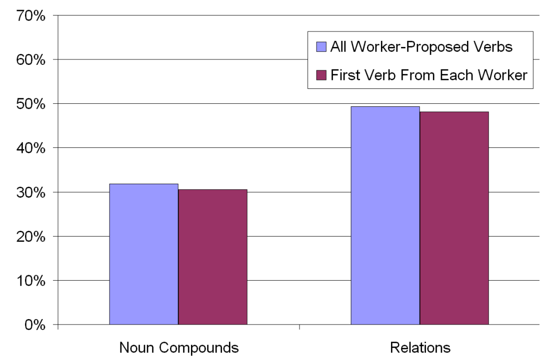


Fig. 4. Average cosine correlation (in %s) between the human- and the programme-generated verbs for the *Levi-250 dataset* calculated for each noun compound (left) and aggregated by relation (right): using all human-proposed verbs vs. only the first verb from each worker.

We further compared the human- and the programme-generated verbs aggregated by relation. Given a relation like $\mathtt{HAVE}_1$, we collected all verbs belonging to noun-noun compounds from that relation together with their frequencies. From a vector-space model point of view, we summed their corresponding frequency vectors. We did this separately for the human- and the programme-generated verbs, and we then compared the corresponding pairs of summed vectors separately for each relation.

Figure 3 shows the cosine correlations for each of the 16 relations using all human-proposed verbs and only the first verb from each worker. We can see a very-high correlation (mid-70% to mid-90%) for relations like $\mathtt{CAUSE}_1$, $\mathtt{MAKE}_1$, `BE`, but low correlation 11-30% for reverse relations like $\mathtt{HAVE}_2$ and $\mathtt{MAKE}_2$, and for most nominalisations (except for `NOM:AGENT`). Interestingly, using only the first verb improves the results for highly-correlated relations, but damages low-correlated ones. This suggests that when a relation is more homogeneous, the first verbs proposed by the workers are good enough, and the following verbs only introduce noise. However, when the relation is more heterogeneous, the extra verbs are more likely to be useful. As Figure 4 shows, overall the average cosine correlation is slightly higher when all worker-proposed verbs are used vs. the first verb from each worker only: this is true both when comparing the individual noun-noun compounds and when the comparison is performed for the 16 relations. The figure also shows that while the cosine correlation for individual noun-noun compounds is in the low-30%, for relations it is almost 50%.

Finally, we tested whether the paraphrasing verbs are good features to use in a nearest-neighbour classifier. Given a noun-noun compound, we used the human-proposed verbs as features to predict Levi's RDP for that compound. In this experiment, we only used those noun-noun compounds which are not nominalisations, i.e., for which Levi has an RDP provided; this left us with 214 examples (*Levi-214 dataset*) and 12 classes. We performed leave-one-out

Model	**Accuracy**	**Coverage**	**Avg #feats**	**Avg Σfeats**
Human: all v	78.4±6.0	99.5	34.3	70.9
Human: first v from each worker	72.3±6.4	99.5	11.6	25.5
Web: $v+p+c$	50.0±6.7	99.1	216.6	1716.0
Web: $v+p$	50.0±6.7	99.1	208.9	1427.9
Web: $v+c$	46.7±6.6	99.1	187.8	1107.2
Web: v	45.8±6.6	99.1	180.0	819.1
Web: p	33.0±6.0	99.1	28.9	608.8
Web: $p+c$	32.1±5.9	99.1	36.6	896.9
Baseline (majority class)	19.6±4.8	100.0	–	–

Table 5. Predicting Levi's RDP on the *Levi-214 dataset* using verbs v, prepositions p, and coordinating conjunctions c as features: leave-one-out cross-validation. Shown are micro-averaged accuracy and coverage in %s, followed by average number of features and average sum of feature frequencies per example.

cross-validation experiments with a 1-nearest-neighbor classifier (using TF.IDF-weighting and the Dice coefficient[5] as a similarity measure, as in [15]), trying to predict the correct RDP for the testing example. The results are shown in Table 5. We achieved 78.4% accuracy using all verbs, and 72.3% with the first verb from each worker. This result is very strong for a 12-way classification problem, and supports the hypothesis that the paraphrasing verbs are very important features for the task of noun-noun compound interpretation.

Table 5 also shows the results when verbs, prepositions and coordinating conjunctions automatically extracted from the Web are used as features. As we can see, using prepositions alone only yields about 33% accuracy, which is a statistically significant improvement over the majority-class baseline, but is well below the classifier performance when using verbs. Overall, the most important Web-derived features are the verbs: they yield 45.8% accuracy when used alone, and 50% when used together with prepositions. Adding coordinating conjunctions helps a bit with verbs, but not with prepositions. Note however that none of the differences between the different feature combinations involving verbs are statistically significant. However, the difference between using Web-derived verbs and using human-proposed verbs (78.4% vs. 50%) is very statistically significant, and suggests that the human-proposed verbs could be considered an upper bound on the accuracy that could be achieved with automatically extracted features.

Table 5 also shows the average number of distinct features and the sum of feature counts per example. As we can see, for Web-derived features, there is a strong positive correlation between number of extracted features and classification accuracy, the best result being achieved with more than 200 features per example. Note however, that using human-proposed verbs yields very high accuracy with seven times less features on average.

[5] Given two TF.IDF-weighted frequency vectors A and B, we compare them using the following generalised Dice coefficient: $Dice(A,B) = \frac{2\times\sum_{i=1}^{n}\min(a_i,b_i)}{\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}b_i}$

5 Discussion

Interpreting noun compounds in terms of sets of fine-grained verbs that are directly usable in paraphrases of the target noun-noun compounds can be useful for a number of NLP tasks, e.g., for noun compound translation in isolation [3, 16, 17], for paraphrase-augmented machine translation [18–21], for machine translation evaluation [22, 23], for summarisation evaluation [24], etc.

As we have shown above (see [11, 12, 15] for additional details and discussion on our experiments), assuming annotated training data, the paraphrasing verbs can be used as features to predict abstract relations like `CAUSE`, `USE`, `MAKE`, etc. Such coarse-grained relations can in turn be helpful for other applications, e.g., for recognising textual entailment as shown by Tatu&Moldovan [25]. Note however, that, for this task, it is possible to use our noun compound paraphrasing verbs directly as explained in Appendix B of [11].

In information retrieval, the paraphrasing verbs can be used for index normalisation [26], query expansion, query refinement, results re-ranking, etc. For example, when querying for *migraine treatment*, pages containing good paraphrasing verbs like *relieve* or *prevent* could be preferred.

In data mining, the paraphrasing verbs can be used to seed a Web search that looks for particular classes of NPs such as diseases, drugs, etc. For example, after having found that *prevent* is a good paraphrasing verb for *migraine treatment*, we can use the query[6] `"* which prevents migraines"` to obtain different treatments/drugs for migraine, e.g., *feverfew*, *Topamax*, *natural treatment*, *magnesium*, *Botox*, *Glucosamine*, etc. Using a different paraphrasing verb, e.g., using `"* reduces migraine"` can produce additional results: *lamotrigine*, *PFO closure*, *Butterbur Root*, *Clopidogrel*, *topamax*, *anticonvulsant*, *valproate*, *closure of patent foramen ovale*, *Fibromyalgia topamax*, *plant root extract*, *Petadolex*, *Antiepileptic Drug Keppra (Levetiracetam)*, *feverfew*, *Propranolol*, etc. This is similar to the idea of a relational Web search of Cafarella&al. [27], whose system TextRunner serves four types of relational queries, among which there is one asking for all entities that are in a particular relation with a given target entity, e.g., "*find all X such that X prevents migraines*".

6 Conclusion and Future Work

In this paper, we explored and experimentally tested the idea that, in general, the semantics of a given noun-noun compound can be characterised by the set of all possible paraphrasing verbs that can connect the target nouns, with associated weights. The verbs we used were fine-grained, directly usable in paraphrases, and using multiple of them for a given noun-noun compound allowed for better approximating its semantics.

Using Amazon's *Mechanical Turk*, we created a new resource for noun-noun compound interpretation based on paraphrasing verbs, and we demonstrated

[6] Here "*" is the Google star operator, which can substitute one or more words. In fact, it is not really needed in this particular case.

experimentally that verbs are especially useful features for predicting abstract relations like Levi's RDPs. We have already made the resource publicly available [13]; we hope that by doing so, we will inspire further research in the direction of paraphrase-based noun compound interpretation, which opens the door to practical applications in a number of NLP tasks including but not limited to machine translation, text summarisation, question answering, information retrieval, textual entailment, relational similarity, etc.

The present situation with noun compound interpretation is similar to that with word sense disambiguation: in both cases, there is a general agreement that the research is important and much needed, there is a growing interest in performing further research, and a number of competitions are being organised, e.g., as part of SemEval [28]. Still, there are very few applications of noun compound interpretation in real NLP tasks (e.g., [19] and [25]). We think that increasing this number is key for the advancement of the field, and we believe that turning to paraphrasing verbs could help bridge the gap between research interest and practical applicability for noun compound interpretation.

Acknowledgments. This research was supported in part by the FP7 project AsIsKnown, by FP7-REGPOT-2007-1 SISTER and by NSF DBI-0317510.

References

1. Downing, P.: On the creation and use of English compound nouns. Language (53) (1977) 810–842
2. Levi, J.: The Syntax and Semantics of Complex Nominals. Academic Press, New York (1978)
3. Baldwin, T., Tanaka, T.: Translation by machine of compound nominals: Getting it right. In: Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing. (2004) 24–31
4. Warren, B.: Semantic patterns of noun-noun compounds. In: Gothenburg Studies in English 41, Goteburg, Acta Universtatis Gothoburgensis. (1978)
5. Nastase, V., Szpakowicz, S.: Exploring noun-modifier semantic relations. In: Fifth International Workshop on Computational Semantics (IWCS-5), Tilburg, The Netherlands (2003) 285–301
6. Girju, R., Moldovan, D., Tatu, M., Antohe, D.: On the semantics of noun compounds. Journal of Computer Speech and Language - Special Issue on Multiword Expressions **4**(19) (2005) 479–496
7. Rosario, B., Hearst, M.: Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. In: Proceedings of EMNLP. (2001) 82–90
8. Lauer, M.: Designing Statistical Language Learners: Experiments on Noun Compounds. PhD thesis, Dept. of Computing, Macquarie University, Australia (1995)
9. Finin, T.: The Semantic Interpretation of Compound Nominals. PhD thesis, University of Illinois, Urbana, Illinois (1980)
10. Nakov, P., Hearst, M.: Using verbs to characterize noun-noun relations. In: AIMSA. Volume 4183 of LNCS., Springer (2006) 233–244
11. Nakov, P.: Using the Web as an Implicit Training Set: Application to Noun Compound Syntax and Semantics. PhD thesis, EECS Department, University of California, Berkeley, UCB/EECS-2007-173 (2007)

12. Nakov, P., Hearst, M.: Solving relational similarity problems using the web as a corpus. In: Proceedings of ACL'08: HLT, Columbus, OH (2008)
13. Nakov, P.: Paraphrasing verbs for noun compound interpretation. In: Proceedings of the LREC'08 Workshop: Towards a Shared Task for Multiword Expressions (MWE'08), Marrakech, Morocco (2008)
14. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. 2 edn. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ (1988)
15. Nakov, P., Hearst, M.: UCB: System description for SemEval Task #4. In: Proceedings of SemEval, Prague, Czech Republic (2007) 366–369
16. Grefenstette, G.: The World Wide Web as a resource for example-based machine translation tasks. In: Translating and the Computer 21: Proceedings of the 21st International Conference on Translating and the Computer. (1999)
17. Tanaka, T., Baldwin, T.: Noun-noun compound machine translation: a feasibility study on shallow processing. In: Proceedings of the ACL 2003 workshop on Multiword expressions. (2003) 17–24
18. Callison-Burch, C., Koehn, P., Osborne, M.: Improved statistical machine translation using paraphrases. In: Proceedings of Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. (2006) 17–24
19. Nakov, P.: Improved statistical machine translation using monolingual paraphrases. In: Proceedings of the European Conference on Artificial Intelligence (ECAI'08), Patras, Greece (2008)
20. Nakov, P., Hearst, M.: UCB system description for the WMT 2007 shared task. In: Proceedings of the Second Workshop on Statistical Machine Translation, Prague, Czech Republic (June 2007) 212–215
21. Nakov, P.: Improving English-Spanish statistical machine translation: Experiments in domain adaptation, sentence paraphrasing, tokenization, and recasing. In: Proceedings of the Third Workshop on Statistical Machine Translation (WMT), Columbus, OH (2008)
22. Russo-Lassner, G., Lin, J., Resnik, P.: A paraphrase-based approach to machine translation evaluation. Technical Report LAMP-TR-125/CS-TR-4754/UMIACS-TR-2005-57, University of Maryland (August 2005)
23. Kauchak, D., Barzilay, R.: Paraphrasing for automatic evaluation. In: Proceedings Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. (2006) 455–462
24. Zhou, L., Lin, C.Y., Hovy, E.: Re-evaluating machine translation results with paraphrase support. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia (July 2006) 77–84
25. Tatu, M., Moldovan, D.: A semantic approach to recognizing textual entailment. In: Proceedings of HLT. (2005) 371–378
26. Zhai, C.: Fast statistical parsing of noun phrases for document indexing. In: Proceedings of the fifth conference on Applied natural language processing, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1997) 312–319
27. Cafarella, M., Banko, M., Etzioni, O.: Relational Web search. Technical Report 2006-04-02, University of Washington, Department of Computer Science and Engineering (2006)
28. Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., Yuret, D.: Semeval-2007 task 04: Classification of semantic relations between nominals. In: Proceedings of SemEval, Prague, Czech Republic (2007) 13–18