Uncovering Noun-Noun Compound Relations by Gamification

Johan Bos
Center for Language and Cognition
Oude Kijk in 't Jatstraat 26
University of Groningen
johan.bos@rug.nl

Malvina Nissim
Center for Language and Cognition
Oude Kijk in 't Jatstraat 26
University of Groningen
m.nissim@rug.nl

Abstract

Can relations described by English noun-noun compounds be adequately captured by prepositions? We attempt to answer this question in a data-driven way, using gamification to annotate a set of about a thousand noun-noun compound examples. Annotators could make a choice out of five prepositions generated with the help of paraphrases found in the Google n-gram corpus. We show that there is substantial agreement among the players of our linguistic annotation game, and that their answers differ in about 50% of raw frequency counts of the Google n-gram corpus. Prepositions can be used to describe the majority of the implicit relations present in noun-noun compounds, but not all relations are captured by natural prepositions and some compounds are not easy to paraphrase with the use of a preposition.

1 Introduction

English noun-noun compounds express a relation between the two nouns involved, but this relation isn't made linguistically explicit. So we can have *war crime* meaning a crime **in** a war, or *safety violations* meaning violations **of** safety, or *security guarantees*, meaning guarantees **for** security. In short: the relation between two nouns in a compound expression isn't specified and can take many different roles. This situation introduces an interesting problem for meaning interpretation: what semantic relation is expressed in a noun-noun compound?

There are mainly three different approaches that deal with this problem. The first family of approaches take a (usually small) fixed inventory of relations and use it to describe compounds based on well-established ontologies. The second line of research takes a set of English prepositions to describe compounds (in a way similar as we did above). This makes sense, as prepositions naturally describe a relation between two entities. The seminal work following this tradition is Lauer (Lauer, 1995), who, inspired by Levi's work on fixing a set of possible predicates for interpreting noun-noun compounds (Levi, 1978), developed an inventory comprising eight different prepositions: *of*, *for*, *with*, *in*, *on*, *at*, *about*, and *from*. The third set of attempts views compound interpretation as a paraphrasing task (Nakov, 2007). This would yield interpretations such as "a crime committed during a war" for our earlier example *war crime*.

None of the three approaches show clear advantages. On the one side of the spectrum, the fixed-vocabulary-approach faces the problem of being too strict. On the other end of it, paraphrasing is hard to control. Attempts at combining more than one approach for English (Girju, 2009) or German (Dima et al., 2014) still rely heavily on pre-constructed sets of relations/prepositions, the latter advocating a hybrid approach combining a semantic-relation and preposition-based method.

Given that the preposition-approach lies somewhere between these other two approaches, and can be taken in such a way that is entirely data driven, this is the approach that we will consider and use in this paper. While we are aware of its expressive limitations (prepositions might not be sufficient, and they might preserve some ambiguity of the compound), we still think it is interesting to test to what extent it can be carried out in (i) a completely data-driven fashion and (ii) using judgments by multiple speakers without linguistic training, thus making it extremely inexpensive and light, yet useful. To comply with (i), we make sure that prepositions are not derived from a fixed pre-compiled list, but rather acquired automatically, *case by case*, exploiting Google's n-grams to generate candidates. The compounds themselves are

taken from an existing semantically annotated corpus, the Groningen Meaning Bank (Basile et al., 2012). Regarding (ii), we exploit crowd-sourcing and develop a game-with-a-purpose setting to collect data. The acquired data can then be analysed to investigate more closely the use of prepositions for interpreting noun noun compounds and the extent to which different people agree. Moreover, the data can be used to collect descriptive statistics on preposition use in this context that might give new insight into this approach.

2 Method

In this section we describe how we selected noun-noun compounds from a corpus (Step 1), generated potential prepositional relations for each compound (Step 2), and then manually annotated the preposition resembling the underlying meaning relation (Step 3). In what follows we will describe each step in further detail.

The first step is pretty straightforward and makes use of an existing parsed corpus of English texts, and simply looks for a sequence of exactly two nouns (i.e., the words before and after are not tagged as nouns). This excludes compounds comprising three of more nouns but this would only complicate the task (dealing with issues such as internal bracketing) and therefore this limitation allows us to put more focus on our key objectives. On a more detailed note, we take sequences that are tagged NN NN or NN NNS, as English grammar restricts the first noun to be of singular case.

The aim of the second step is to find a set of most likely prepositions that can be used to describe a noun-noun compound expression. This process is carried out with the aid of the Google n-gram corpus. Our starting point are 26 common English prepositions (this is considered to be a closed set, disregarding compound prepositions):

> *of, for, in, on, with, from, by, at, through, into, about, after, between, per, against, over, under, without, before, within, among, via, across, towards, toward*, and *around*.

Next, given a pair N_1–N_2 extracted from the corpus in Step 1, we compute the frequencies of the 4-gram N_2(s)–PREPOSITION–ARTICLE–N_1(s) in the four different singular/plural formations. We use MORPHA and MORPHG to generate all inflected forms of the nouns (Minnen et al., 2001). The articles that we insert in the 4-gram are *a*, *an* and *the*. For instance, the compound *expansion plan* would generate the following 4-gram patterns:

> plan of a expansion
> plan of an expansion
> plan of the expansion
> plans of a expansion
> ...
> plans for an expansions
> plans for the expansions

In case the number of resulting instances was lower than five, the empty places were filled up with the most frequently used prepositions overall computed for all compounds extracted from the Google n-gram corpus. These were: *of, from, on, for* and *by*. The total for a preposition given a compound is the sum of all frequencies obtained for each single query.

The third step is using the data generated in Step 2 in a GWAP, a *game with a purpose*, in order to collect human judgements. Wordrobe (Venhuizen et al., 2013), an existing internet-based GWAP architecture was used to launch a noun-noun compound annotation task in the shape of a game named *burgers* at `www.wordrobe.org`. Players of this game, not necessarily knowing anything about linguistic annotation, received a snippet of a text with the relevant noun-noun compound marked up in bold face, and were asked to select the most appropriate prepositions of the five candidates generated in Step 2. They were awarded points relative to the agreement of other players' choices for the same question (using add-1 smoothing initially). Players were instructed to hit the *skip* button in case none of the choices seemed to make sense.

A total of 1,296 game questions were generated on March 7, 2013 and released to the GWAP. We did not actively solliciated players, but instead relied solely on regular Wordrobe players or new players that found the game via social media or web links. This way, we gathered a total of 5,368 responses by 187 different GWAP players in the period between release and now (January 26, 2015).

3 Results

The number of annotations in our dataset is 5,195, for a total of 965 different compounds. This yields an average number of 5.4 annotations per com-

pound (min=1, max=138). Most examples had between one and six GWAP players.

A small number of examples were *skipped* by the GWAP players (see previous section): 170 times, for a total of 75 different noun-noun compouds. In most cases these were ill-formed expressions caused by POS-tagging mistakes. Consider for instance the following compounds that were skipped by more than five different players: *capital city, attack north, camp north, c-130 aircraft*, and *accident north*. Except for *c-130 aircraft*, a name-noun compound, these are all mistakenly parsed as noun-noun compounds. This shows that the *skip* function in our annotation game does its job.

To get an idea of the effect of gamification, we took the 100 most frequently answered GWAP questions for further investigation. Within this set, we found that 51 times a preposition formed the majority class that was different from the most frequent preposition in the n-gram corpus found for the corresponding 4-gram patterns (see previous section). This indicates that the GWAP makes a real difference in choice of preposition for a compound.

Prep.	#selected	#majority	Example
about	46	8	security concerns(12)
across	7		border police(2)
after	3		capital city(2)
against	18	2	missile shield(11)
among	56		bird flu(53)
around	12	2	capital city(2)
at	122	19	border checkpoint(19)
before	8		bird flu(5)
between	6	1	government lines(2)
by	143	25	bomb attack(12)
for	1279	248	news agency(65)
from	296	31	bird flu(62)
in	592	65	car bomb(87)
into	17	2	cell research(5)
of	1879	344	death toll(62)
on	308	34	roadside bomb(37)
over	28	3	radio address(10)
per	12	2	capita income(9)
through	13	1	export trade(2)
toward	2		peace process(2)
towards	9		peace process(6)
under	21	1	car bomb(12)
via	7	2	audio messages(4)
with	300	44	bomb attack(26)
within	11		war crimes(2)
without	0		

Table 1: Choice of Prepositions by GWAP players.

In the whole dataset, 25 different prepositions were chosen by GWAP players, but obviously not all were used equally frequently. Its distribution is shown in Table 1. The second column in this table shows the total number of times a given preposition was chosen by a GWAP player. The third column shows the number of times the preposition had the majority of votes. The example in the fourth column is the one where the preposition was chosen in its highest score.

Perhaps unsurprisingly, *of* was picked most frequently. The least common prepositions selected by GWAP players were *across* (7), *between* (6), *after* (3), and *toward* (2). Perhaps this is because these prepositions express quite complex spatial or temporal relations. What Table 1 also shows is the number of times a preposition formed a majority class for a certain noun-noun compound. Relative majority has proven to be a simple but effective method for selected gold-standard values for word sense disambiguation in a GWAP setting (Venhuizen et al., 2013).

Recall that the GWAP players could select one preposition out of a set of five (extracted as described in Section 2). In the large majority of cases, either one (368 compounds) or two (374 compounds) prepositions were chosen. Three different ones were selected in 156 cases, four in 62, and five in 5 cases. Overall, we think this agreement is encouraging.

4 Discussion

It is hard to quantify the results that we obtained in terms of annotator accuracy. But taking a closer look at the results reveals some interesting and promising patterns. First of all, we show some examples of compounds that had unanimous decisions among various annotators (Table 2). Even relatively non-frequent prepositions like *against* were selected in complete agreement by the GWAP players.

Compound	Preposition	# Players
government forces	of	16
agriculture development	of	12
missile shield	against	11
agency chief	of	11
rescue teams	for	9

Table 2: Compounds with unanimous decisions.

Examples of compounds with only two different prepositions chosen by the players are shown in Table 3. In the top part of the table we report cases where one preposition is nevertheless dominant, while in the bottom part more difficult, ambiguous cases can be found.

Compound	prep1(#)	prep2(#)	prep3(#)	prep4(#)	prep5(#)	selected(%)	**Total**
bird flu	among(53)	before(5)	from(62)	in(16)	against(2)	from(0.45)	138
chemical company	for(5)	from(1)	in(3)	of(3)	within(1)	for(0.38)	13
death toll	in(2)	of(62)	on(6)	for(3)	with(1)	of(0.84)	74
defense official	from(3)	of(7)	at(2)	in(1)	with(1)	of(0.5)	14
background checks	for(1)	in(1)	into(1)	of(2)	on(1)	of(0.33)	6

Table 4: Cases where GWAP players picked at least one of each possible preposition candidate.

Compound	prep1	prep2	selected	**# Players**
roadside bomb	on(37)	in(2)	on(0.95)	39
assassination plan	for(16)	with(1)	for(0.94)	17
basketball game	in(1)	of(16)	of(0.94)	17
security concerns	about(12)	over(3)	about(0.8)	15
missile strike	of(7)	with(12)	with(0.63)	19
air strike	by(8)	from(9)	from(0.53)	17
army uniform	for(6)	of(7)	of(0.54)	13
army prison	for(5)	of(7)	of(0.58)	12
cell research	into(5)	on(6)	on(0.54)	11

Table 3: Compounds with two different choices.

The examples shown in this table show that in various cases more than one preposition seems accurate: *bomb **on** the roadside* or *bomb **in** the roadside* both express a spatial relation, and both prepositions would probably be accursate. Similarly, *concerns **about** security* or *concerns **over** security* seem both appropriate paraphrases of the compound. It also illustrates the fact that the more underspecified *of* is often chosen together with other, more specific prepositions. This shows both the advantage and disadvantage of using prepositions as relations: like other words, prepositions are ambiguous, and there is substantial overlap in meaning between the prepositions one finds in the English language. This makes them more flexible, but also less formal.

The data also supports the observation that prepositions that carry some logical meaning (such as negation) are unsuitable to describe relations between two entities found in noun-noun compounds. A clear case is *without*, that was never selected by a GWAP player (Table 1), whose lexical meaning could be paraphrased as *not with*. As negation is not related to any general kind of concept, a noun-noun compound would never be able to catch this aspect of meaning. Similarly, *after* and *before* carry some negation in their lexical meaning (as in *not at the same time*), and again the data supports this as they never formed the majority class.

Finally, in Table 4 we list those cases where all GWAP players selected at least one of each possible answers. Such a situation could be evidence that the relational meaning of a compound is hard to catch by a preposition. Certainly, the more players answered a specific question, the higher the chance that all possible candidates were selected at least once (for instance, taking into account the fact that players make mistakes). This seems to be the case for the first example in Table 4 where *before* and *against* are odd choices, but they are clearly outperformed by *among* and *from* that both seem adequate choices.

5 Conclusion

We showed that a data-driven approach to finding prepositions describing noun-noun compound relations is feasible. Simple raw frequencies of prepositional paraphrases aren't likely to get useful results. We demonstrated that a game with a purpose yields good results to find appropriate prepositions for this task. The results will be used to improve the Groningen Meaning Bank, a large corpus of semantically-annotated texts (Basile et al., 2012).

Compared to Lauer, we opted for a more data-driven choice of prepositions, rather than restricting ourselves to Lauer's set of eight prepositions. None of these prepositions would fit the compound *missile shield* but in our approach *against* would be selected as relation (see Table 2). We clearly benefit from such cases.

In future work it would be worthy to try to map prepositions to unambiguous relations, or attempt to group prepositions that bear similar meanings. One interesting way is to look at answer patterns in the data to disambiguate the very general preposition *of*, by taking into account other answers as well instead of just considering the majority class. Similarly, it would be interesting to see if one can predict idiosyncratic compounds such as *suicide bomber*, whose implicit relation is hard to catch by a preposition.

References

Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. A platform for collaborative semantic annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 92–96, Avignon, France.

Corina Dima, Verena Henrich, Erhard Hinrichs, and Christina Hoppermann. 2014. How to Tell a Schneemann from a Milchmann: An Annotation Scheme for Compound-Internal Relations. In Calzolari et al., editor, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Roxana Girju. 2009. The syntax and semantics of prepositions in the task of automatic interpretation of nominal phrases and compounds: A cross-linguistic study. *Computational Linguistics*, 35(2):185–228.

Mark Lauer. 1995. Corpus statistics meet the noun compound: Some empirical results. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 47–54, Cambridge, Massachusetts, USA, June. Association for Computational Linguistics.

Judith Levi. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press.

Guido Minnen, John Carroll, and Darren Pearce. 2001. Applied morphological processing of English. *Journal of Natural Language Engineering*, 7(3):207–223.

Preslav Ivanov Nakov. 2007. *Using the Web as an Implicit Training Set: Application to Noun Compound Syntax and Semantics*. Ph.D. thesis, University of California, Berkeley.

Noortje Venhuizen, Valerio Basile, Kilian Evang, and Johan Bos. 2013. Gamification for Word Sense Labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*, pages 397–403, Potsdam, Germany.