A Hybrid Approach for the Interpretation of Nominal Compounds using Ontology

Sruti Rallapalli, Soma Paul
Language Technologies Research Center
International Institute of Information Technology
Hyderabad
`sruti@students.iiit.ac.in`
`soma@iiit.ac.in`

Abstract

Understanding and interpretation of nominal compounds has been a long-standing area of interest in NLP research for various reasons. (1) Nominal compounds occur frequently in most languages. (2) Compounding is an extremely productive word formation phenomenon. (3) Compounds contain implicit semantic relations between their constituent nouns. Most approaches that have been proposed so far concentrate on building statistical models using machine learning techniques and rely on large-scale, domain-specific or open- domain knowledge bases. In this paper we present a novel approach that combines the use of lexical hierarchies such as PurposeNet and WordNet, with WordNet-based similarity measures for the interpretation of domain-specific nominal compounds. We aim at building a robust system that can handle most of the commonly occurring English bigram nominal compounds within the domain.

1 Introduction

Understanding and interpretation of nominal compounds has been a long-standing area of interest in NLP research. The main reasons that make understanding compound nouns an interesting and challenging task are: (1) Compound nouns are a frequent phenomenon in many languages, occurring in different languages with varying frequencies. English and Sanskrit are two languages that display great flexibility in compounding (Ó Séaghdha, 2008). About 3.9 % of the words in Reuters are bigram nominal compounds (Baldwin and Tanaka, 2004). (2) Compounding is a recursive process that can lead to formation of large and complex compounds, that are difficult for comprehension. (3) Compounds usually carry an implicit meaning that may sometimes differ significantly from that of the combining concepts. Consider the example of a *garden knife*. A *garden knife* is interpreted as a knife used in the garden. Here the modifier *garden* modifies the locative information of the head noun *knife*. Alternatively, consider the example of a *gamma knife*. A *gamma knife* is a device used to treat brain tumors by administering gamma radiations in a particular manner. Here, the modifier does not necessarily modify the head, instead they both combine together to denote a different concept. This understanding of the difference in the structure and purpose between *gamma knife* and other kinds of knife cannot be achieved by any means of statistical predictions or morphological and syntactic analyses of the compound.

The most common representations adopted for the interpretation of nominal compounds involve an inventory of verbs, prepositions or abstract semantic relations. Verb and preposition paraphrases bring in lexical ambiguity in the interpretation of the compound, essentially owing to the polysemous nature of verbs and prepositions. For example, *morning tea* and *bar lights* would both be paraphrased using the preposition *in*. However, the paraphrase 'tea in the morning' conveys the *temporal* aspect of the compound, while the paraphrase 'lights in the bar' describes the *location* information in the compound. Due to this polysemous behaviour of prepositions and verbs, a restricted inventory of abstract semantic

relations is more favorable for the compound interpretation problem.

Most of the approaches proposed for interpreting nominal compounds fall into one of the two classes (a) supervised machine learning approaches, and (b) unsupervised data driven approaches. These approaches fail to handle the sparseness of data, which is a major issue in case of noun compounds. They collect statistics that use occurrence frequencies of the compounds. Therefore, rarely occurring compounds lead to wrong estimations of probabilities and thereby unreliable interpretations. A third and less frequently adopted alternative involves the use of large-scale, domain-independent, lexical and conceptual hierarchies that provide detailed natural language semantics. Such ontologies promise reliability and accuracy of data but fail to cover equally, lexical items and semantic relations. Moreover, construction of such ontologies is extremely time-consuming, due to which manually built ontologies are never up to date with changes in the language. This motivates us to argue that the most optimal approach to compound interpretation would be the combination of a lexical hierarchy for the frequent and idiosyncratic compounds (Johnston and Busa, 1996) and WordNet-based similarity for those that are not listed in the hierarchy. We show in this paper, that adopting our hybrid approach helps us achieve significant results (70% accuracy) in ontology-based compound interpretation, irrespective of the size, coverage and domain of the ontology. We perform all our experiments using PurposeNet (KiranMayee et al., 2008), which is a purpose-centric ontology of artifacts and semantic relations.

The rest of the paper is divided into the following sections. In section 2, we discuss some related works that use ontologies and also motivate our choice of a hybrid approach using ontology. In section 3, we discuss the architectural design of PurposeNet in brief. We then proceed to explain our hybrid approach in section 4, and discuss the preparation and analysis of data in section 5. We finally produce in section 6, the results for the compound interpretation experiments performed, and then discuss the scope of improvement and future work in section 7.

2   Related Work

The most common approaches to handle compound interpretation are broadly categorized under supervised and unsupervised approaches. The supervised approaches combine machine learning techniques with lexical taxonomies to classify the nominal compounds into one of a set of pre-defined semantic relations. The unsupervised approaches are usually data-driven probabilistic methods that collect statistics on the occurrence frequency of compounds in the corpus and use them to predict the most probable interpretation for the compounds. Other approaches have evolved which focus on the use of large, lexical and conceptual hierarchies, both domain specific and domain independent, for this task. The earliest such approach is by Johnston and Busa (1996). They make use of the Generative Lexicon model proposed by Pustejovsky (1991), that couples lexical semantic representations with mechanisms to capture the relations between those representations and their syntactic expressions. Their lexicon consists of a type, argument, event, and qualia structure for every lexical entry. Phrase structure schemata were developed to compositionally understand the links between the qualia of head nouns and their corresponding modifiers. However, this approach works only for those nouns and noun compounds whose qualia are listed in the lexicon. It also restricts the nominal compounds that can be interpreted by the type of the modifier and the action prescribed in the qualia of the head.

Most of the approaches that followed depend on supervised machine learning and domain specific lexical hierarchies. Rosario and Hearst (2001) mapped the nominal compounds into unique concept IDs and into terms in the MeSH medical ontology. They built different models based on these MeSH descriptor terms and trained artificial neural networks to classify every nominal compound into one of the different semantic relations. Kim and Baldwin (2005) introduced a machine learning approach that used WordNet for classifying compounds based on abstract semantic relations. They built a training set of 1088 manually annotated compounds, and interpreted the test cases using WordNet-based similarity. They calculated the similarity between a given test instance and every

training instance in the training set, and predicted the semantic relation of the most similar training instance. They used different similarity measures such as WUP (Wu and Wu, 1994), LCH (Leacock and Chodorow, 1998), JCN (Jiang and Conrath, 1997)and LIN (Lin, 1998) and obtained a good result of 53% accuracy over open domain. The only bottleneck however, is that it requires sufficient training data distributed over the different abstract semantic relations. We therefore increase the robustness of our system by extending ontology search with a module that performs word similarity measurement for compounds.

3 PurposeNet Ontology

3.1 Architectural Design

PurposeNet is a purpose-centric ontology of artifacts, where the artifacts are organized in a multiple inheritance hierarchy. The ontology contains artifacts and relations between them. There are two types of features:

- Descriptive features

- Action features

and three types of relations:

- Subtype

- Component

- Accessory

Every artifact is expressed in terms of 20 descriptive features and 7 action features and is connected to the other artifacts via one or more of the above mentioned relations. While the descriptive features capture information pertaining to the physical nature of the artifact, the action features capture information about the actions performed on and by the artifact, such as its birth, maintenance and destruction. Table 1 shows the descriptive features for *Butter Knife* while Table 2 shows the action features for the *Car*.

3.2 Schema for Handling Nominal Compounds in the Ontology

A nominal compound is a complex construction where two or more different concepts combine together to form a single concept. While on the con-

Descriptive Features	Possible Values	Butter Knife
Color	Black, White, Green	any
Constituition	Metal, Plastic, Foam, Rubber	Steel,Metal
Fluidity	Fluid, Non-fluid	Non-fluid
Heaviness	Light-Weight, Moderate-Weight, Heavy-Weight	Light Weight
Inertness	Inert, Reactive, Alkaline, Acidic	Inert
Mobility	Mobile, Immobile	Immobile
Oiliness	Oily, Non-oily	Non-oily
Physical State	Solid, Liquid, Gaseous	Solid
Shape	Cubical, Cuboidal, Cylindrical, Flat	Flat
Size	Big, Small, Huge	Small
Sliminess	Slimy, Non-slimy	Non-slimy
Smell	Pleasant, Unpleasant, Odourless	Odourless
Smoothness	Smooth, Rough	Smooth
Softness	Soft, Hard	Soft
Sound	Silent, Bearable, Unbearable, Noisy	Silent
Stability	Stable, Non-stable	Stable
Subtleness	Subtle, Non-subtle	Non-subtle
Taste	Sweet, Sour, Bitter	Tasteless
Temperature	Hot, Cold, Room-temperature	Room-temperature
Transparency	Transparent, Translucent, Opaque	Opaque
Viscidity	Viscous, Non-viscous	Non-viscous

Table 1: Descriptive features for *Butter Knife*.

ceptual level, a nominal compound represents a single concept, it manifests as a set of ordered lexical items, due to which representation of a nominal compound in a knowledge base becomes a challenge by itself. Compounds can be represented by a single unit or a single node in the ontology. They can also be broken into their respective constituents and the constituents be placed under appropriate classes with appropriate relations ascribed between them. We discuss in this section how compounds are handled in different ontologies, and the motivation for the schema adopted in PurposeNet.

WordNet (Fellbaum, 1998) is one of the most notable of the available, large scale, general pur-

Action Feature	Subtype	Definition	Some Values for Car
Birth		Manufacture of artifact	Fix Chassis to Body, Attach Seats, Attach Tyres
Purpose		Purpose of artifact	Transport Human
Maintenance	General Maintenance Repair Maintenance	Maintenance of artifact	Clean Car, Clean Engine Repair Car, Repair Engine
Wear and Tear		Wear and tear of artifact	Burst Tyre, Overheat Engine
ProcessRel		Actions the artifact can perform	Board Passengers, Move from A to B, Alight Passengers
Set up	First time Set up General Set up	Set up the artifact for functioning	Check Ignition System, Check Brake Check Tyre, Check Brake
Result On Destruction		Results on destruction of artifact	Engine recycled to metal, Seats - reused

Table 2: Action features for *Car*.

pose, domain-independent ontologies. Although it focuses mainly on the taxonomies of words, it does not attach any significance to the representation of multi-word expressions (MWEs) such as compounds. Most of the compounds listed in WordNet are represented as a single node in the lexical hierarchy. Ex : wildfire, orange juice and mailman. Such a representation is suitable only to compound nouns that are commonly occurring and in which the relationship between the constituents is unambiguous and easily comprehensible (Mahesh, 1996).

ConceptNet (Havasi, 2007) is a large-scale, commonsense knowledge base, similar in structure to WordNet, but the nodes in ConceptNet are mostly semi-structured English fragments or compound concepts connected to each other by semantic relations. Since all the compound formations are covered as individual nodes in the ontology, ConceptNet fails to explicate the relations within the constituents of the compound, much like WordNet. Further, such a representation has led to redundancy in data. Similar compound constructions are represented as different nodes in the ontology. This fails to capture the similarity between different constructions and leaves little scope for interpreting new compounds made from similar constructions. For example, ConceptNet captures *orange juice*, *lemon juice*, and *fruit juice* as different nodes. It fails to capture the information that a fruit juice is made from a fruit, and the fruit can be orange, or lemon or any other.

Yet another ontology that we have surveyed is the Brandeis Semantic Ontology (Pustejovsky et al., 2006) built on the basis of the Generative Lexicon approach (Johnston and Busa, 1996). This ontology uses a type structure, argument structure, qualia structure and an event structure for every entry, and

couples them with phrase structure schemata. However, as discussed in section 2, this approach works only for nouns whose qualia are defined in the lexicon and which adhere to the type of the modifiers and the action prescribed in the qualia. The above discussed drawbacks have motivated us to adopt a new schema for representing nominal compounds in the ontology.

1. Compounds that are incomprehensible by themselves and cannot be predicted from similar constructions are defined as unique nominals. Even compounds that are significantly different from similarly constructed compounds, in terms of their physical nature or in their purpose, manufacture etc, are unique.

2. All unique nominals must be represented by a single node in the ontology hierarchy. Example : *gamma knife* and *garden knife*.

3. All similarly constructed compounds must be represented by a generic compound in such a way that the similarity between the constructions is captured while leaving scope for new constructions to be captured.

Consider the examples of *wheat bread*, *rice bread* and *ginger bread*. Compounds such as *wheat bread*, *rice bread*, *oat bread*, *barley bread* and all other similar constructions can be captured using the feature {component, cereal} in the generic class *bread*. However, gingerbread is a confectionery. It is a unique nominalisation, and must be represented by a single node in the ontology. Such a representation therefore shrinks the ontology from polynomial to linear space without any loss of information.

557

4 APPROACH

We propose a hybrid approach that combines the use of ontology with word similarity measures to interpret nominal compounds. The approach includes the following two phases: (1) ontology search and (2) word-similarity based interpretation. In the first phase, given an nominal compound, we search the ontology to locate the node corresponding to the head or the nominal compound, and then extract the corresponding descriptive and action features of the artifact represented by it. We use pattern matching techniques to match the modifier in these features, and then with the use of a pre-constructed mapping between the features and semantic relations, we map the feature to its corresponding semantic relation in our inventory. This phase can only interpret a compound when both its constituent nouns are covered in the ontology. It fails for the rest of the cases when the head and/or the modifier are not covered in the ontology. To increase the robustness of the system, we adopt in the second phase, a word similarity measurement technique to handle the remaining compounds where the head or the modifier are not covered in the ontology.

4.1 Pre-Processing

The first step in our approach is the pre-processing of the test data and the ontology. Pre-processing of the test instances includes stemming of constituent nouns in the given compounds using Porter Stemmer. In every compound that is made up of one or more plural constituent nouns, we modify Porter Stemmer to stem those plurals and rebuild the compound from the new constituents.

In order to access the ontology in an efficient way, we need to index the nodes in the ontology. We therefore adopt the Dewey Encoding scheme [1] to index the nodes in our ontology hierarchy. This type of indexing helps in quick accessing of the nodes and also makes it easy for traversal from one node to another within the hierarchy. We apply the indexing mechanism starting from the root node of the ontology which is 'Entity' and label it '0'. Every other node in the ontology is given an index containing the path from the root to the node, and each path uniquely identifies the absolute position of the node within the hierarchy (Tatarinov et al., 2002). Fig-

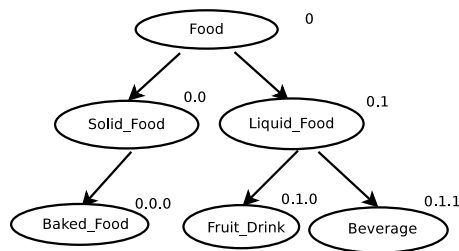ure 1 shows a small part of the indexed PurposeNet hierarchy.

Figure 1: Dewey Order Indexing

4.2 Ontology Search

Given a compound $\langle N1N2 \rangle$ to be interpreted, we first check for the following different possibilities regarding its coverage in the ontology, each of which can be handled in our hybrid approach:

(a) the compound is unique and occurs as a single node in the ontology.

(b) the compound is non-unique and both the head (N2) and the modifier (N1) are encoded as different nodes in the ontology.

(c) the compound is non-unique and only the head (N2) is present in the ontology.

(d) the compound is non-unique and only the modifier (N1) is covered in the ontology.

Case (c) and (d) are handled in the next phase using word similarity measures. However, the case where neither of the constituents of the nominal compound are covered in our ontology is currently not handled in our hybrid approach.

Our strategy is to adopt different search traversals suited to each of the different cases, and use pattern matching to identify the right features in the ontology. For the compound $\langle N1N2 \rangle$, we first define a Left end (LE) and a Right end (RE) as boundaries of our search traversal. They are essentially the nodes representing the compound or its constituents in the ontology. We then obtain the indexes corresponding to N1 and N2 by reading from the index table. Then, we extract all the action features $\langle A \rangle$, and all the descriptive features $\langle D \rangle$ of the RE, and match the LE using simple pattern matching expressions. When the compound is unique and occurs as a single node

in the ontology (case (a)), the LE is the modifier while the RE is the node that defines the compound as an artifact. In the case when both the constituents of the compound occur as different nodes in the ontology (case (b)), the LE is the modifier, while the RE is the head of the compound.

Now, given the LE and the RE, we implement a robust search mechanism between the two ends, using different search traversals postulated below:

- One level search - Consider the example of *lemon tea*, where *tea* has a feature {*component,emon*}. In such cases where the LE and the RE are directly related to each other, we perform a single level search through the features of the RE and extract that feature whose value matches the LE.

- Multi Level Search - This search is called when the one level search fails to return any non-empty feature for the compound. Consider the example of *bedroom light*. We define the family of *bedroom* as its parents and siblings, that is the immediate super class in the hierarchy(parent), and all those nodes that are sub classes of the parent (siblings). We replace the LE (*bedroom*) with each member of its family and repeat the one level search with each of the new LE. Here, we also take into consideration the super class of the LE from WordNet (Hypernym) in the family of LE. The intuition behind this multi level search is that every modifier that belongs to the same family modifies the head noun in the same way. Ex: A *lemon juice* is juice made from *lemon*, and so is a *fruit juice*, juice made from *fruit*.

- If the above search mechanisms fail to retrieve a semantic relation for the compound, we replace the RE with the values of each of its features in turn, and repeat the One level and Multi Level search with the new RE. This search mechanisms particularly holds for compositional compounds that follow the law of transitivity. If *A* is a component of *B* and *B* is a component of *C*, then *A* is a component of *C*.

When all the above search traversals fail to retrieve any feature, then ontology search fails to interpret the given compound.

4.3 Word Similarity Measurement

This phase of our experiments is built on a hypothesis that states that 'a compound can be interpreted, at least in part, by knowledge about the meanings of similar compounds' (Ó Séaghdha, 2008). Therefore, we define our experiments on the unpredicted compounds of phase 1 in such a way that by understanding and interpreting similar compounds formed by each of their constituents taken individually, we can interpret the underlying semantic relation for our target compound instance.

Most of the word similarity measures can be classified into one of the following three classes: (a) Approaches that use knowledge resources such as ontologies and thesauri for extracting information such as glosses of the lexical items, or hierarchy information such as the Is-A from WordNet. (b) Approaches that acquire context information for each of the words and check the overlap between the contexts to calculate the similarity. Here, words that occur in similar contexts are intuitively more similar. (c) Approaches that are specifically built for similarity between word pairs. These approaches consider, for each word pair, the contexts in which the constituents of the word pairs occur together. The intuition behind this approach is that when both the constituents occur together in a particular context, the context will most likely yield information about the relation between the constituents (Ó Séaghdha, 2008).

In this paper, we use the Extended gloss overlap measure that uses gloss information from WordNet (Banerjee and Pedersen, 2003). The extended gloss overlap measure calculates the relatedness between two lexical items by comparing their glosses, along with the glosses of the synsets that are related to these lexical items. As mentioned earlier, cases (c) and (d) are handled in this phase, along with other compounds that can not be predicted in phase 1 due to lack of coverage in the ontology. Our strategy consists of building a set of compounds which we call the base set, and measuring the similarity between a given test instance and each base instance. We choose the k most similar base instances, and interpret them using ontology search. We first explain the steps involved in building the base set for cases (c) and (d).

559

Given a nominal compound whose head is covered as a node in the ontology hierarchy, we extract all the descriptive features of the corresponding node, as well as the descriptive features of every action feature, as each of these can be potential modifiers of the head in a compound. For example, consider a nominal compound $\langle N1, N2 \rangle$, where N1 and N2 are the modifier and the head respectively, and N2 is covered in the ontology. We extract the descriptive features $\langle D \rangle$ and the action features $\langle A \rangle$ of N2. However, since the action features are verbs and cannot act as modifiers in a nominal compound, we extract the descriptive features $\langle D' \rangle$ of the node corresponding to every A $\epsilon \langle A \rangle$ and append $\langle D \rangle$ and $\langle D' \rangle$ to the list of potential modifiers $\langle M \rangle$. In the next step, we add the family members(parent, siblings) of every M $\epsilon \langle M \rangle$. We then run a POS tagger on each of these modifiers and prune away those modifiers that are NNP or are not tagged as nouns. For each of the remaining modifiers M in $\langle M \rangle$, we construct compounds with the head N2 and form the base set of compounds.

For a compound whose modifier is covered in the ontology, we follow a similar strategy as above, building a set of base compounds based on on the occurrences of the modifier in the ontology. Consider again, a nominal compound $\langle N1, N2 \rangle$, where N1 and N2 are the modifier and the head respectively and N1 is covered in the ontology. In this case, we extract all the nodes in the ontology $\langle N \rangle$ and for every N $\epsilon \langle N \rangle$, we extract its descriptive features $\langle D \rangle$, and action features $\langle A \rangle$. We then check its descriptive features $\langle D \rangle$ as well as the descriptive features $\langle D' \rangle$ of every node corresponding to the action features $\langle A \rangle$ for any occurrence of the modifier N1. For every occurrence of the modifier N1, we construct a compound of N1 with that node N and add to our list of base compounds.

In the next step, we use the *WordNet::QueryData*[2] to query the different senses of the constituents for the test compound $\langle N1, N2 \rangle$ and each of the base set compounds $\langle N1', N2' \rangle$. For each sense of N1 and each sense of N1', we calculate the relatedness using the extended gloss overlap measure, and obtain the most related senses. Similarly, we ob-

---

²`http://search.cpan.org/dist/`
`WordNet-QueryData/QueryData.pm`

Base compound	Similarity of modifiers	Similarity of heads
Fruit Soup	5145	70
Fruit Skin	5145	40
Fruit Drink	5145	37

Table 3: Base compounds and their similarity with *Fruit custard*.

tain the most similar senses of N2 and N2', and then calculate the similarity between the compounds $\langle N1, N2 \rangle$ and $\langle N1', N2' \rangle$ as the product of the similarity between the most related senses of their corresponding constituents N1-N1' and N2-N2'. For example, the 3 most similar base compounds to the nominal compound *fruit custard* were found to be *fruit soup*, *fruit skin* and *fruit drink*. We show in Table 3, the relatedness measures of these base compounds with our test compound. Finally, we use the k-best method and obtain k(3 or 4) base compounds that are most similar to the test compound. We then interpret them using the ontology search. This results in a set of most probable interpretations for the test instance. Human judgement would be required to choose the most appropriate interpretation out of the most probable interpretations.

5 Preparation of Data

5.1 Extraction

In order to perform our experiments on nominal compounds that are within the tourism domain, we compiled a list of those compounds from the web, that were formed by the artifacts listed in our PurposeNet ontology. Firstly, we extracted all the artifacts that are described in the ontology. Those artifacts which are represented using noun compounds in the ontology were further split into their corresponding noun constituents and then appended to the list of nouns for web search. We then used the Bing Search API to search the web for all occurrences of the nouns representing each of the artifacts and extract compounds formed by them. The search was restricted to the top 10 web results obtained from Bing. We used a simple heuristic to identify the noun compounds, similar to that used by Lauer (1996). All those occurrences of the artifact nouns that were preceded or succeeded by nouns,

and which are not flanked by tokens tagged as nouns on either side were extracted as noun compounds. A list of stop words was then used to prune away compounds containing junk words. This extraction does not retrieve hyphenated noun sequences as these noun sequences need further validation before appending to the list of nominal compounds.

There are a total of 616 artifacts in the ontology. These artifacts were used to prepare a list of 400 nouns for the web search. We ran the bing web search API on each of these 400 nouns and a total of 89,578 compounds were extracted. However, this extraction mechanism also resulted in incorrect cases due to false tagging of words as nouns. Therefore, we manually identified and extracted from this list, some compounds which were restricted to our tourism domain, to form a small test set of 600 compounds for our experiment.

5.2 Semantic Relations

In our experiment, we use an inventory of 22 semantic relations proposed by Girju (2006) for interpreting the nominal compounds. We chose this list as it contains clearly defined semantic relations, with clear and well defined boundaries and sufficient coverage of the different possible semantic relations that can exist between two nouns. Moreover, most of these relations are captured as features in PurposeNet. This simplifies the task of mapping features from the ontology to these relations in the PurposeNet experiment.

5.3 Annotation

We used two human annotators for annotating the compounds. Only a list of compounds and the annotation guidelines were provided for the annotation. The compounds were allowed to be annotated with more than one semantic relation, as and when suitable. Table 4 shows the distribution of the compounds among the different relations, for each of the annotators. Rarely occurring relations (<5 times) have not been considered in the table. We observe that different semantic relations in our inventory provide different depths of interpretations for the nominal compounds. For example, the *type* relation has a very 'surfacy' nature, and most of the compounds can be classified into this class. A *wine bottle* is a type of bottle, *glass furniture* is a type of fur-

Relation	Annotator 1	Annotator 2
Part-Whole	184	90
Type	178	192
Purpose	132	122
Source	70	78
Property	25	22
Hypernymy	25	30
Location	34	30
Topic	12	8
Theme	8	8
Temporal	5	5

Table 4: Distribution of annotated data among the relations.

niture, and similarly, *orange juice* is a type of juice. Alternatively, each of these compounds can be interpreted using deeper semantic relations, such as *purpose* and *source*. A *wine bottle* can be interpreted as a bottle used to serve wine, a *glass furniture* is furniture made up of glass, and *orange juice* is juice made from orange. Therefore, in case of such compounds, both *type* and *purpose*, *type* and *source* can be considered as the appropriate annotations. However, all occurrences of *purpose*, *source* and other relations cannot be replaced using *type*. Since there is no clear method of distinguishing the agreements from the disagreements in annotations involving *type*, we choose to calculate the inter annotator agreement in two ways. The first calculation counts all the mismatches between *type* and *purpose*, *type* and *source* etc as disagreements, while the second calculation counts all the mismatches containing *type* as one of the annotations as an agreement between the annotators. In the first case, the inter-annotator agreement on the 600-nominal compounds set was 65.6% with a moderate kappa score of 0.57. The second calculation, on the other hand, produced a high inter annotator agreement of 89% with a kappa score of 0.87. The ideal inter-annotator agreement can be defined as a value belonging to range bound by these two limits.

It is evident from analysis that *part-whole(meronymy)* and *purpose* exhibit very little agreement with each other. This can be mainly attributed to the possibility of more than one correct interpretations for a given compound. For example,

Annotation1	Annotation 2	Disagreement count
Part-Whole	Type	90
Purpose	Type	22
Source	Type	14
Hypernymy	Type	12
Part-Whole	Source	10
Property	Type	6
Part-Whole	Purpose	6
Purpose	Theme	6
Purpose	Location	6
Source	Hypernymy	6

Table 5: Distribution of the disagreements in annotation[3].

Relation	Predicted		Unpredicted
	Phase1	Phase2	
Meronymy	22.5	14	3
Purpose	16.5	3.5	6
Type	22.5	12	10
Location	0	2.5	1
Source	2.5	2.5	1
Hypernymy	6	5	2
Property	3	0	2
Beneficiary	3.5	1	0

Table 6: Distribution of the predicted and unpredicted nominal compounds.

a *glass furniture* can be interpreted as 'furniture made of glass' (*part-whole*) or 'furniture made from glass' (*source*) or 'a type of furniture' (*type*). All such instances of disagreements that occurred in our data set were solved using a third annotator whose judgement was chosen as final.

6 Experiment and Results

We conducted our experiments on the set of 600 nominal compounds extracted from web using the Bing Search API. Each nominal compound was allowed to be interpreted using more than one feature from PurposeNet. These features were in turn mapped to their corresponding semantic relations in our inventory using a set of rules that were built manually based on the definitions of the features and semantic relations. In order to evaluate the performance of the hybrid approach, we adopt a single-label evaluation method where compounds with atleast one correctly predicted label are considered to be correctly interpreted by our system. However, we disregard the less informative labels such as *hypernymy* and *type* as correct interpretations for any compound, and do not consider them in our evaluation. The results of our experiment on the 600 nominal compounds are reported in detail in Table 6. The first column lists all the semantic relations that were found in our data set. The second column reports the distribution of the compounds that were successfully interpreted by our model, with detailed contribution of each phase, for each semantic

relation. The last column gives the distribution of those compounds that failed to be interpreted by our hybrid model. Compounds that were annotated with multiple labels were counted under each of the labels. We observe that our system has precision and recall values of 0.76 and 0.92 respectively, while its overall accuracy (calculated as the ratio of the number of correctly predicted compounds to the total number of compounds) is 0.70. As shown in the table, most of the uninterpreted compounds belong to *type* and *purpose* relation. We also observe that of all the nominal compounds that were predicted, 55% of the compounds were predicted in phase 1 of the approach, while the remaining 39% of the compounds were predicted at the end of phase 2. This indicates that the addition of lexical word similarity measures to our ontology search has caused a significant improvement in the results of compound interpretation.

7 Conclusion and Future Work

We have observed that most of the approaches proposed so far for understanding nominal compounds implement machine learning techniques or statistical prediction methods to classify nominal compounds into different semantic relations. In this paper, we described a hybrid, ontology-based approach for the understanding and labeling nominal compounds with semantic relations. It is a unique system that combines lexico-semantic information from a domain-specific hierarchy with gloss information from WordNet for interpreting two word nominal compounds. It implements an efficient look-up

---

[3]Minor disagreements (<5) have not been shown in the table.

mechanism that uses minimised search space for searching nominal compounds in the ontology. To increase the robustness of the system, we use lexical similarity measures based on gloss information from WordNet to handle compounds beyond the scope of our ontology.

We presented the experimental results of our hybrid approach, and compare the contribution of each phase of our system in successfully interpreting nominal compounds. Our system has achieved an accuracy of about 70% on domain-specific nominal compouns and is comparable in its performance to Girju's state-of-the-art best performing system for domain-independent nominal compounds (Girju et al., 2005) that reports, on an average, an accuracy of 75%. This motivates us to further experiment our hybrid approach our hybrid model on different ontologies (such as ConceptNet and WordNet) and different lexical and relational word similarity measures and compare their performance for the task of compound interpretation.

References

Timothy Baldwin and Takaaki Tanaka. 2004. Translation by machine of complex nominals: Getting it right. In *Proceedings of the ACL04 Workshop on Multiword Expressions: Integrating Processing*.

Satanjeev Banerjee and Ted Pedersen. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pages 805–810.

Christiane Fellbaum, editor. 1998. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London, May.

Roxana Girju, Dan Moldovan, Marta Tatu, and Daniel Antohe. 2005. On the semantics of noun compounds. *Comput. Speech Lang.*, 19(4):479–496, October.

Roxana Girju. 2006. Out-of-context noun phrase semantic interpretation with cross-linguistic evidence. In *Proceedings of the 15th ACM international conference on Information and knowledge management*.

Catherine Havasi. 2007. Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge. In *the 22nd Conference on Artificial Intelligence*.

Jay J. Jiang and David W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy.

Michael Johnston and Federica Busa. 1996. Qualia structure and the compositional interpretation of compounds.

Kim and Baldwin. 2005. Automatic interpretation of noun compounds using wordnet similarity. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing, Jeju Island, South Korea, 11–13*, pages 945–956.

P. KiranMayee, Rajeev Sangal, Soma Paul, and Navjyoti Singh. 2008. An ontological resource organized around purpose. In *Proceedings of 6th International Conference on Natural Language Processing, Pune*.

Mark Lauer. 1996. Designing statistical language learners: Experiments on noun compounds. *CoRR*, cmp-lg/9609008.

C. Leacock and M. Chodorow, 1998. *Combining local context and WordNet similarity for word sense identification*, pages 305–332. In C. Fellbaum (Ed.), MIT Press.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304. Morgan Kaufmann.

Kavi Mahesh. 1996. Ontology development for machine translation: Ideology and methodology. Technical Report MCCS-96-292, CRL, New Mexico State University.

Diarmuid Ó Séaghdha. 2008. *Learning compound noun semantics*. Ph.D. thesis, Computer Laboratory, University of Cambridge. Published as University of Cambridge Computer Laboratory Technical Report 735.

James Pustejovsky, Catherine Havasi, Jessica Littman, Anna Rumshisky, and Marc Verhagen. 2006. Towards a generative lexical resource: The brandeis semantic ontology. In *Proceedings of the Fifth Language Resource and Evaluation Conference*.

James Pustejovsky. 1991. The generative lexicon. *Comput. Linguist.*, 17(4):409–441, December.

Barbara Rosario and Marti Hearst. 2001. Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP-01*, pages 82–90.

Igor Tatarinov, Kevin Beyer, and Jayavel Shanmugasundaram. 2002. Storing and querying ordered xml using a relational database system. In *In SIGMOD*, pages 204–215.

Zhibiao Wu and Zhibiao Wu. 1994. Verb semantics and lexical selection.