

Unsupervised Verb Inference from Nouns Crossing Root Boundary

Soon Gill Hong Department of KSE KAIST Daejeon, Republic of Korea hsoongil@gmail.com	Sin-Hee Cho Department of KSE KAIST Daejeon, Republic of Korea chosinhee@kaist.ac.kr	Mun Yong Yi Department of KSE KAIST Daejeon, Republic of Korea munyi@kaist.ac.kr
---	---	---

Abstract

Inference about whether a word in one text has similar meaning to another word in the other text is an essential task in order to understand whether two texts have similar meaning. However, this inference becomes difficult especially when two words do not share a lexical root, do not have the same argument structure, or do not have the same part-of-speech. This paper presents an unsupervised approach for inferring verbs from nouns along with a new online resource PreDic (PREdicate DICtionary) that contains verbs inferred from nouns sharing similar concepts but not the root. The verbs in PreDic are categorized into three groups, enabling applications to target precision-oriented, recall-oriented, or harmony-oriented results as needed. The experiment results show that the proposed unsupervised approach performs similar to or better than WordNet and NOMLEX. Furthermore, a new domain-verb association measure is presented to show the association relationships between inferred verbs and domains to which the verbs are possibly applied.

1 Introduction

The variability of expression is an underlying phenomenon in natural language, and the recognition of the variability serves as the foundation of understanding natural language. Recognizing textual entailment is a research area that seeks to understand this variability, and thus to identify, generate, or extract textual entailment relations from texts. Textual entailment describes a relation of texts where the meaning of one text can be inferred plausibly from another text (Dagan et al., 2010). As a related term, paraphrases refer to expressions that deliver almost the same information using different words (Androutopoulos and Malakasiotis, 2009). As a lighter form of textual entailment, inference rules refer to expressions that carry not only the same meanings but also similar meanings and could be useful to question answering (Lin and Pantel, 2001).

Much research in recent years has focused on recognizing textual entailment pairs in natural language texts. For example, consider the following sentences:

- (1) Emily Bronte *wrote* Wuthering Heights.
- (2) Emily Bronte *authored* Wuthering Heights.

Given that these two sentences deliver the same meaning, the verbs *wrote* and *authored* are in a textual entailment relation.

Textual entailment plays a very important role in many areas. For example, in question answering, paraphrases from bilingual parallel corpus were used to expand the original questions (Lin and Pantel, 2001; Duboue and Chu-Carroll, 2006; Riezler et al., 2007); in information extraction, paraphrases were extracted and then used to find entities to fill the slots of binary relations (Shinyama and Sekine, 2003); in machine translation, paraphrases were captured and used as part of reference translations (Madhani et al., 2007; Marton et al., 2009).

This work is licenced under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

Among textual entailment relations, recognizing the textual entailment of the *predicate* part of a sentence is a hard task, especially when two sentences use different words of different parts-of-speech and different argument structures. This difficulty becomes worse if the predicates of two sentences share neither any lexical root nor have proper chains from thesauri. For example, consider the following sentences:

(3) The *ingredients* of pasta are flour, eggs, and a little bit of water.

(4) Pasta is *made from* flour, eggs, and a little bit of water.

(5) What is the main *ingredient* of pasta?

As example (3) and (4) deliver the same meaning, the words *ingredients* and *made from* have a textual entailment relation. Moreover, example (4) can be an answer to example (5). However, those text pieces share neither any lexical root (*make* vs. *ingredient*) nor any syntactic structure (*X predicate Y* vs. *predicate preposition (of) X linking-verb (is) Y*) nor part-of-speech (*verb* vs. *noun*), so recognizing them as a textual entailment relation is harder than between examples (1) and (2). These inferences from nouns to verbs crossing root boundary remain unclear, and no resources have been published so far, to the best of our knowledge.

This paper presents a new unsupervised approach of inferring verbs from nouns, which share concepts but do not share roots, from glosses of multiple dictionaries. Unsupervised verb inference from nouns crossing root boundary, which covers the variable expressions between nouns and verbs, can be used to help recognize textual entailment relations. PreDic implemented the new approach and can be accessed online at <http://lod.kaist.ac.kr/predic>. PreDic only works for English nouns.

2 Related Work

Collecting similar words from a text is largely based on the *Distributional Hypothesis* (Harris, 1981). The basic idea is that words that occur in the same contexts tend to have similar meanings. Many studies in the literature acquired inference rules or paraphrases based on this hypothesis (Lin, 1998; Lin and Pantel, 2001; Bhagat and Ravichandran, 2008). If we apply that idea to the glosses of dictionaries, then we obtain many similar or relevant words in the glosses for entry words in the dictionaries.

Using dictionary glosses to understand natural language has been a popular approach. Lesk (1986) tried to identify the correct sense of each of two adjacent words, each of which having more than one gloss in the dictionary, by counting overlaps among the combinations of each gloss of each word. Glosses also have been used for extending the functionalities of another resource. Extended WordNet was built by analyzing glosses and extracting extra relations for WordNet synsets (Harabagiu et al., 1999).

Nominalization is a way of inferring nouns mainly from verbs or adjectives, especially when they share the same root. Macleod et al. (1998) built a dictionary of nominalization, NOMLEX (NOMinalization LEXicon). NOMLEX contains the nominalizations of verbs with additional information to relate the complements of nouns to the arguments of the corresponding verbs. This dictionary can be used to capture the following textual entailment relation (Bedaride and Gardent, 2009).

(6) Rome's *destruction of* Carthage.

(7) Rome *destroy(ed)* Carthage.

Argument-Mapped WordNet (Szpektor and Dagan, 2009) provides explicit mappings of arguments between verbs to alleviate the difficulty of tracking argument changes. They manually built or automatically captured rules to augment WordNet's inference capability, which permits inference over predicates only on substitution relations, such as synonyms and hypernyms, e.g. *buy* \Rightarrow *acquire*. The Argument-Mapped WordNet defined only unary rules for verb-nominalization relations and verb-verb relations (e.g., X_{obj} 's *employment* \Leftrightarrow *employ* X_{obj} as a nominalization-verb relation or X_{subj} *break*_{intrans} \Rightarrow *damage*_{trans} X_{obj} as a verb-verb relation).

However, neither the resources of nominalization nor the mappings of argument changes can recognize examples (3) and (4) as textual entailment relations. Nonetheless, we may find a clue by chaining in WordNet. WordNet (Miller, 1995) and Extended WordNet (Harabagiu et al., 1999) contain links among synsets, so paraphrases that cross lexical root boundaries can be captured by chaining (i.e., *noun A* → *verb form B of noun A* → *verb synonym C of verb B*). We adopted this approach to compare PreDic to WordNet in the experiment.

3 Methodology

Among many entailment relations, our methodology has focused on the relations between nouns and verbs that represent similar concepts without sharing their roots. Specifically, verbs for nouns that do not share the same root are collected and then used to recognize textual entailment relations, such as the examples (3) and (4) above. The following sections describe how we collect noun-verb entailment relations crossing roots and how we categorize them for applications.

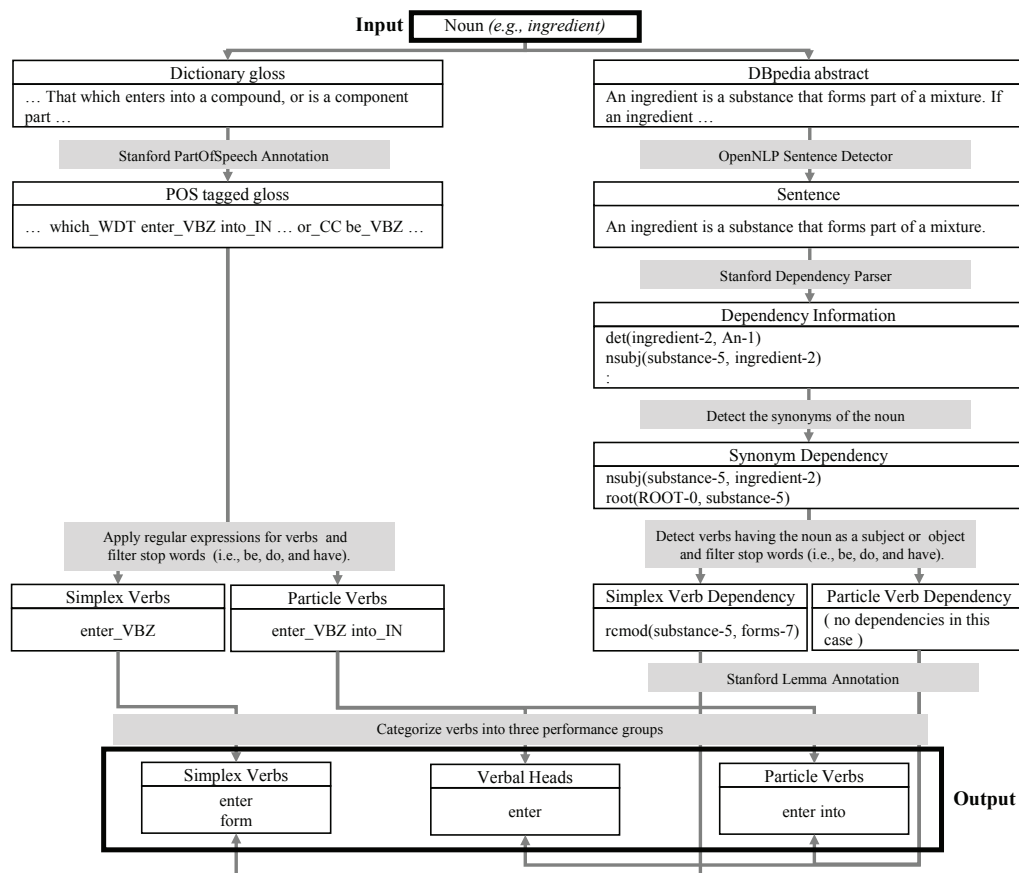


Figure 1: Algorithm consists of two major steps for generating verbs from nouns: acquisition and categorization. Verbs are detected by matching regular expression patterns against POS tagged glosses and by analyzing dependency information, and then categorized into three verb groups (Simplex, Particle, and Verbal Head).

3.1 Acquisition

Verbs that can express a similar concept of a noun can be extracted by analyzing the dictionary glosses of the noun. For example, *The Collaborative International Dictionary of English* describes *ingredient* as “...That which *enters into* a compound, or is a component part of any combination, recipe, or mixture; an element; a constituent...” This example shows that verbs used in the dictionary glosses for a noun can be regarded as having entailment relationships with the noun. Encyclopedias, such as Wikipedia¹, can

¹<http://www.wikipedia.org>

also be used as a source for collecting such verbs. Here is an excerpt from Wikipedia article on the word *ingredient*: “An ingredient ... *forms...used ... purported ... required ... listed ... consists of ...*” Table 1 shows sample collected verbs that have entailment relationship with the noun *ingredient*.

Noun	Dictionary Gloss	Collected Verb
ingredient	An ingredient is a substance that <i>forms</i> part of a mixture (in a general sense)... If an ingredient itself <i>consists of</i> more than one... (Wikipedia)	<i>form, consist of, enter into</i>
	... which <i>enters into</i> a compound, or is a component part of any combination, recipe, or mixture; an element; a constituent... (The Collaborative International Dictionary of English)	

Table 1: An example of how verbs are collected from glosses. Simplex verbs and particle verbs are collected from the glosses.

Our approach uses five freely available online resources to infer verbs: The Collaborative International Dictionary of English Version 0.48 (which is also referred to as GCIDE), WordNet 3.0, DBpedia version 3.8² (which is a structured version of Wikipedia), dictionary.cambridge.org (especially, Cambridge Learners Dictionary and Cambridge Advanced Learners Dictionary), and www.merriam-webster.com (especially, Merriam-Webster’s Collegiate Dictionary and Merriam-Webster’s Learners Dictionary).

Fig. 1 shows the algorithm for generating and categorizing verbs. Dictionary glosses and texts from DBpedia are processed in different ways in the algorithm. As most of the dictionary glosses are phrases rather than sentences, they have simple syntax and few numbers of verbs. Therefore, after tagging parts-of-speech to every word in the glosses, regular expressions are used to capture verbs. However, texts from DBpedia are composed of several sentences and contain comparatively large numbers of verbs. Thus, dependency parsing is used to capture the verbs that have “close” relations to the noun.

The detailed procedures for generating verbs from dictionary glosses are described here. At the beginning, Stanford CoreNLP³ adds a part-of-speech tag to each word in the gloss. Then, a regular expression captures simplex verbs of which part-of-speech tag is either one of the “VB”, “VBD”, “VBG”, “VBN”, “VBP”, or “VBZ”. Another regular expression captures particle verbs of which verb’s part-of-speech tag is one of the listed above and particle’s part-of-speech tag is either “RP” or “IN”. Then, verbs that are too commonly used such as *have*, *be*, and *do* are filtered out. At the end, the captured verbs are categorized into three verb groups. A detailed explanation of the categorization is described at section 3.2.

The detailed procedures for generating verbs from DBpedia texts are described here. At the beginning, OpenNLP Sentence Detector⁴ splits the text into sentences. Next, Stanford Dependency Parser⁵ generates dependency information about the words in each sentence. Then, a list of noun synonyms are gathered from *nsubj* and *root* tags (for more information about the tags or relation names, see de Marnette et al. (2008)). Next, simplex verbs are captured. That is, if the noun or any of the noun synonyms appears at the head position with any of *rmod*, *ccomp*, *parataxis*, *vmod*, *partmod*, and *infmod* tag, then the word in the dependent position is captured. Similarly, if any of the noun synonyms appears at the dependent position with any of *nsubj*, *nsubjpass*, *xsubj*, and *doobj*, then the word at the head position is captured. In case of *pobj* that has “preposition” as a head and “object” as a dependent, the verb located at a different dependency relation is extracted by recursively tracing dependency relations. Afterwards, particle verbs are captured by finding particles for each of the simplex verb. That is, if a dependency relation has a *prep* tag and has any one of the simplex verbs at the head position, then the word at the dependent position is regarded as a particle candidate. When the part-of-speech tag for the particle candidate is either “RP” or “IN”, then the particle candidate is regarded as a particle of interest. Consequently, the combination of the simplex verb and the particle is generated as a particle verb. Finally, the captured

²<http://wiki.dbpedia.org/Downloads38?v=6c5>

³<http://nlp.stanford.edu/software/corenlp.shtml>

⁴<https://opennlp.apache.org>

⁵<http://nlp.stanford.edu/software/stanford-dependencies.shtml>

verbs are categorized into three verb groups.

3.2 Categorization

We pay special attention to particle verbs. Particle verbs are a combination of a verb usually with an adverb or a preposition (Blaheta and Johnson, 2001). The adverbs or prepositions, when combined with simplex verbs, generate another concept that simplex verbs alone do not carry. For example, by adding the second word to *shoot*, various concepts can be produced: *shoot up*, *shoot off*, etc (Meyer, 1975). Hence, the definition of particle verb we use here is, to a certain degree, similar to the definition of multi-word verbs that, at the least way, carry extra meaning, or some of the words have a restricted or modified meaning when they go together.

Thus, we assume that particle verbs in text play more important roles than simplex verbs by delivering the author’s intention more specifically. Based on this assumption, we built a performance group model that categorizes each verb into up to three groups. Fig. 2 shows how collected verbs are assigned to three different performance groups: (1) group of simplex verbs and verbal heads from particle verbs, (2) group of particle verbs, and (3) group of verbal heads from particle verbs. A simplex verb can be assigned to only simplex group while a particle verb, as a whole or only as a verbal head, can be assigned up to three groups. Fig. 2 formalizes the concept of categorization.

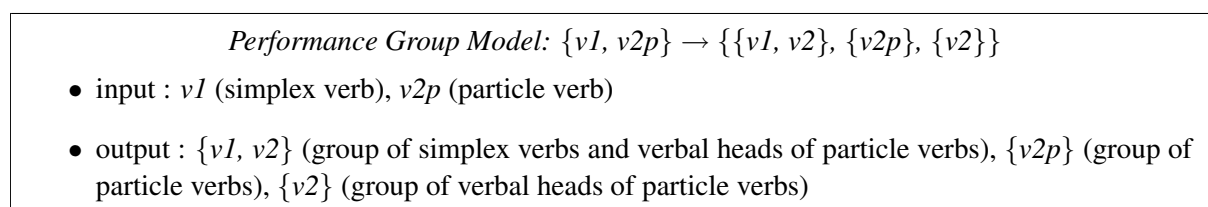


Figure 2: Performance group model that assigns collected verbs into three verb groups ($v1$: simplex verb, $v2$: verbal head of particle verb, $v2p$: particle verb). A simplex verb can be assigned to only simplex group while a particle verb, as a whole or only as a verb part, can be assigned up to three groups.

For example, when *form*, *consist of*, and *enter into* are collected for *ingredient*, they are categorized as follows: *form*, *consist*, and *enter* are assigned to the simplex group; *consist of* and *enter into* are assigned to the particle group; *consist* and *enter* are assigned to the verbal head group. Table 2 shows an example of categorization in detail.

Collected Verb	Verb Group		
	Simplex	Particle	Verbal Head
form, consist of, enter into	form, consist, enter	consist of, enter into	consist, enter

Table 2: Examples of how verbs are categorized into up to three verb groups (Simplex: group of simplex verbs, Particle: group of particle verbs, Verbal Head: group of verbal heads of particle verbs).

3.2.1 Simplex Verb Group

Only simplex verbs among collected verbs are assigned to the simplex group. For example, if the following verbs are collected from the glosses on the word *ingredient* (*form*, *consist of*, *enter into*), then *form*, “*consist*” of *consists of*, and “*enter*” of *enter into* are assigned to the simplex group. As the number of verbs in the simplex group is the largest among the three verb groups, chances are that the number of recognized texts in entailment relations using verbs in this group would be the largest among the three groups. Therefore, verbs in this group should be used to recognize as much relevant information as possible in spite of low precision. In other words, this group is suitable for recall-oriented tasks.

3.2.2 Particle Verb Group

Only particle verbs composed of two words are assigned to the particle group. For example, *consists of* and *enter into* from the collected verbs in Table 2 are assigned to the particle group. As the verbs in

this group are all particle verbs, chances are that the number of recognized texts in entailment relations using verbs in this group would be the smallest among the three groups. Therefore, verbs in this group should be used to recognize as much accurate information as possible at the expense of low recall. In other words, this group is suitable for precision-oriented tasks.

3.2.3 Verbal Head Group

The verbal head of a particle verb is the word that determines the syntactic type or the nature of that particle verb. Only verbal heads of particle verbs are assigned to the verbal head group. For example, “*consist*” of *consist of* and “*enter*” of *enter into* from the collected verbs in Table 2 are assigned to the verbal head group. This group comes between the simplex group and the particle group in terms of both precision and recall. For example, if one searched for *consist* in a text, then texts with *consist of* and *consist in* would be retrieved. It is not clear whether *consist in* fits the search needs, but it is reasonable to think that the word *consist* is common in both of the two types of search results, and therefore, all of the results would share some meaning to a certain extent. Consequently, verbs in this group should be used as a compromise between precision and recall. In other words, this group is suitable for harmony-oriented tasks.

4 Experiment

The experiment aimed at proving three things: the application performance of PreDic compared to NOM-LEX that is regarded as a baseline system, the application performance of PreDic compared to WordNet, and the efficiency of the performance group model in real use. We will discuss the application performance at sections 5.1 and the efficiency of performance group model at section 5.2, respectively. In this section, we describe how the experiment was designed and performed.

4.1 Task: Textual Entailment for Relation Extraction

Relation extraction is one of the application areas that uses textual entailment as a core function. PreDic was used to extract binary relations that have textual entailment. Binary relation, $relation(X,Y)$, is one of the typical relation types, and extracting binary relation can be classified into three tasks: given two instances of X and Y (e.g., *pizza* and *dough*), find relations (e.g., *ingredient*); given one instance of X (e.g., *pizza*) and a relation (e.g., *ingredient*), find the other instances of Y (e.g., *dough*); and given a relation (e.g., *ingredient*), find instances of X and Y (e.g., *pizza* and *dough*) (Sarawagi, 2008).

As the second type (i.e., given one instance of X and a relation, find the other instances of Y) can have predefined noun relations, an experiment with this type can show how noun relations and verb relations are used interchangeably. Therefore, the experiment was performed with a predefined list of subject instances and noun relations.

4.2 Test Data

A PASCAL RTE (Recognizing Textual Entailment) dataset would be the best choice for experiment. However, as a PASCAL RTE dataset for information extraction is composed of pairs of texts, rather than a text and a structured template like the second type mentioned above (Dagan et al., 2009), it was difficult to validate the proposed approach’s capability of inferring verbs from nouns.

Therefore, we decided to use pairs of templates and texts from Wikipedia because they are easily found in Wikipedia. Article names were used as subject instances, and the property names of the infobox were used as noun relations (Wikipedia’s infobox is a fixed-format table provided by the system, and people populate the table to present a summary of an article text). However, we used DBpedia instead of Wikipedia in the experiment. This is because DBpedia is easier to access from application viewpoint. That is, it already captured infobox property names from Wikipedia’s article. Furthermore, DBpedia provides first few sentences as abstracts from Wikipedia’s article rather than full text that is sometime too long and complex to process.

Templates were built for *Cuisine* and *Country* domains. For the *Cuisine* domain, 1,029 cuisine instance names were prepared based on the top 10 countries that have the largest number of cuisine related pages in Wikipedia’s cuisine category. For the *Country* domain, 206 country instance names were prepared.

Domain	Relation	Definition
Cuisine	ingredient	Substance of the cuisine
	origin	The country or period of origin
	serving	Temperature or dishes served with
Country	border	Geographical units such as countries, rivers, or mountain, etc.
	language	Official or unofficial spoken languages
	population	The number of people living in the borders of the country

Table 3: Noun relations and definitions about relations used for the experiment.

Each domain had three noun relations (*ingredient*, *origin*, and *serving* for *Cuisine*, and *border*, *language*, and *population* for *Country*). These relations were chosen according to the frequencies of infobox property names. Thus, a total of 3,087 (1,029 instances multiplied by three noun relations) templates were prepared for the *Cuisine* domain, and a total of 618 (206 instances multiplied by three noun relations) templates were prepared for the *Country* domain. Table 3 shows the noun relations and their descriptions used in this experiment.

Total 7,996 sentences were prepared as texts from DBpedia for the *Cuisine* domain, and 3,062 sentences were prepared as texts from DBpedia for the *Country* domain. Three human raters read these sentences and marked whether each sentence expressed a similar concept to the prepared templates. For example, if a rater read “Typically pasta is made from an unleavened dough of a durum wheat flour ...”, then the rater was supposed to mark the sentence as “relevant” to the template of *ingredient* (X, Y). The agreement could be subjective, so we adopted a majority vote from three raters for each sentence. Hence, the sentences upon which the two raters agreed were annotated as relevant and put into the answer set. Each rater worked independently and was not aware of how our proposed algorithm worked.

4.3 Execution

For a given template (e.g., *ingredient* (*Pasta*, Y), a number of patterns were generated by substituting the noun relation with verbs from PreDic (e.g., *made from* (*Pasta*, Y), *contain* (*Pasta*, Y), etc.). When the subject and predicate of each sentence matched the subject instance and verb of each pattern, the sentence was marked as “retrieved”. If the retrieved sentence exists in the answer set, then it is marked as “retrieved and relevant”.

The performances of PreDic was compared to the performances of NOMLEX. The verbs from NOMLEX were manually collected for the experiment. We also compared the performances of PreDic to the performances of WordNet. However, getting similar verbs of PreDic from WordNet was hard because WordNet does not directly provide verbs for a noun unless the noun itself also has a verb form. Hence, we adopted to collect verbs chaining words by navigating relations in WordNet (Szpektor and Dagan, 2009). We performed chaining up to a certain level until we could collect a similar number of verbs to PreDic. For example, when a similar number of verbs were extracted in the first search for the noun, then all verbs were collected and stopped (level 1). If a similar number of verbs were not extracted, then the extracted noun synonyms were searched again for verbs, and so on. MIT Java WordNet Interface (Finlayson, 2013) was used to collect verbs for the six noun relations from locally installed WordNet 3.1 (see Appendix for the complete list of the acquired verbs from PreDic and WordNet for the experiment. Simplex verbs are omitted if they can be generated by particle verbs).

5 Result and Discussion

5.1 Comparison to NOMLEX and WordNet

As NOMLEX provides verbs as long as nouns have their verbal forms, the performances of the two nouns (i.e., *ingredient* and *language*) could not be measured. Moreover, NOMLEX provides only simplex verbs, so only the performances using simplex verbs could be measured. Table 4 shows that PreDic is better at recall for all relations. In terms of F1, PreDic is better for four relations (i.e., *ingredient*,

origin, *language*, and *population*) while NOMLEX is better for two relations (i.e., *serving* and *border*). However, PreDic is better at precision for only two relations (i.e., *ingredient* and *language*) while NOMLEX is better at precision for four relations (i.e., *origin*, *serving*, *border*, and *population*). Although NOMLEX performs more precisely, its limited coverage degraded the overall performance of the resource.

Relation (R.S.)	Verb Group	PreDic					NOMLEX				
		Ret	R.R.	Pre	Rec	F1	Ret	R.R.	Pre	Rec	F1
ingredient (1413)	simplex	1104	571	0.52	0.40	0.45	-	-	-	-	-
origin (636)	simplex	1298	242	0.19	0.38	0.25	97	81	0.84	0.13	0.22
serving (505)	simplex	1124	344	0.31	0.68	0.42	407	285	0.70	0.56	0.63
border (233)	simplex	259	112	0.43	0.48	0.45	105	105	1.00	0.45	0.62
language (80)	simplex	245	9	0.04	0.11	0.06	-	-	-	-	-
population (133)	simplex	216	7	0.03	0.05	0.04	3	1	0.33	0.01	0.01

Table 4: Application Performance of PreDic and NOMLEX. The coverage of NOMLEX is limited to the nouns that have verbal forms. R.S.: number of relevant sentences, Ret: number of retrieved sentences, R.R.: number of retrieved & relevant sentences, Pre: precision (%), Rec: recall (%), F1: F1 score (%). The best scores for each relation are printed in **bold**.

Table 5 shows the performance comparison between PreDic and WordNet. According to Table 5, PreDic is better at recall for all relations except *border*. PreDic is better at precision for three relations (*ingredient*, *origin*, and *population*) and WordNet is better for three relations (*serving*, *border*, and *language*). In terms of F1, PreDic is better for four relations (*ingredient*, *origin*, *border*, and *population*), and WordNet is better for two relations (*serving* and *language*).

Relation (R.S.)	Verb Group	PreDic					WordNet				
		Ret	R.R.	Pre	Rec	F1	Ret	R.R.	Pre	Rec	F1
ingredient (1413)	simplex	1104	571	0.52	0.40	0.45	798	424	0.53	0.30	0.38
	particle	293	245	0.84	0.17	0.29	0	0	-	0	-
	verbal head	971	527	0.54	0.37	0.44	49	6	0.12	0.00	0.01
origin (636)	simplex	1298	242	0.19	0.38	0.25	111	16	0.14	0.03	0.04
	particle	86	58	0.67	0.09	0.16	3	0	0.00	0.00	0.00
	verbal head	207	117	0.57	0.18	0.28	65	7	0.11	0.01	0.02
serving (505)	simplex	1124	344	0.31	0.68	0.42	393	272	0.69	0.54	0.61
	particle	83	6	0.07	0.01	0.02	1	0	0.00	0.00	0.00
	verbal head	963	292	0.30	0.58	0.40	384	272	0.71	0.54	0.61
border (233)	simplex	259	112	0.43	0.48	0.45	184	122	0.66	0.52	0.59
	particle	15	8	0.53	0.03	0.06	0	0	-	0	-
	verbal head	153	105	0.69	0.45	0.54	55	5	0.09	0.02	0.03
language (80)	simplex	245	9	0.04	0.11	0.06	16	6	0.38	0.08	0.13
	particle	54	4	0.074	0.05	0.06	0	0	-	0	-
	verbal head	122	8	0.066	0.10	0.08	7	0	0.00	0.00	0.00
population (133)	simplex	216	7	0.03	0.053	0.04	129	6	0.05	0.045	0.05
	particle	7	1	0.14	0.01	0.01	0	0	-	0	-
	verbal head	90	3	0.03	0.023	0.27	42	2	0.05	0.015	0.02

Table 5: Application Performance of PreDic and WordNet. PreDic shows better or similar performances than WordNet (refer to Table 4 for the acronyms in the table header). The best scores for each relation are printed in **bold**.

If we compile all counts and scores of each relation into verb groups by micro average and macro average, we can get more straightforward comparisons. Micro-averaging assigns equal weight to each instance (e.g., each retrieval) regardless of classes, whereas macro-averaging assigns equal weight to each class (e.g., each predicate). Table 6 shows that PreDic is the best at recall for both micro average (i.e., 0.43) and macro average (i.e., 0.42), while NOMLEX is best at precision for both micro average (i.e., 0.77) and macro average (i.e., 0.48). However, the large difference in precision for NOMLEX between micro and macro average (i.e., 29 percent) shows that NOMLEX performs well on some nouns but not on other nouns. In contrast, PreDic provides not only the better recall and broader coverage than NOMLEX and WordNet, but also competitive macro average precision (i.e., 0.39 vs. 0.41) compared to NOMLEX or even better micro average precision (i.e., 0.60 vs. 0.52) compared to WordNet.

	Verb Group	Precision			Recall		
		PreDic	WordNet	NOMLEX	PreDic	WordNet	NOMLEX
Micro Average	simplex	0.30	0.52	0.77**	0.43	0.28	0.16**
	particle	0.60	0.00*	N/A	0.11	0.00	N/A
	verbal head	0.42	0.49	N/A	0.35	0.10	N/A
Macro Average	simplex	0.25	0.41	0.48**	0.42	0.25	0.19**
	particle	0.39	0.00*	N/A	0.07	0.00	N/A
	verbal head	0.37	0.18	N/A	0.28	0.10	N/A

Table 6: Micro and Macro Average Performances of PreDic, WordNet, and NOMLEX. Scores for WordNet (*) were calculated by using only two relations (*origin* and *servicing*) and scores for NOMLEX (**) were calculated by using only four relations. The best scores at precision and recall for each average are printed in **bold**.

The results imply that an unsupervised approach can outperform over hand-crafted resources. This also implies that unsupervised approaches can contribute to building diverse lexical resources and cover more variability of expressions in natural language as well.

5.2 Efficiency of Performance Group Model for PreDic

If we narrow the scope of performance to PreDic, we can see from the performances of PreDic in Table 5 that the verbs from the simplex group are best at recall for all relations as expected, the verbs from the particle group are best at precision for four relations (i.e., *ingredient*, *origin*, *language*, and *population*), and the verbs from the verbal head group are best at F1 for four relations (i.e., *origin*, *border*, *language*, and *population*). These results are consistent with the results in Table 6. The performances of PreDic in Table 6 show that the particle group is best at precision for micro and macro averages (i.e., 0.60 and 0.39, respectively) and the simplex group is best at recall (i.e., 0.43 and 0.42, respectively).

These results assure that our assumption for performance group model is convincing. This implies that performance group model can be adopted for implementing tasks as a precision-oriented, recall-oriented, or even harmony-oriented as needed. That is, as the variability of natural language is hard to predict, this performance group model plays a very important role in guiding applications on whether to focus on getting high-quality information at the expense of large quantities of information, large quantities at the expense of high quality, or a compromise between these two extremes given limited available time and cost. For example, when acquiring more verbs from another source, particle verb group can be used since it provides few but accurate seed verbs, whereas simplex verb group can be preferable when extracting information from texts because it offers more verbs.

6 Domain Verb Association

The relationship between nouns and inferred verbs can be measured by counting co-occurrences using web search engines (Soderland et al., 2004), and in this paper we used Google to collect the frequencies of the co-occurrences. According to Table 7, *prepare*, *contain*, *make from*, *form*, and *consist of* show

Verb	Hits w/ <i>ingredient</i>
prepare	56,100,000
contain	46,400,000
make from	37,800,000
form	33,700,000
consist of	15,500,000
attract	3,440,000
occupy with	3,420,000
display in	3,260,000
list by	2,050,000
use with	839,000
enter into	661,000

Table 7: Noun-Verb co-occurrence counts. These numbers provide conceptual relationships between the noun *ingredient* and the inferred verbs.

Verb	Hits w/ Cuisine	DVA
make from	7,452,978	1.00
consist of	1,454,047	0.20
prepare	718,686	0.10
form	406,778	0.05
use with	191,220	0.03
contain	101,411	0.01
enter into	22,749	0.00
display in	8,714	0.00
attract	4,298	0.00
list by	4,605	0.00
occupy with	9	0.00

Table 8: Domain Verb Associations (DVA) for *Cuisine* domain and the inferred verbs. The verbs with DVA above 0.00 (printed in **bold**) seem to be more associated with the *Cuisine* domain.

Verb	Hits w/ Drug	DVA
contain	1,651,933	1.00
make from	594,561	0.36
use with	490,588	0.30
consist of	54,163	0.03
form	42,963	0.03
prepare	30,366	0.02
attract	121	0.00
display in	51	0.00
enter into	12	0.00
list by	9	0.00
occupy with	0	0.00

Table 9: Domain Verb Associations (DVA) for *Drug* domain and the inferred verbs. The verbs with DVA above 0.00 (printed in **bold**) seem to be more associated with *Drug* domain.

much more co-occurrences with *ingredient* than with the other verbs (with a threshold of 10 million, for example). Although the co-occurrences do not consider the distance between two words, they must reveal the degree of relationships between the concept of nouns and their actual verbal forms in texts.

However, what matters more is how much inferred verbs are used with the words of interest rather than a noun itself. Furthermore, if we can rank preferred verbs by domains, inferred verbs can be more useful to applications that focus on specific domains. Hence we defined Domain Verb Association (DVA) to measure how frequently inferred verbs are used with domain instances that can be used as subjects or objects for the verbs. Let D denote a set of domain instances, V a set of verbs inferred from a predicate P , v_f a verb form of a (base form) verb v . Domain Verb Association measures a normalized association score for an ordered combination of a domain and a verb by summing the co-occurrences of each domain instance in the domain and each verb form of the base form of the verb:

$$DVA(D, v|P) = \frac{\sum_{d_i \in D} \sum_{v_f \in v} hits(d_i || v_f)}{\max_{v \in V} DVA(D, v|P)} \quad (1)$$

where *hits* is the number of search engine hits for query and $d_i || v_f$ is a concatenation of two words enclosed by “ and ”.

For the experiment, we defined present simple, past simple, and simple present passive voice as a set of verb forms, without taking the argument structures of the verbs into account for simplicity. We chose *hamburger*, *pasta*, and *sandwich* as a set of sample representative instances of the domain *Cuisine*. We also selected *Advil*, *Aspirin*, and *Benadryl* as a set of sample representative instances of the domain *Drug*. Consequently, queries $d_i || v_f$ were built like “*pasta makes from*”, “*pasta made from*” or “*pasta is made from*” for each combination of a domain instance and an inferred verb. DVA scores for the association of the inferred verbs from *ingredient* and the two domains (*Cuisine* and *Drug*) using Eq. (1) are shown in Table 8 and Table 9, respectively. The results show that each domain prefers some verbs to other verbs in that *make from* is the most frequent in *Cuisine* domain but *contain* is the most frequent in *Drug* domain. *Make from* is used about 70 times more often than *contain* in *Cuisine* domain, while *contain* is used about two and a half times more often than *make from* in *Drug* domain. Certainly we believe that this measure will help to improve the application performance of using PreDic.

7 Conclusion

We have presented an unsupervised approach for inferring verbs from nouns crossing root boundary and introduced a new lexical resource, *PreDic*, which is an implementation of the approach and contains verbs inferred from nouns that share neither a root nor argument structure nor a part-of-speech. We have also demonstrated a performance group model that arranges verbs into three groups is practical enough to guide applications to pursue recall-oriented, precision-oriented, or harmony-oriented results. Furthermore, the Domain Verb Association measure was introduced to show the relationships between inferred verbs and domains to which the inferred verbs are possibly applied.

Many researchers have suggested effective approaches for verb entailment acquisition and built valuable lexical resources with which the variability of natural language expression can be understood more systematically. However, unsupervised verb inference from nouns that can deliver similar meaning without shared roots has not been explicitly addressed so far. This research presents compelling evidence that the proposed approach can be a stepping stone for such applications as information extraction or natural language question answering in understanding the variability of natural language expression and recognizing such relations in text. Our future research needs to incorporate more syntactic and external knowledge, and to learn more verbs using some of the inferred verbs as seeds. Moreover, the inference over composite nouns and other parts of speech will also be investigated. Notwithstanding these future research issues, the present research findings provide clear evidence that utilizing verb inference from nouns is a fruitful textual inference approach.

Acknowledgements

This research was conducted by the International Collaborative Research and Development Program (Creating Knowledge out of Interlinked Data) and funded by the Korean Ministry of Knowledge Economy.

References

- Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barrett, and Ruth Reeves. 1998. Nomlex: A lexicon of nominalizations. In *Proceedings of EURALEX* (Vol. 98, pp. 187–193).
- Dekang Lin. 1998. Extracting collocations from text corpora. In *First workshop on computational terminology*, (pp. 57-63).
- Dekang Lin and Patrick Pantel. 2001. Discovery of inference rules for question-answering. *Natural Language Engineering*, 7(4):343–360.
- De Marneffe, Marie-Catherine, and Christopher D. Manning. 2008. Stanford typed dependencies manual. URL http://nlp.stanford.edu/software/dependencies_manual.pdf.
- Don Blaheta and Mark Johnson. 2001. Unsupervised learning of multi-word verbs. In *Proc. of the ACL/EACL 2001 Workshop on the Computational Extraction, Analysis and Exploitation of Collocations* (pp. 54-60).
- George A. Meyer. 1975. *The two-word verb: A dictionary of the verb-preposition phrases in American English* (No. 19). Walter de Gruyter.
- George A. Miller. 1995. Wordnet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.
- Idan Szpektor and Ido Dagan. 2009. Augmenting wordnet-based inference with argument mapping. In *Proceedings of the 2009 Workshop on Applied Textual Inference*.
- Ido Dagan, Bill Dolan, Bernardo Magnini, and Dan Roth. 2010. Recognizing textual entailment: Rational, evaluation and approaches. *Natural Language Engineering*, 16(01):105–105.
- Ion Androutsopoulos and Prodromos Malakasiotis. 2009. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38:135–187.
- Mark Alan Finlayson. 2013. Code for Java Libraries for Accessing the Princeton Wordnet: Comparison and Evaluation. In *Proceedings of the 7th Global Wordnet Conference*.

- Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation* (pp. 24–26). ACM.
- Nitin Madnani, Necip Fazil Ayan, Philip Resnik, and Bonnie J. Dorr. 2007. Using paraphrases for parameter tuning in statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 120–127). Association for Computational Linguistics.
- Pablo Ariel Duboue and Jennifer Chu-Carroll. 2006. Answering the question you wish they had asked: The impact of paraphrasing for question answering. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers* (pp. 33–36). Association for Computational Linguistics.
- Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, (pp. 113–120). Association for Computational Linguistics.
- Paul Bedaride and Claire Gardent. Noun/verb inference. 2009. *Human Language Technologies au a Challenge for the Computer Science and Linguistic*, 311–315.
- Ravichandran Bhagat and Deepak Ravichandran. 2008. Large scale acquisition of paraphrases for learning surface patterns. In *ACL* (Vol. 8, pp. 674–682).
- Sanda Harabagiu, George Miller, and Dan Moldovan. 1999. Wordnet 2—a morphologically and semantically enhanced resource. In *Proceedings of SIGLEX*, (vol. 99, pp. 1–8).
- Stefan Riezler, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu Mittal, and Yi Liu. 2007. Statistical machine translation for query expansion in answer retrieval. In *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS* (Vol. 45, No. 1, p. 464).
- Stephen Soderland, Oren Etzioni, Tal Shaked, and Daniel S. Weld. 2004. The use of Web-based statistics to validate information extraction. In *AAAI-04 Workshop on Adaptive Text Extraction and Mining* (pp. 21–26).
- Sunita Sarawagi. 2007. Information extraction. In *Foundations and trends in databases* (Vol. 1, No. 3, pp. 261–377).
- Yusuke Shinyama and Satoshi Sekine. 2003. Paraphrase acquisition for information extraction. In *Proceedings of the second international workshop on Paraphrasing-Volume 16* (pp. 65–71). Association for Computational Linguistics.
- Yuval Marton, Chris Callison-Burch, and Philip Resnik. 2009. Improved statistical machine translation using monolingually-derived paraphrases. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 381–390). Association for Computational Linguistics.
- Zelig S. Harris. 1981. *Distributional structure* (pp. 3–22). Springer Netherlands.

Appendix. List of verbs used for the experiment from PreDic and WordNet

PreDic				WordNet						
ingredient	servicing	language	ingredient	ingredient	ingredient	servicing	border	border	population	
attract	make	communicate by	use by	allot	neuter	vivify	serve well	evade	stick to	face up
consist of	serve of	compare	use of	alter	part	origin	service	exhibit	surround	follow
contain	take as	consist of	utter	amend	pay back	begin	suffice	fence in	take a hop	front
display in	border	convey	want	animate	pay off	blood	swear out	frame in	telephone	live
enter into	approach	create for	write	assign	buy in	wait on	wait on	frame up	throttle	look
form	arrange along	depend on	population	break up	posit	carry	border	hedge	tie down	make up
list by	border on	descend from	begin	broke up	prepare	commence	abut	hedge	tie up	man
make from	come near	describe	belong	bushel	quicken	delineate	adhere	hem in	touch	personify
need	confine within	distinguish by	cause	castrate	readily	describe	adjoin	hold fast	trammel	population
occupy with	contest	evolve	clothe	change	reanimate	draw	attach	inch	truss	present
prepare use	cross	example in	come	compensate	recompense	get down	band	jump	wall	represent
use with	define	execute	control	cook	recreate	lead off	bandage	knell	language	
origin	divide	express in	convict	define	rectify	line	beleaguer	leap	address	
bear	form	garble	define	depart	remediate	origin	besiege	limit	articulate	
begin	foster of	group of	deplete	deposit	remedy	root	bind	march	formulate	
cause by	furnish with	include of	draw	desex	renovate	rootle	bond	meet	give voice	
come from	grow up	introduce	educate	desexualise	repair	roul	border	obligate	language	
create with	indicate	involve in	entail	desexualize	resort	run along	bounce	oblige	lyric	
derive	live in	man	experience	determine	restore	set about	bound	parade	mouth	
describe	live in	name	feed	disunite	revive	set out	bunt	parade	phrase	
exist	make	originate	give	divide	revivify	settle down	butt against	parry	sound	
fix during	open	produce	go	doctor	secure	source	butt on	peal	speak	
give	plant	record	hire	factor in	separate	sprout	call up	phone	talk	
know	print	refer	increase	factor out	set forth	start out	cast	process	utter	
make	separate	related with	inhabit	fasten	set off	steady down	circumvent	put off	verbalise	
name for	settle	rely on	interbreed	fix	set out	stock up	compose	rebound	verbalize	
originate in	touch at	represent	keep down	fix	set up	stockpile	confine	recoil	vocalise	
proceed	use before	see as	live in	fixate	situate	take root	constipate	redact	vocalize	
rise	walk	set	make up	furbish up	spay	trace	contact	resile	voice	
spring into	language	speak by	occupy	fixate	specify	servicing	couch	resound	word	
start	achieve	speak in	populate	get	split up	answer	demonstrate	restrict	population	
use	arrange	speak throughout	refer in	heal	start out	assist	dodge	reverberate	comprise	
servicing	articulate by	specify	remain	indemnify	sterilise	attend to	draw up	ricochet	confront	
accept	associate by	start with	represent	ingredient	sterilize	dish out	duck	ring	constitute	
accord	base for	study of	seem	interpolate	take off	dish up	echo	set up	cost	
deliver	base on	take	take for	limit	tighten	function	elude	sidestep	earth	
eat	belong	teach	use before	make	touch on	help	ensnare	skirt	embody	
employ at	call	think		mend	unsex	process	entrap	smother	equal	
help of	combine	understand		modify	vary	serve up	environ	spring	exist	