Semantic Labeling of Compound Nominalization in Chinese

Jinglei Zhao, Hui Liu & Ruzhan Lu
Department of Computer Science
Shanghai Jiao Tong University
800 Dongchuan Road Shanghai, China
{zjl,lh_charles,rzlu}@sjtu.edu.cn

Abstract

This paper discusses the semantic interpretation of compound nominalizations in Chinese. We propose four coarse-grained semantic roles of the noun modifier and use a Maximum Entropy Model to label such relations in a compound nominalization. The feature functions used for the model are web-based statistics acquired via role related paraphrase patterns, which are formed by a set of word instances of prepositions, support verbs, feature nouns and aspect markers. By applying a sub-linear transformation and discretization of the raw statistics, a rate of approximately 77% is obtained for classification of the four semantic relations.

1 Introduction

A nominal compound (NC) is the concatenation of any two or more nominal concepts which functions as a third nominal concept (Finin, 1980). (Leonard, 1984) observed that the amount of NCs had been increasing explosively in English in recent years. NCs such as *satellite navigation system* are abundant in news and technical texts. In other languages such as Chinese, NCs have been more productive since earlier days as evidenced by the fact that many simple words in Chinese are actually a result of compounding of morphemes.

Many aspects in Natural Language Processing (NLP), such as machine translation, information retrieval, question answering, etc. call for the automatic interpretation of NCs, that is, making explicit the underlying semantic relationships between the constituent concepts. For example, the semantic relations involved in *satellite communication system* can be expressed by the conceptual graph (Sowa, 1984) in Figure 1, in which, for instance, the semantic relation between *satellite* and *communication* is MANNER. Due to the productivity of NCs and the lack of syntactic clues to guide the interpretation process, the automatic interpretation of NCs has been proven to be a very difficult problem in NLP.

In this paper, we deal with the semantic interpretation of NCs in Chinese. Especially, we will focus on a subset of NCs in which the head word is a verb nominalization. Nominalization is a common phenomenon across languages in which a predicative expression is transformed to refer to an event or a property. For example, the English verb *communicate* has the related nominalized form *communication*. Different from English, Chinese has little morphology. Verb nominalization in Chinese has the same form as the verb predicate.

Nominalizations retain the argument structure of the corresponding predicates. The semantic relation between a noun modifier and a verb nominalization head can be characterized by the semantic role the modifier can take respecting to the corresponding verb predicate. Our method uses a Maximum Entropy model to label coarse-grained semantic roles in Chinese compound nominalizations. Unlike most approaches in compound interpretation and semantic role labeling, we don't exploit features from any parsed texts or lexical knowledge sources. Instead, features are acquired using web-based statis-

Proceedings of the Workshop on A Broader Perspective on Multiword Expressions, pages 73–80,
Prague, June 2007.

[satellite]←(MANNER)←[communication]←(TELIC)←[system]

Figure 1: The conceptual graph for *satellite communication system*

tics (PMI-IR) produced from paraphrase patterns of the compound Nominalization.

The remainder of the paper is organized as follows: Section 2 describes related works. Section 3 describes the semantic relations for our labeling task. Section 4 introduces the paraphrase patterns used. Section 5 gives a detailed description of our algorithm. Section 6 presents the experimental result. Finally, in Section 7, we give the conclusions and discuss future work.

2 Related Works

2.1 Nominal Compound Interpretation

The methods used in the semantic interpretation of NCs fall into two main categories: rule-based ones and statistic-based ones. The rule-based approaches such as (Finin, 1980; Mcdonald, 1982; Leonard, 1984; Vanderwende, 1995) think that the interpretation of NCs depends heavily on the constituent concepts and model the semantic interpretation as a slot-filling process. Various rules are employed by such approaches to determine, for example, whether the modifier can fill in one slot of the head.

The statistic-based approaches view the semantic interpretation as a multi-class classification problem. (Rosario and Hearst, 2001; Moldovan et al., 2004; Kim and Baldwin, 2005) use supervised methods and explore classification features from a simple structured type hierarchy. (Kim and Baldwin, 2006) use a set of seed verbs to characterize the semantic relation between the constituent nouns and explores a parsed corpus to classify NCs. (Turney, 2005) uses latent relational analysis to classify NCs. The similarity between two NCs is characterized by the similarity between their related pattern set.

(Lauer, 1995) is the first to use paraphrase based unsupervised statistical models to classify semantic relations of NCs. (Lapata, 2000; Grover et al., 2005; Nicholson, 2005) use paraphrase statistics computed from parsed texts to interpret compound nominalization, but the relations used are purely syntactic. Lapata(2000) only classifies syntactic relations of subject and object. Grover(2005) and Nicholson (2005) classify relations of subject, object and prepositional object.

2.2 Semantic Role Labeling of Nominalization

Most previous work on semantic role labeling of nominalizations are conducted in the situation where a verb nominalization is the head of a general noun phrase. (Dahl et al., 1987; Hull and Gomez, 1996) use hand-coded slot-filling rules to determine the semantic roles of the arguments of a nominalization. In such approaches, first, parsers are used to identify syntactic clues such as prepositional types. Then, rules are applied to label semantic roles according to clues and constraints of different roles.

Supervised machine learning methods become prevalent in recent years in semantic role labeling of verb nominalizations as part of the resurgence of research in shallow semantic analysis. (Pradhan et al., 2004) use a SVM classifier for the semantic role labeling of nominalizations in English and Chinese based on the FrameNet database and the Chinese PropBank respectively. (Xue, 2006) uses the Chinese Nombank to label nominalizations in Chinese. Compared to English, the main difficulty of using supervised method for Chinese, as noted by Xue (2006), is that the precision of current parsers of Chinese is very low due to the lack of morphology, difficulty in segmentation and lack of sufficient training materials in Chinese.

2.3 Web as a large Corpus

Data sparseness is the most notorious hinder for applying statistical methods in natural language processing. However, the World Wide Web can be seen as a large corpus. (Grefenstette and Nioche, 2000; Jones and Ghani, 2000) use the web to generate corpora for languages for which electronic resources are scarce. (Zhu and Rosenfeld, 2001) use Web-based n-gram counts for language modeling. (Keller and Lapata, 2003) show that Web page counts and n-gram frequency counts are highly correlated in a log scale.

3 Semantic Relations

Although verb nominalization is commonly considered to have arguments as the verb predicate, Xue(2006) finds that there tend to be fewer arguments and fewer types of adjuncts in verb nominalizations compared to verb predicates in Chinese. We argue that this phenomenon is more obvious in compound nominalization. By analyzing a set of compound nominalizations of length two from a balanced corpus(Jin et al., 2003), we find the semantic relations between a noun modifier and a verb nominalization head can be characterized by four coarse-grained semantic roles: Proto-Agent (PA), Proto-Patient (PP), Range (RA) and Manner (MA). This is illustrated by Table1.

Relations	Examples	
PA	血液 循环	(Blood Circulation)
	鸟类 迁徙	(Bird Migration)
PP	企业 管理	(Enterprise Management)
	动物 分类	(Animal Categorization)
MA	激光 存储	(Laser Storage)
	卫星 通信	(Satellite Communication)
RA	全球 定位	(Global Positioning)
	长期 发展	(Long-time Development)

Table 1: Semantic Relations between Noun Modifier and Verb Nominalization Head.

Due to the linking between semantic roles and syntactic roles (Dowty, 1991), the relations above overlap with syntactic roles, for example, Proto-Agent with Subject and Proto-Patient with Object, but they are not the same, as illustrated by the example 动物分类(*Animal Categorization*). Although the predicate 分类(*categorize*) in Chinese is an intransitive verb, the semantic relation between 动物(*animal*) and 分类(*categorization*) is Proto-Patient.

4 Paraphrase Patterns

4.1 Motivations

Syntactic patterns provide clues for semantic relations (Hearst, 1992). For example, Hearst(1992) uses the pattern "NP such as List" to indicate that nouns in List are hyponyms of NP. To classify the four semantic relations listed in section 3, we propose some domain independent surface paraphrase patterns to characterize each semantic relation. The patterns we adopted mainly exploit a set of word instances of prepositions, support verbs, feature nouns and aspect markers.

Prepositions are strong indicators of semantic roles in Chinese. For example, in sentence 1), the preposition 把(*ba*) indicates that the noun 门(*door*) and 张三(*Zhangsan*) is the Proto-Patient and Proto-Agent of verb 锁(*lock*) respectively.

1) a. 张三把门锁上
   b. *Zhangsan ba door locked.*
   c. *Zhangsan locked the door.*

The prepositions we use to characterize each relation are listed in table 2.

Relations	Prepositional Indicators
PP	被(bei), 让(rang), 叫(jiao), 由(you)
PA	把(ba), 将(jiang), 所(suo), 对(dui)
MA	通过(tongguo), 用(yong), 以(yi)
RA	在(zai), 于(yu), 从(cong)

Table 2: Prepositional indicators of different relations in Chinese.

Support verbs such as 进行(*conduct*), 加以(*put-to*) can take verb nominalizations as objects. When combined with prepositions, they could be good indicators of semantic roles. For example in 2), the verb 进行(*conduct*) together with the preposition 对(*dui*) indicate that the relation between 分类(*categorization*) and 动物(*animal*) is PA.

2) a. 对动物进行分类
   b. *dui animal conduct categorization.*
   c. *conduct categorization regarding animal.*

Nouns such as 方法(*method*), 方式(*manner*), 范围(*range*) and 地点(*place*) can be used as features when co-occurring with the compound nominalizations under consideration. For example, if 全球范围(*global range*) co-occurs frequently with 定位(*positioning*), it will indicate a possible RA relation between 全球(*global*) and 定位(*positioning*).

Another set of word instances we use is aspect, tense and modal markers. As we have mentioned, verb nominalizations have the same form as the corresponding verb predicates in Chinese. Aspect，tense and modal markers make a good indicator for recognizing a verb predicate. For example if a verb is directly followed by an aspect marker such as 了(*le*), which indicates a finished state, it could be safely viewed as a predicate. Such markers are very useful in paraphrase patterns. This can be illustrated by 3), in which, the tense marker 开始(*start*) indicates a strong agentive meaning of the noun 鸟类(*bird*) and provides good clues of the relation PP between鸟类(*bird*) and 迁徙(*migration*) in the compound 鸟类迁徙(*bird migration*).

3) a. 鸟类开始迁徙
   b. *Bird start migrate.*
   c. *Birds start to migrate.*

4.2 Paraphrase Pattern Templates

We use the set of word instances above to form pattern templates which could be instantiated by the compound nominalization under consideration to form paraphrase patterns. The templates are expressed using the employed search engine's query language. Currently, we employ totally 30 feature templates for the four semantic relations. A sample of the pattern templates is listed in Tabel 3, in which, x, y is the variable which need to be instantiated by the noun modifier and verb nominalization respectively.

Relations	Paraphrase Pattern Templates
PP	”对x进行y” (”dui x $conduct$ y”)
	”把x” ”y” (”ba x” ”y”)
	”y着x” (”y zhe x”)
	”x被” ”y” (”x bei” ”y”)
PA	”被x” ”y” (”bei x” ”y”)
	”x开始y” (”x $start$ y”)
	”x” ”可以y” (”x” ”can y”)
	”x所y” (”x suo y”)
MA	”通过x” ”y” -”通过xy”
	(”$tongguo$ x” ”y” -”$tongguo$ xy”)
	”x方法” ”y” (”x $method$” ”y”)
RA	”在x” ”y” -”在y”(”zai x” ”y” -”zai y”)
	”从x” ”y” (”$cong$ x” ”y”)
	”x范围” ”y” (”x $range$” ”y”)

Table 3: A Sample Set of the Paraphrase Pattern Templates.

5 System Description

5.1 Data Source

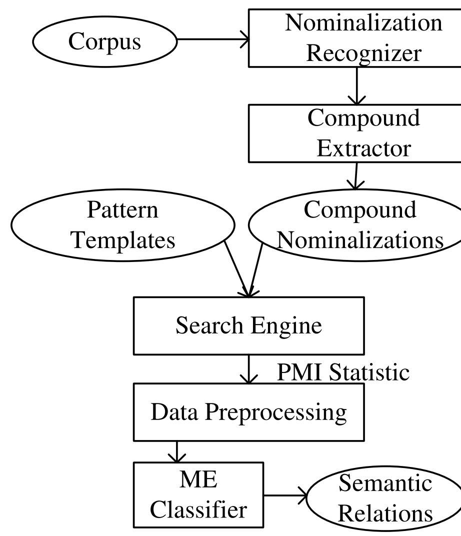

Figure 2: System Architecture

Figure 2 illustrates the system architecture of our approach. We view the semantic labeling of compound nominalization as a data-driven classification problem. The data used for the experiment is auto-extracted from the Chinese National Corpus (Jin et al., 2003), which is a balanced segmented and POS tagged corpus with 8M characters. Because the corpus doesn't distinguish verb predicates with verb nominalizations, a verb nominalization recognizer is first used to recognize all the verb nominalizations in the corpus, and then, a compound extractor identifies all the compound nominalizations having a noun modifier and a verb nominalization head in the corpus. We manually examined a sample of the result set and finally randomly select 300 correct noun-nominalization pairs as our training and testing set for semantic interpretation.

One PHD student majored in computer science and one in linguistics were employed to label all the 300 data samples simultaneously according to the relation set given in section 3. The annotator's agreement was measured using the Kappa statistic (Siegel and Castellan, 1988) illustrated in (1), of which $Pr(A)$ is the probability of the actual outcome and $Pr(E)$ is the probability of the expected outcome as predicted by chance. The Kappa score of the annotation is 87.3%.

$$K = \frac{Pr(A) - Pr(E)}{1 - Pr(E)} \quad (1)$$

After discussion, the two annotators reached agreement on a final version of the data sample labeling. In which, the proportion of relations PP, PA, MA, RA is 45.6%, 27.7%, 16.7% and 10% respectively, giving a baseline of 45.6% of the classification problem by viewing all the relations to be PP. Finally, the 300 data instances were partitioned into a training set and a testing set containing 225 and 75 instances respectively.

5.2 Maximum Entropy Model

We use the Maximum Entropy (ME) Model (Berger et al., 1996) for our classification task. Given a set of training examples of a random process, ME is a method of estimating the conditional probability $p(y|x)$ that, given a context x, the process will output y. In our task, the output corresponds to the four relation labels PP, PA, MA and RA.

The modeling of ME is based on the Maximum Entropy Principle, that is, modeling all that is known and assuming nothing about what is unknown. The computation of $p(y|x)$ is illustrated as the formula (2). $f_i(x, y)$ are binary valued feature functions with the parameter λ_i used to express the statistics of the data sample. $Z_\lambda(x)$ is a normalization factor.

$$p_\lambda(y|x) = \frac{1}{Z_\lambda(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right) \quad (2)$$

5.3 PMI-IR Score as Features

The feature functions we adopted for ME differentiate from most other works on the semantic labeling task, which mainly exploited features from well-parsed text. Instead, we use a web-based statistic called PMI-IR which mainly measures the co-occurrence between the data to classify and the set of paraphrase pattern templates we stated in section 4. The PMI-IR measure was first adopted by (Turney, 2001) for mining synonyms from the Web. (Etzioni et al., 2004) uses the PMI-IR measure to evaluate the information extracted from the Web.

Given a compound nominalization pair $p(x, y)$ and a set of paraphrase pattern templates t_1, t_2, ···, t_n, the PMI-IR score between p and t_i can be computed by formula (3).

$$PMI(p, t_i) = \frac{Hits(p, t_i)}{Hits(p)} \quad (3)$$

In which, $PMI(p, t_i)$ is the co-occurrence web page counts of $p(x, y)$ and t_i. For example, if the template t is "对(*dui*) x 进行(*conduct*) y" and the compound nominalization is the pair p(动物(*animal*),分类(*categorization*)), then $Hits(p, t)$ is the web counts returned from the search engine for the pattern "对(*dui*) 动物(*animal*) 进行(*conduct*) 分类(*categorization*)".

5.4 Scaling of PMI Features

Web counts are inflated which need to be scaled to attain a good estimation of the underlying probability density function in ME. In our approach, first, a log sub-linear transformation is used to preprocess the raw PMI-IR feature function for the ME model. Then, a discretization algorithm called CAIM (Kurgan and Cios, 2004) is used to transform the continuous feature functions into discrete ones.

CAIM is a supervised discretization algorithm which can discretize an attribute into the smallest number of intervals and maximize the class-attribute interdependency. Suppose that the data set consists of M examples and each example belongs to only one of the S classes. F indicates the continuous feature functions produced from paraphrase patterns in our task. D is a discretization scheme on F, which discretizes F into n non-overlapping discrete intervals. The class variable and the discretization variable of attribute F are treated as two random varibles defining a two-dimensional frequency matrix(called quanta matrix) that is shown in Table 4, in which, q_{ir} is the total number of continuous values belonging to the i^{th} class that are within interval $(d_{r-1}, d_r]$, while M_{i+} is the total number of values belonging to the i^{th} class, and M_{+r} is the total number of values of attribute F that are within the interval $(d_{r-1}, d_r]$, for $i = 1, 2, ..., S$ and $r = 1, 2, ..., n$. The CAIM algorithm uses a greedy search to find the specific discretization sechme D according to the Class-Attribute Interdependency Maximization (CAIM) criterion defined as(4), where max_r is the maximum value among all q_{ir} values.

Class	$[d_0,d_1]$	...	$[d_{r-1},d_r]$	...	$[d_{n-1},d_n]$	Class Total
C_1	q_{11}	...	q_{1r}	...	q_{1n}	M_{1+}
:	:	...	:	...	:	:
C_i	q_{i1}	...	q_{ir}	...	q_{in}	M_{i+}
:	:	...	:	...	:	:
C_s	q_{S1}	...	q_{Sr}	...	q_{Sn}	M_{S+}
Interval Total	M_{+1}	...	M_{+r}	...	M_{+n}	M

Table 4: The Quanta Matrix for Attribute F and Discretization Scheme D

$$CAIM(C,D|F) = \frac{1}{n}\sum_{r=1}^{n}\frac{\max_r^2}{M_{+r}} \qquad (4)$$

6 Results and Discussion

In this section, we present our experimental results on the semantic relation labeling of our Compound Nominalization Dataset. We compared the performance between two different engines, also, between the raw PMI and the scaled one.

Two search engines, Google (www.google.com) and Baidu (www.baidu.com) are used and compared to obtain the PMI scores between a verb nominalization pair and the set of paraphrase patterns. The result of using Google and Baidu are comparable. For example, when using raw PMI score as the features of ME classification model, Google based algorithm obtains a correct classification rate of 65.3%, while Baidu based algorithm obtains a correct classification rate of 62.7%. The main difference between the two search engines is their indexing and rating algorithm of the web pages. Compared to Google, Baidu uses a stop wordlist, including empty markers such as 了(*le*), to filter the queries. While this is beneficial for common users, it hurts our algorithm which depends heavily on such information.

Compared with using raw PMI as the classification features, feature scaling improves much on the classification result. Using Log transformation, Both Google based and Baidu based algorithm increase about 4 percent on the correct classification rate and when CAIM algorithm is employed to preprocess the data, both algorithm's correct classification rates increase more than 8 percent. We think that the usefulness of log sub-linear transformation is mainly due to the fact that the Web is extremely biased and inflated. The compression of the inflated feature space can enable the ME model to give a good estimation of the underlying probability density function of the data. As to the usefulness of the discretization of the data, we think that it is mainly because that the web-based statistics contain much noise and the features produced from paraphrase patterns are highly correlated with specific classes. CAIM discretization algorithm can maximize the class-attribute interdependence in the data and can be seen as a noise pruning process in some sense.

Among the four semantic relations labeled, PP gets the best precision and recall overall and relations such as RA gets a lower F-score. We think that this is mainly due to the difficulty in selecting paraphrase patterns for RA compared to PP. Some patterns are not as indicative as others for the relations considered. For example, the paraphrase patterns "在x" "y" -"在y" ("in x" "y" -"in y") for RA is not as indicative as the pattern "对x进行y" (*dui x conduct y*) for PP. Discovering and selecting the most indicative patterns for each relation is the key element for our algorithm.

We can make a rough comparison to the related works in the literature. In syntactic relation labeling of compound nominalization in English, Lapata (2000) and Grover et al. (2005) both apply parsed text and obtains 87.3%, 77% accuracy for the subject-object and subject-object-prepositional objects classification tasks respectively. Nicholson (2005) uses both the parsed text and the web for the classification of subject-object-prepositional objects and the result is comparatively poor. Compared to such works, the relations we exploited in the labeling task is purely semantic which makes the classification task more difficult and we don't use any parsed text as input. Considering the difficulty of

	Google			Baidu		
	Precision	Recall	F-Score	precision	Recall	F-Score
			Raw PMI			
PP	72.5	82.9	77.3	65.3	88.9	75.2
PA	47.6	50.0	48.8	50.0	42.1	45.7
MA	75.0	50.0	60.0	50.0	27.3	35.3
RA	66.7	50.0	57.1	80.0	44.4	57.1
Rate	65.3			62.7		
			Log			
PP	66.7	85.7	75.0	68.2	83.3	75.0
PA	64.7	55.0	59.5	60.0	47.4	52.9
MA	80.0	66.7	72.7	66.7	54.5	60.0
RA	100	37.5	54.5	71.4	55.5	62.5
Rate	69.3			66.7		
			Log+Discretization			
PP	82.5	94.3	88.0	80.9	94.4	87.2
PA	81.3	65.0	72.2	64.7	57.9	61.1
MA	75.0	50.0	60.0	87.5	63.6	73.7
RA	54.5	75.0	63.2	64.5	55.6	58.8
Rate	77.3			76.0		

Table 5: Results comparing different search engines, raw PMI as features vs. scaled features. Rate is the correct classification rate for the four semantic relations overall.

the problem and the unsupervised nature of our algorithm, the results (accuracy 77.3%) are very encouraging.

7 Conclusions and Future Work

In this paper, we view the semantic relation labeling of compound nominalization as a classification problem. We propose four coarse-grained semantic roles of the noun modifier for the verb nominalization head. A Maximum Entropy model is applied for the classification task. The features used for the model are web-based statistics acquired via class related paraphrase patterns, which mainly use a set of word instances of prepositions, support verbs, feature nouns and aspect markers. The experimental result illustrates that our method is very effective.

We believe that the method we proposed is not only limited in the semantic interpretation of compound nominalizations, but can also be used as a way to compensate the low accuracy of the more general task of semantic role labeling of nominalization phrases caused by the inefficiency of Chinese parsers.

The major limitation of our approach is that the paraphrase pattern templates we use now are hand-coded according to the linguistic theory. To achieve more generality of our method, in the future, we should study automatic template induction and feature selection algorithms for the classifier to select the set of most indicative pattern templates for each semantic relation.

8 Acknowledgements

This work is supported by NSFC Major Research Program 60496326: Basic Theory and Core Techniques of Non Canonical Knowledge.

References

A.L. Berger, V.J. Della Pietra, and S.A. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

D.A. Dahl, M.S. Palmer, and R.J. Passonneau. 1987. Nominalizations in PUNDIT. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics, Stanford University, Stanford, CA, July.*

DR Dowty. 1991. Thematic ProtoRoles and Argument Selection. *Second Conference on Maritime Termonology, Turku*, 33:31–38.

O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.M. Popescu, T. Shaked, S. Soderland, D.S. Weld, and A. Yates. 2004. Web-scale information extraction in knowitall:(preliminary results). *Proceedings of the 13th international conference on World Wide Web*, pages 100–110.

T.W. Finin. 1980. The semantic interpretation of compound nominals. *Dissertation Abstracts International Part B: Science and Engineering[DISS. ABST. INT. PT. B- SCI. & ENG.],*, 41(6):1980.

G. Grefenstette and J. Nioche. 2000. Estimation of English and non-English Language Use on the WWW. *Arxiv preprint cs.CL/0006032*.

C. Grover, A. Lascarides, and M. Lapata. 2005. A comparison of parsing technologies for the biomedical domain. *Natural Language Engineering*, 11(01):27–65.

M.A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545.

R.D. Hull and F. Gomez. 1996. Semantic interpretation of nominalizations. *AAAI Conference*, pages 1062–1068.

Guangjin Jin, Shulun Guo, Hang Xiao, and Yunfan Zhang. 2003. Standardization for Corpus Processing. *Applied Linguistics*, pages 16–24.

R. Jones and R. Ghani. 2000. Automatically building a corpus for a minority language from the web. *Proceedings of the Student Research Workshop at the 38th Annual Meeting of the Association for Computational Linguistics*, pages 29–36.

F. Keller and M. Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics*, 29(3):459–484.

S.N. Kim and T. Baldwin. 2005. Automatic interpretation of noun compounds using WordNet similarity. *Proc. of IJCNLP-05*, pages 945–956.

S.N. Kim and T. Baldwin. 2006. Interpreting Semantic Relations in Noun Compounds via Verb Semantics. *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 491–498.

LA Kurgan and KJ Cios. 2004. CAIM discretization algorithm. *Knowledge and Data Engineering, IEEE Transactions on*, 16(2):145–153.

M. Lapata. 2000. The automatic interpretation of nominalizations. *Proceedings of AAAI*.

M. Lauer. 1995. *Designing Statistical Language Learners: Experiments on Compound Nouns*. Ph.D. thesis, Ph. D. thesis, Macquarie University, Sydney.

R. Leonard. 1984. *The interpretation of English noun sequences on the computer*. North-Holland.

D.B. Mcdonald. 1982. Understanding noun compounds. *Carnegie-Mellon University*.

D. Moldovan, A. Badulescu, M. Tatu, D. Antohe, and R. Girju. 2004. Models for the semantic classification of noun phrases. *Proceedings of HLT/NAACL-2004 Workshop on Computational Lexical Semantics*.

J. Nicholson. 2005. *Statistical Interpretation of Compound Nouns*. Ph.D. thesis, University of Melbourne.

S. Pradhan, H. Sun, W. Ward, J.H. Martin, and D. Jurafsky. 2004. Parsing Arguments of Nominalizations in English and Chinese. *Proc. of HLT-NAACL*.

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

S. Siegel and NJ Castellan. 1988. Nonparametric statistics for the behavioral sciences. *McGraw-HiU Book Company, New York*.

JF Sowa. 1984. *Conceptual structures: information processing in mind and machine*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

P.D. Turney. 2001. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the Twelfth European Conference on Machine Learning*, pages 491–502.

P.D. Turney. 2005. Measuring semantic similarity by latent relational analysis. *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 1136–1141.

L.H. Vanderwende. 1995. *The analysis of noun sequences using semantic information extracted from online dictionaries*. Ph.D. thesis, Georgetown University.

N. Xue. 2006. Semantic Role Labeling of Nominalized Predicates in Chinese. *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*.

X. Zhu and R. Rosenfeld. 2001. Improving trigram language modeling with the World Wide Web. *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, 1.