Learning Semantic Relationships in Compound Nouns with Connectionist Networks

Stefan Wermter
Department of Computer and Information Science
University of Massachusetts

Abstract

This paper describes a new approach for understanding compound nouns. Since several approaches have demonstrated the difficulties in finding detailed and suitable semantic relationships within compound nouns, we use only a few basic semantic relationships and provide the system with the additional ability to learn the details of these basic semantic relationships from training examples. Our system is based on a backpropagation architecture and has been trained to understand compound nouns from a scientific technical domain. The test results demonstrated that a connectionist network is able to learn semantic relationships within compound nouns.

Introduction

Understanding compound nouns plays an important role in understanding natural language. In the past, different approaches for understanding compound nouns have been investigated in artificial intelligence, linguistics, and cognitive science ((Marcus 80) (Finin 80) (McDonald 82) (Lehnert 86) (Arens 87) (Dahl 87)). Most approaches relied on a representation of the words in compound nouns as frames or semantic features and contained fixed control structures which determined the semantic relationships between the words. For example, Finin (Finin 80) used frames to predict the semantic relationships between words and a hierarchy of rules to identify the best relationship. McDonald's system (McDonald 82) is based on Fahlman's parallel semantic network (Fahlman 79) and used marker passing to find the semantic relationships between word concepts.

These approaches try to understand compound nouns by coding as much knowledge as possible about the words, semantic relationships, and control structures. In this paper we investigate a different approach for understanding compound nouns consisting of two words. We use only a few basic semantic relationships and provide the system with the ability to learn the details of the basic semantic relationships from training examples. Instead of encoding knowledge structures and control structures for understanding compound nouns, basic semantic relationships in compound nouns are learned using a connectionist architecture.

The Domain and the Basic Semantic Relationships

Compound nouns are frequently used in almost every domain. Our domain is the NPL[1] corpus (Sparck-Jones 76) which contains abstracts and queries from the physical sciences. From this corpus

[1] National Physics Laboratory

we randomly chose 108 compound nouns consisting of two words, e.g. "heat effect". Each word is represented as a binary vector of 16 semantic features, which were extracted by using the NASA thesaurus (NASA 85). For a more detailed description of the process of feature extraction see (Wermter and Lehnert 89). Figure 1 illustrates the semantic features for the compound nouns.

Semantic Features	Examples
MEASURING-EVENT	Observation, Investigation, Research
CHANGING-EVENT	Amplification, Acceleration, Loss
SCIENTIFIC-FIELD	Mechanics, Ferromagnetics
PROPERTY	Intensity, Viscosity, Temperature
MECHANISM	Experiment, Technique, Theorem
ELECTRIC-OBJECT	Transistor, Resistor, Amplifier
PHYSICAL-OBJECT	Earth, Crystal, Vehicle, Room
RELATION	Cause, Dependence, Interaction
ORGANIZATION-FORM	Layer, Level, Stratification, F-Region
GAS	Air, Oxygen, Atmosphere, Nitrogen
SPATIAL-LOCATION	Antarctic, Earth, Range, Region, Source
TIME	June, Day, Time, History
ENERGY	Radiation, Ray, Light, Sound, Current
MATERIAL	Aluminium, Water, Carbon, Vapour
ABSTRACT-REPRESENTATION	Note, Data, Equation, Term, Parameter
EMPTY	Cavity, Vacua

Figure 1: Semantic Features of the Nouns and Examples

To represent basic semantic associations between words we use 7 basic semantic relationships. We specify a **Basic Semantic Relationship** as a preposition paraphrase (see figure 2). For example, a "room experiment" has the basic semantic relationship IN-P since the experiment is "in" a room, and an "excitation mechanism" has the basic semantic relationship FOR-P since it is a mechanism "for" excitation. Each compound noun can have different basic semantic relationships; for instance, a "feedback circuit" is a "circuit FOR-P feedback" or a "circuit WITH-P feedback". Each basic semantic relationship can have several meanings; for instance, the IN-P is different for "storage IN-P computer" and "disturbance IN-P atmosphere".

Basic Semantic Relationships	Examples for the Basic Semantic Relationships
BY-P	Impurity Conduction
FOR-P	Excitation Mechanism
FROM-P	Space Vehicle
IN-P	Room Experiment
OF-P	Oxygen Emission
ON-P	Skin Effect
WITH-P	Amplifier Circuit

Figure 2: The Basic Semantic Relationships

We consider the basic semantic relationships as a first step to differentiate semantic relationships according to their main properties. This general concept of classifying semantic relationships according to preposition paraphrases has been found useful in several studies on compound nouns (e.g., (Lee 60) (Levi 78) (Finin 80)), since preposition paraphrases contain general relationships; e.g., FROM-P expresses a source, FOR-P expresses a purpose, and IN-P expresses inclusion. Our goal here is to specify basic semantic relationships as preposition paraphrases and to build a system which learns the underlying semantic relationships from training examples.

The Architecture

The architecture for learning semantic relationships is a backpropagation network with three layers (see figure 3). The bottom layer consists of 32 binary input units for the semantic features of the two words in the compound noun. The hidden layer is a 7 x 12 array of hidden units, 12 hidden units for each of the 7 basic semantic relationships. The top layer consists of 7 real-valued output units, one for each of the 7 basic semantic relationships.

Each output unit is connected only to all hidden units of the same basic semantic relationship. All hidden units are connected to all input units. This modular organization has two advantages: (1) training and testing for each basic semantic relationship can be done independently, and (2) adding, deleting and modifying a basic semantic relationship does not require retraining the whole network.

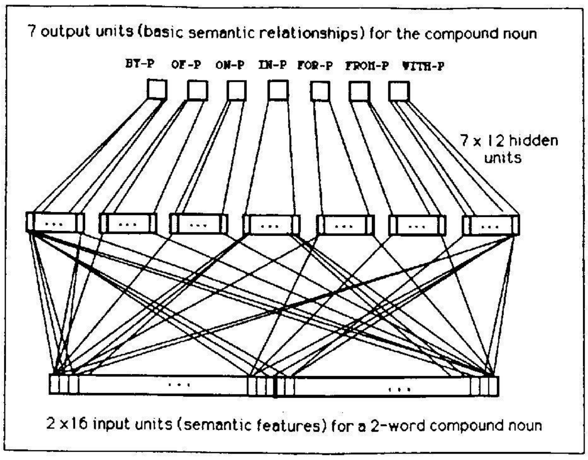

Figure 3: The Structure of the Backpropagation Network

Training the Network

First, 108 compound nouns consisting of two words were randomly selected from the NPL corpus. Each compound noun was represented with 32 binary features, 16 for each word. The 108 compound nouns were divided into 88 compound nouns for a training set and 20 compound nouns for a test set. Because of the modular architecture the network can be trained in separate modules for the different basic semantic relationships. For each of the 7 basic semantic relationships the feature representations of the 88 compound nouns were presented as the input together with a desired binary plausibility value as the output. The plausibility value indicates if the basic semantic relationship between the two words is plausible (value 1) or not plausible (value 0). The following example shows two of the 88 training examples for the basic semantic relationship IN-P: "Plasma layer" in the sense of "layer IN-P plasma" is plausible, while "sunspot number" in the sense of "number IN-P sunspot" is not plausible.

```
PLASMA LAYER      ->    LAYER IN-P PLASMA       1
SUNSPOT NUMBER    ->    NUMBER IN-P SUNSPOT     0
```

For each of the 7 basic semantic relationships the semantic features and plausibility values of the 88 compound nouns were presented for 800 cycles (that is 70400 training examples). The backpropagation algorithm (Rumelhart et. al. 86) was used to learn the plausibility of each basic semantic relationship[2]. To be independent of the random start initialization of the network, three different runs (each with the 70400 training examples) were conducted for each of the 7 basic semantic relationships. Within this learning phase the average of the total sum squared error for all training examples over all 21 runs decreased from 23.2 at the start of the training to 1.4 at the end of the training.

Evaluation of the Test Results

After training, the network was tested on the training set of 88 compound nouns and the test set of 20 compound nouns. The semantic feature representation of the compound nouns in the test set had not been part of the training set. The network was tested by presenting the feature representation of a compound noun, and the system computed the plausibility value for each basic semantic relationship. A basic semantic relationship is considered correct, if the computed plausibility value deviates less than 0.49 from the desired value 1 for a plausible basic semantic relationship and from the desired value 0 for an implausible basic semantic relationship.

[2]The learning rate η was set to 0.01, the weight change momentum α was 0.9 for all experiments.

Basic Semantic Relationships	Correct in the Training Set	Correct in the Test Set
BY-P	94%	83%
FOR-P	97%	73%
FROM-P	94%	82%
IN-P	96%	73%
OF-P	93%	77%
ON-P	98%	95%
WITH-P	98%	88%

Figure 4: Basic Semantic Relationships in Training Set and Test Set

Figure 4 illustrates the overall system performance on the training set and on the test set for each basic semantic relationship. The average percentage of correctly learned training examples for the three different learning runs is between 93% and 98%, the percentage of correctly generalized test examples is between 73% and 95%.

Figure 5 shows a more detailed interpretation of representative examples from the test set of new compound nouns. Each compound noun is shown with the computed plausibility values for each basic semantic relationship[3]. We say that a basic semantic relationship for a compound noun exists if the computed plausibility value is greater than or equal to 0.5.

Compound Nouns	BY-P	FOR-P	FROM-P	IN-P	OF-P	ON-P	WITH-P
Heat Exchange	0.3	0.0	0.0	0.3	0.9	0.0	0.0
Transistor Life	0.0	0.4	0.0	0.1	1.0	0.0	0.1
Writing Method	0.0	0.8	0.1	0.0	1.0	0.0	0.0
Wing Motion	0.0	0.1	0.3	1.0	0.5	0.0	0.0
Waveform Solution	0.1	0.4	0.0	0.1	0.4	0.1	0.0
Earth Satellite	0.0	0.0	0.9	0.9	0.0	0.0	0.7
Transport Theory	0.3	0.1	0.0	0.0	0.9	0.6	0.1
Water Vapour	0.0	0.0	0.6	1.0	0.4	0.0	0.6
Wave Propagation	0.1	0.0	0.2	0.7	0.7	0.1	0.0
Microwave Emission	0.7	0.1	0.0	0.0	1.0	0.0	0.0

Figure 5: Examples for the Interpretation of Compound Nouns (see text for explanation)

In the first two examples in figure 5 a single basic semantic relationship exists between the two words in the compound noun: "heat exchange" is interpreted as "exchange OF-P heat", and "transistor life" as "life OF-P transistor" (only these basic semantic relationships have a plausibility value greater or equal 0.5).

[3] Again, as in figure 4, the plausibility values shown are the averages over the three different runs for each basic semantic relationships.

Although only one basic semantic relationship exists in the first two examples, most test examples have more than one existing basic semantic relationship. For instance, "writing method" has the existing relationships "method FOR-P writing" and "method OF-P writing". The other basic semantic relationships for "writing method", like "method BY-P writing" and "method FROM-P writing", do not exist. Another example of multiple basic semantic relationships is "wing motion" (as in airplanes) which is interpreted as "motion OF-P wing" and "motion IN-P wing". This example illustrates ambiguous interpretations and context is needed to determine if the wing is the object which is moving (motion OF-P wing) or the location of a motion (motion IN-P wing).

The plausibilty values in Figure 5 indicate unsure interpretations as well. For instance, the plausibility values of the compound noun "waveform solution" are lower than 0.5 for all basic semantic relationships. The network can not find a basic semantic relationship because similar relationships had not been in the training set. The results show examples with some incorrect basic semantic relationships as well. For instance "water vapour" is interpreted with 3 existing relationships: "vapour FROM-P water", "vapour WITH-P water", and "vapour IN-P water". While the first two relationships FROM-P and WITH-P are plausible, the third is not plausible.

Although our corpus is still fairly small our test results demonstrate the extent to which the learned basic semantic relationships generalize for new compound nouns. The basic semantic relationships in our network generalize well for compound nouns whose first and second noun are characterized with subsets of the following semantic features: Noun1: ENERGY PROPERTY ORGANIZATION-FORM and Noun2: CHANGING-EVENT PROPERTY MECHANISM. Examples for this class of compound nouns are "heat exchange" and "wave propagation". Another class of compound nouns with good generalizations are subsets of the following features: Noun1: ELECTRIC-OBJECT PHYSICAL-OBJECT and Noun2: TIME PROPERTY, like in "transistor life".

Besides these classes of compound nouns with good generalizations, compound nouns with subsets of the following features do not generalize well: Noun1: PHYSICAL-OBJECT SPATIAL-LOCATION MATERIAL and Noun2: PHYSICAL-OBJECT GAS MATERIAL. Examples with subsets of these feature combinations are "earth satellite" and "water vapour". The reason for the decrease in the generalization performance for this last class is the restricted use of only 16 semantic features. To generalize relationships between two physical objects more features are needed. For instance, a network with a SIZE feature could generalize the WITH-P relationships between physical objects so that "earth satellite" could not be interpreted as "satellite WITH-P earth" since the earth has a bigger size than a satellite. The identification of these incorrectly generalized basic relationships is important for deciding which semantic features and basic semantic relationships might be modified. We make no claim for a "right" classification of semantic features and basic semantic relationships for our domain but we claim that the adaptive process of identifying better suitable semantic features and semantic relationships is supported by the learning ability and the modular architecture.

Related work

Comparing the performance of our system with existing systems for compound noun analysis is somewhat difficult, because the techniques, the level of the semantic relationships, and the domains are fundamentally different. McDonald reports about 54% to 64% correct interpretations for his compound noun system (McDonald 82) using detailed semantic relationships and fixed control strategies. The performance of Finin's system is similar to McDonald's system. Our system determines plausible basic semantic relationships for unknown compound nouns. Although our basic semantic relationships are not as detailed as McDonald's or Finin's, our basic semantic relationships are automatically acquired. As far as we know there is currently no system which has the ability to *learn* the semantic relationships between compound nouns.

Our system has the advantage of learning knowledge for the semantic relationships, while this knowledge is difficult to acquire in other compound noun systems (e.g. (Finin 80) (McDonald 82) (Arens 87) (Gay 88)). The knowledge about semantic relationships is represented uniformly in modular networks. On the other hand, these systems allow compound nouns with more than two words while we need additional mechanisms to understand longer compound nouns. Currently, we are investigating the use of recursive autoassociative network architectures ((Pollack 88), (St John 88)) and relaxation networks (Wermter 89) to understand compound nouns of arbitrary length.

Conclusions

One way to approach compound noun analysis is the use of extensive knowledge engineering, as demonstrated in several computational models. Because of the difficulties of identifying the semantic relationships and control structures, we presented a new approach for understanding compound nouns. Using a modular connectionist architecture we showed that basic semantic relationships within compound nouns can be learned. The general concepts of basic semantic relationships, learning, and modular network architectures demonstrate how uniform memory models can be built for natural language understanding.

Acknowledgements

I would like to thank Wendy Lehnert, Steve Bradtke, Claire Cardie, and Ellen Riloff for discussions and comments on earlier drafts of this paper. This research was supported by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract #N00014-87-K-0238, the Office of Naval Research, under a University Research Initiative Grant, Contract #N00014-86-K-0764 and NSF Presidential Young Investigators Award NSFIST-8351863.

References

1. Arens Y., Granacki J.J., Parker, A.C. 1987. Phrasal Analysis of Long Noun Sequences. *Proceedings of the ACL 87.*

2. Fahlman S.E. 1979. *NETL: A system for representing and using real-world knowledge.*

3. Finin T.W. 1980. *The semantic Interpretation of Compound Nominals.* PhD Thesis. University of Illinois at Urbana-Champaign.

4. Dahl A.D., Palmer M.S., Passonneau R.J. 1987. Nominalizations in PUNDIT. *Procedings of the ACL 1987.*

5. Gay L.S. 1988. *Interpreting Nominal Compounds for Information Retrieval.* Internal Report, GE Research and Development Center.

6. Lee R.B. 1960. The grammar of English Nominalizations. *International Journal of American Linguistics, Vol. 26, 3.*

7. Lehnert W.G. 1986. The analysis of compound nouns. *VERSUS, December, 1986.*

8. Levi J.N. 1978. The syntax and semantics of complex nominals.

9. Marcus M.P. 1980. *A theory of syntactic recognition for natural language.*

10. McDonald D.B. 1982. *Understanding Noun Compounds.* PhD thesis, Carnegie Mellon University.

11. NASA Thesaurus 1985.

12. Pollack J. 1988. Recursive Auto-associative Memory: Devising Compositional Distributed Representations. *Proc. of the Tenth Annual Conference of the Cognitive Science Society.*

13. Rumelhart D.E., Hinton G.E., Williams R.J. 1886. Learning Internal Representations by Error Propagation. In: McClelland J.L., Rumelhart D.E. (Ed.) *Parallel Distributed Processing Vol 1.* MIT Press.

14. Sparck-Jones K., VanRijsbergen C.J. 1976. Information Retrieval Test Collections. *Journal of Documentation, vol 32, no 1.*

15. St.John M.F., McClelland J.L. 1988. Applying contextual constraints in sentence comprehension. *Proc. of the Tenth Annual Conference of the Cognitive Science Society.*

16. Wermter S., Lehnert W.G. 1989. Noun phrase analysis with connectionist networks. In: Sharkey N., Reilly R.: *Connectionist Approaches to Language Processing.* (in press).

17. Wermter S. 1989. Integration of semantic and syntactic constraints for structural noun phrase disambiguation. *Proc. of the Eleventh International Joint Conference on Artificial Intelligence 89.*