Title
Pattern-Based Parsing for Word-Sense Disambiguation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 10(0)

Authors
Virkar, Rajesh S.
Roach, John W.

Publication Date
1988

Peer reviewed

Pattern-Based Parsing for Word-Sense Disambiguation

Rajesh S. Virkar & John W. Roach
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

Abstract

In the study of natural language understanding, the reductionist approach has been commonly used by A. I. researchers. Here we develop a technique for parsing based on this approach. We use a set of semantic primitives to represent word meanings and utilize patterns of sentences for mapping sentences onto meaning structures. To assist the parsing process, we develop semantic mappings for primitive sentences, semantic transformations for decomposing complex sentences using function words and axioms that encode world knowledge. We then explore the application of our approach to the word polysemy problem.

1. Introduction

Automated word-sense disambiguation is an unsolved problem in computational linguistics, and it forms the goal of the work presented here. We develop a pattern-based parsing technique and give illustrations for word-sense disambiguations using only the sentential context. Our solution to this problem utilizes the semantic primitives approach and appears to be a significant step in determining the ultimate consequences of this approach.

We hypothesize that most of the every-day lexicon can be represented by a large, yet finite, number of semantic primitives. Intuitively, we foresee two kinds of connections between these primitives: a 'tangled hierarchy' of primitives and 'semantic connections'. The tangled hierarchy is used to establish properties of an object inherited all the way down from the root-level nodes, while semantic connections are those relationships that capture world knowledge and provide us with clues in understanding linguistic expressions. Sentence forms can now be classified by the patterning of semantic primitives and function words. Function words are certain words (namely, prepositions, conjunctives, etc.) that may appear in a sentence. We hypothesize that all simple sentence forms can be captured by a very large, yet finite, set of patterns.

We employ a multi-tiered approach to parsing linguistic expressions. This approach accounts for the three different representations of expressions, namely 1. the surface representation of sentences, 2. the verbal representation of semantic primitive patterns, and 3. the deep meaning representation. Processing, in this approach, involves transforming the surface structure into its deep meaning representation.

2. Semantic Primitives and Their Relationship to Meaning

The set of semantic primitives we use is quite large (over 120 semantic primitives). These primitives can be broadly grouped under four classes: ENTITY, EVENT, ABSTRACT and RELATIONAL, and this categorization forms the foundation of the deep meaning representation, as will be seen later in the 'sentence meaning structures'. This classification scheme is adapted from Nida's work on componential analysis (Nida, 1975).

The primitives approach to developing a lexicon has been used by several computational linguists, including Laffal (Laffal, 1973). Laffal's work resulted in a set of 118 semantic categories that were combined to express around 23,500 words of the English language. Each word in this 'concept dictionary' was expressed by one category or by concatenating at most two categories.

Definition 1: A semantic primitive is a set that represents a class of words that refer to the same concept.

Example semantic primitives of the class ENTITY are PERSON (e.g. John) and WRITING-ARTIFACT (e.g. pen); examples of EVENT are DIRECTIONAL-MOTION (e.g. come) and DISTRIBUTE (e.g. give); examples of ABSTRACT are HEALTH-N (e.g. polio) and INTELLECT (e.g. smart); while examples of RELATIONAL are SPATIAL (e.g. inside) and TEMPORAL (e.g. before).

Definition 2: Interpretation 'I' refers to mapping a lexeme onto one of the possible senses of that lexeme. Here, we are acknowledging that a word or a sentence can have several interpretations. Senses are represented by semantic primitives and hence, we can say that $I : w \rightarrow P$. The set of interpretations $\mathbf{I}_W$ of a word $\mathbf{w}$ is equal to the set of semantic primitives $\mathbf{P}_W$ that can represent the given word in different situations.

Definition 3: The meaning of a word is the intended sense of the word in a given sentence. Thus, the meaning of a word $\mathbf{w}$ in a given sentence refers to one interpretation i_X represented by the semantic primitive $\mathbf{P}_X$ in the set $\mathbf{P}_W$ (i.e. $\mathbf{P}_X \in \mathbf{P}_W$). Similarly, the meaning of a sentence is the intended sense of the sentence.

Primitive Cohesion refers to that aspect of language which enables us to create newer (and/or larger) meaningful expressions by combining words. For example, by knowing the meanings of words such as 'John', 'drinks' and 'water', we can combine them to create a meaningful expression such as 'John drinks water'. Meaning has the property of preserving cohesion, i.e. meaning is that sense of a word that conforms to the semantic and linguistic constraints and contributes to the meaningfulness of the sentence.

Definition 4: A meaningful expression is an expression that describes an event with the help of other basic classes such as entities, abstracts and relationals. It is implied that to describe an event, one or more of these basic classes may not be required. A meaningful expression is neither anomalous, indeterminable, nor contradictory (Allan, 1986).

Suppose that we denote the set of all semantic primitives by Π. Then, Π^+ denotes the set of all strings, of length ≥ 1, of semantic primitives. The language L, the set of all meaningful expressions, can now be viewed as a subset of the set of all possible strings derived from Π. In other words, $L \subset \Pi^+$. Natural language understanding can now be roughly defined as a mapping, U, that translates all meaningful expressions in L onto the set of "sentence meaning structures" (SMSs), S; i.e., $U: L \rightarrow S$.

Definition 5: A function word is a word that, in its position in the sentence, signals the beginning of a new (possibly primitive) meaningful expression. For example, 'Jack and Jill went up the hill'. The set of function words **F** contains logical connectives such as 'and' and 'or', and prepositions such as 'for', 'of', etc.

Definition 6: A primitive sentence is a sentence that conveys only one "meaningful expression" (definition 4) in a language. A complex sentence, on the other hand, is a sentence that expresses two or more (related) meaningful expressions connected by one or more function words.

A simple complex sentence is one that has two meaningful expressions connected by one function word. Hence, by recognizing the function word, the two expressions can be separated into two primitive sentences.

Development of some sort of a mapping scheme is essential to achieve the deep meaning representation from the verbal representation. This mapping scheme should accept an incoming sentence and its pattern, and return an SMS (sentence meaning structure) for that sentence.

Definition 7: A semantic mapping is a function that translates a sentence, based on its corresponding semantic pattern, into one of the possible interpretations. An example of a mapping can be given as follows:

σ_{s1} : PERSON DIRECTIONAL-MOTION DIRECTION-RELATIONAL ARTICLE LOCATION

$\rightarrow$	[SMS	**(ENTITY**	PERSON)	
		(EVENT	*DIRECTIONAL-MOTION*)	
		(RELATIONAL	LOCATION)	]

σ_{s1} can be applied to a fairly large set of primitive sentences but we contend that the same meaning is expressed by all sentences belonging to this set. Here, the three slots (in bold) correspond to 3 of the ~120 primitives. It can be shown that any meaningful primitive natural language sentence can be reduced to its meaning in the form of an SMS.

Definition 8: An axiom is a rule based on the pattern of semantic primitives representing a sentence that allows the correct translation of the sentence into its corresponding meaning structure. An axiom helps the parsing process in eliminating the incorrect interpretations of a sentence. Execution of a semantic mapping must be guided by axioms to realize the correct meaning representation.

Let $S= w_1 w_2 .....w_r$ be a sentence with r words and $P= p_1 p_2 ......p_r$ be the pattern of semantic primitives based on the sentence S. An axiom a_i, as it governs the execution of a semantic mapping m_j, can be given in terms of S and P as follows:

$$a_i[m_j(w_1 w_2 ..... w_r, p_1 p_2 ...... p_r)] \rightarrow \alpha(w_k, p_k, << w_1 w_2 ..... w_r >>), \forall k \in \{1,2, ... ,r\}$$

Here, α is an admissibility procedure that checks the validity of the selected interpretation of a word in the context of the presented sentence. The enclosing symbols (<<,>>) are used to represent the sentence as a single entity, i.e. a meaningful expression consisting of a string of words.

$$\alpha(w_i, p_i, << w_1 w_2 ..... w_r >>) \rightarrow << w_1 w_2 ..... w_r >>_{wi,pi} \Vert p_i \subset \chi, \text{ where } \chi \in \Pi$$

Here, '$\Vert$' is a formal relation named 'is supported by', and the formula is read as follows: for the i^{th} word (w_i) in the sentence $<< w_1 w_2 ..... w_r >>$, the instantiation of the chosen semantic primitive (p_i) that represents its meaning is supported by the set inclusion property that relates p_i to one of the four basic classes of semantic primitives. The sentential context is used for guiding α in a given axiom.

The set of all axioms **A** can be developed by adding a new axiom (or rule) for every new interpretation of a word, produced in different patterns of primitives containing at least one primitive that represents that word. In other words, each axiom in **A** represents "primitive cohesion" or the links between primitive concepts.

We can claim that if $\mathbf{i_x}$ is a newly encountered (from the standpoint of **A**) interpretation of a word $\mathbf{w_k}$ in a sentence **S** (represented by a pattern **P**), and that an axiom governing such a semantic mapping does not exist, then the set **A** can be augmented by adding the appropriate axiom that oversees this semantic mapping.

<u>**Definition 9:**</u> A <u>semantic transformation</u> can be defined as the process of decomposing a complex sentence into two primitive sentences using the function word appearing in the complex sentence, without altering the 'meaning' expressed by the original complex sentence.

Let **T** be the set of semantic transformations and t_i be some semantic transformation. Let w_j ($j \leq m$, m is the length of the sentence S) be the function word. This means that S is a simple complex sentence that is a combination of two meaningful expressions. Then we can say that

$$\exists\, t_i \in T \text{ such that } t_i(S,P,w_j) \Rightarrow < (S_1,P_1)\ (S_2,P_2) >$$

where $< (S_1,P_1)\ (S_2,P_2) >$ preserves the meaning expressed by (S,P).

It should be noted, however, that several semantic transformations may exist for a given function word, and the one that is applied is selected based on the semantic pattern of the sentence. It should also be noted that a semantic transformation on a sentence that does not contain any function words is equivalent to applying a semantic mapping.

It is possible that after applying a semantic transformation, one or both of the resulting sentences will be complex. This can occur only in the presence of more function words. These function words can now be used to apply other semantic transformations and further reduce the sentences. Since every execution of a semantic transformation reduces the complexity of the sentence, a finite number of semantic transformations guarantees convergence to primitive sentences.

3. Word-Sense Disambiguation

Word-sense disambiguation has been a very important aspect of natural language understanding. From the definition of our semantic primitives, we know that different meanings of any given word refer to different corresponding semantic primitives as interpretations. Thus, polysemy is now reduced to the problem of removing extraneous/illegal interpretations from the set of all possible interpretations until we have reached the appropriate meaning.

A typical case of polysemy occurs in a class of sentences where some word may be used in several different places in a sentence. Here, different interpretations of the word suit different positions in the sentence and, in those positions, help represent a coherent meaning structure. This idea of **meaning distribution of multiple senses** will be explained through the following example.

(1) The pen is in the box.
(2) The box is in the pen.
(3) The ink is in the pen.
(4) The bull is in the pen.
(5) The pen is in the pen.

The analysis of these sentences gives a common meaning structure based on relationships of the constituent words and their interpretations. Since the ambiguous interpretations have not been resolved yet, the structure (not fully instantiated) can be given as:

<u>Common Meaning Representation</u>:

```
[SMS    (STATE      is)
        (?a1        ?a2)
        (RELATION   in)
        (?b1        ?b2)         ]
```

where "?" suggests an uninstantiated variable.

It is obvious that 'pen' is the ambiguous word in the group of sentences (1) through (5). The corresponding common interpretations are a. a pen as a writing instrument, and b. a pen as a small construction used for confinement. Axioms can now be applied to resolve the ambiguity in the meaning of 'pen'. The incorporation of semantic constraints and common world knowledge into these axioms should then be evident. Note that the following analyses are constructed using only the sentential context.

Analysis of (1):

a(S_{p1}.S_{p2}... S_{p6}, [The pen is in the box])

resolves 'pen'

using the function $\alpha(S_{p2},$[pen])

α in turn validates a possible instantiation-

Check_Instantiation(WRITING-ARTIFACT, [pen])

The reasoning can be presented as follows: a confining construction cannot be contained in a box, while a writing instrument can be contained in a box. This results into the resolved meaning structure:

```
                 [SMS    (STATE              is)
------->                 (WRITING-ARTIFACT   pen)
                         (RELATION           in)
                         (CONTAINER          box)        ]
```

Similar analyses can be performed on sentences (2) through (4). The last sentence, (5), is a very interesting example (due to Bar-Hillel). It involves two occurrences of the word 'pen' in which the object referenced by the meaning of one occurrence is contained inside the object referenced by the meaning of the second occurrence. Now, let us look at its analysis.

Analysis of (5):

a($S_{p1}.S_{p2}... S_{p6}$, [The pen is in the pen])

resolves the occurrences of 'pen'

using the functions $\alpha(S_{p2},$[pen]) and $\alpha(S_{p6},$[pen])

α in turn validates two possible instantiations-

Check_Instantiation(WRITING-ARTIFACT, [pen]) ----> for S_{p2}

and **Check_Instantiation**(CONSTRUCTION, [pen]) ----> for S_{p6}

The reasoning here involves the world knowledge about the sizes of the two 'pens'. An instance of a construction cannot be contained inside a writing instrument, while a writing instrument can certainly be kept in a storage construction. We are assuming that any object cannot be stored within itself, and in the event that a smaller 'pen' is kept inside a larger 'pen' (of the same interpretation), the context will dictate the choice. This reasoning results into the resolved meaning structure:

```
                 [SMS    (STATE              is)
------>                  (WRITING-ARTIFACT   pen)
                         (RELATION           in)
------>                  (LOCATION           pen)        ]
```

From the example sentences it should be clear how axioms incorporating semantic constraints can be used to disambiguate multiple senses involving meaning distribution in the same sentence.

4. Results

A natural language parser developed using our approach of semantic primitives and their patterns is operational on a VAX 11/785 with a 1 KLIPS Prolog interpreter. The examples presented in this paper are actual results among many produced by the program.

The lexicon used by the program, in its current state, has over 500 different word senses. The words are divided into four broad classes: (a) Entities, (b) Events, (c) Abstracts, and (d) Relationals. There are over twenty different semantic transformations in the program based on function words and phrases. The transformations support over forty semantic mappings. The program also contains over forty axioms that aid in reducing sentences into their respective sentence meaning structures.

The techniques described in this paper have been used to develop a natural language knowledge assimilation system for expert systems (Virkar & Roach, 1988). This system augments an already existing knowledge base of a drug interaction expert system, DIE (Roach, et. al, 1985), by reading and understanding not only sentences but entire research paper abstracts in the pharmacology domain thus demonstrating the extension of these techniques to the text-level.

5. Conclusion

In this paper we have developed a pattern-based technique for natural language parsing based on semantic primitives and their patterns. We have developed axioms from the semantic primitives for encoding world knowledge.

We have given examples of word-sense disambiguation in sentences using only the sentential context under the class of meaning distribution. We have shown, in detail, how the technique is used to arrive at the correct meaning representations.

Future efforts will focus on encoding entailment rules that represent deep inferencing techniques and semantic context structures that dynamically evaluate the parsing process.

Bibliography

Allan, K. 1986. Linguistic Meaning: Vol. 1. New York: Routledge, Kegan Paul.

Goguen, J. A., Thatcher, J. W. and Wagner, E. G. 1978. *An Initial Algebra Approach to the Specification, Correctness and Implementation of Abstract Data Type.* In Yeh, R. T. (ed.) Current Trends in Programming Methodology, Vol. 4. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Guttag, J. V. 1975. *The Specification and Application to Programming of Abstract Data Types,* Ph. D. Thesis, University of Toronto, Department of Computer Science, available as report CSRG-59.

Laffal, J. 1973. A Concept Dictionary of English. New York: Gallery Press.

Nida, E. A. 1975. Componential Analysis of Meaning: An Introduction to Semantic Structures. The Hague: Moulton.

Roach, J., Lee, S., Wilcke, J. and Ehrich, M. 1985. *An expert system for information on pharmacology and drug interactions.* Computers in Biology and Medicine. Vol. 15, No. 1.

Virkar, R. and Roach, J. 1988. *Direct assimilation of expert-level knowledge by automatically parsing research paper abstractss*International Journal of Expert Systems (in press).