

A Computational Method for Resolving Ambiguities in Coordinate Structures

Haodong Wu and Teiji Furugori
Department of Computer Science
University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182, JAPAN
{wu, furugori}@phaeton.cs.uec.ac.jp

Abstract

This paper describes a method for determining syntactic structure in coordinate constructions. It is based on the information taken from semantic similarities, selectional restrictions, and some other linguistic cues. We discuss the role the information plays in resolving ambiguities that appear in coordinate constructions, describe the means of acquiring the necessary information automatically from two on-line corpora and a lexical database, and devise two algorithms for disambiguating coordinate constructions. An experiment that follows shows effectiveness of our method and its applicability to resolving ambiguities in some other syntactic structures.

1 Introduction

Syntactic ambiguity appears, among others, in coordinate constructions. It is an annoying problem in analyzing structure and meaning of a sentence. A parser, for instance, is to detect the scope of a coordinate structure and identify its inner modification relations. However, the current parsers (e.g., the Link parser) often fail to handle the problem and/or produce a large number of parses.

There are a few computational studies that have tried to resolve ambiguities in coordinate constructions (e.g., Paritong, 1992; Cooper, 1991; Bayer, 1996). For example, Kurohashi and Nagao (1994), in analyzing long Japanese sentences, proposed a syntactic analysis method for detecting conjunctive structures by using lexical similarity and structural parallelism. Mela and Fouqueré (1996) used a direct process to determine the scope of a coordinate structure based on the concept of functor, argument and subcategorization. Unfortunately, neither of them has sufficiently dealt with the syntactic structure of a coordination especially when a coordinator (such as *and*, *or* and comma) has two or possibly more preceding and succeeding constituents.

We in this paper propose a method for determining the structure of a coordinate construction using information on similarities, selectional restrictions, and other linguistic cues. In Section 2 we identify the problem and describe the ideas behind our method in Section 3. We give disambiguation algorithms, show an disambiguation experiment, and evaluate its results in Section 4. In Section 5 we suggest an applicability of our method to resolving other syntactic ambiguities.

2 Modification Relation in Coordination

Resolving ambiguities in a coordinate construction is to determine the way of conjoining constituents (words, phrases, or clauses) and/or to determine the scope of coordination, i.e., immediacy relations among the constituents involved. For instance, in *sweet and sour pork*, the right immediacy relation is *((sweet and sour) pork)* rather than *(sweet and (sour pork))*.

Consider other examples.

- (1) Tom is a ((stock and estate) keeper).
- (2) John is a (student and (chess player)).
- (3) Old men and women were left at the village.

In each of these sentences, a noun or an adjective that appears in the left hand side of coordinator(s) has two (or more) modificant candidates: it may or may not modify the head noun in the right hand side.

In (1), *stock* and *estate* should be conjoined to modify *keeper*¹. In (2), there is not a modification relation between *student* and *player* and its interpretation should be (student and (chess player)).

We are able to make unique interpretation for (1) and (2), but we may have two interpretations for (3).

- (3a) *Women* are left at the village and *old men* were left at the village.
 (3b) *Old men* were left at the village and *old women* were left at the village.

Later in the paper, we try to resolve the ambiguities by determining the modification relation in the coordinate constructions (CCs) such as in (1) to (3).

3 Identifying Modification Relation in Coordination

We have found through linguistic observations that a variety of information supplies important cues for disambiguation. Some of them are computable and effectively used in a computer model of disambiguation. The ones we thought most important include: similarities in syntactic forms and/or meanings, selectional restrictions, and orthographic forms.

3.1 Linguistic Observations

Similarities in Syntactic Forms and Meanings We see that similarities on forms and meanings are crucial to determine the structure of a coordinate construction.

If the modifier is not an adjective, for instance, it is likely that two constituents before and after the coordinator are conjoined when they belong to the same subcategory and match in number:

- (4) ((business and management) sections)
 (5) (businesses and (culture activities))

In the following examples,

- (6) ((research and development) section)
 (7) (researcher and (system engineer))

it is obvious that the *research* and *development* have more in common in meaning than *research* and *section*, and that *researcher* and *engineer* are semantically more similar than *researcher* and *system*. Likewise, we see:

- (8) (lovely (cats and dogs))
 *((lovely cats) and dogs)

Selectional Restrictions Consider the sentence:

- (9) Peter likes ((green vegetables) and (music)).

we know in (9) that *green* as a color can be used to modify concrete entities like *vegetables*, but not abstract one like *music*. This means that selectional restriction (SR), a semantic restriction imposed on lexical items when forming a sentence, is an important factor to determining the structure of coordinate constructions. In this paper, we discuss SR in the context of adj+n1+and+n2 and its extension (e.g., adj n_1, \dots, n_k)².

Other Linguistic Cues Orthographic forms often play an important role in disambiguating the structures of coordinate constructions. It is likely that all nouns can be conjoined when they are in capital forms. An example:

- (10) ((Research and Development) Section)

When the conjoined nouns in a coordinate structure are preceded by a determiner, the usual

¹In this case, the coordination of *stock and estate keeper* is considered to be the reduced form of *stock keeper and estate keeper*.

²Hereafter, *and* in adj+n1+and+n2 and in n1+and+n2+n3 represents a coordinator such as *and*, *or*, *comma*, and the like.

interpretation is that the determiner applies to the conjoins:

- (11) *Old men and women* were left to organize the community.

The structure (old (men and women)) is more likely than ((old men) and women) because there is a tendency that the determiner is not repeated in the noninitial conjoins, e.g., *the (man and women)* (Quirk et al. 1985).

3.2 Computational Measurements of Similarities

Semantic similarity has been measured in a number of ways (e.g., Resnik 1993; Kozima and Furugori, 1993; Dagan et al., 1994). Many researchers opt to investigate methods for deriving measures of semantic similarity among words based on distributional behavior observed in corpora or machine-readable dictionaries. We however employ a hybrid method for combining the quantitative method with an existing broad-coverage source of lexical knowledge.

To capture the similarity between two words in context, we consider two kinds of semantic relations useful and effective: *taxonomic relation* and *co-occurrence relation*.

Computationally, one way of measuring semantic similarity is to use the taxonomic relations in the WordNet (Miller, 1990), a widely used lexical database including four kinds of semantic relationships in a word sense network: *synonymy*, *hyponymy*, *meronymy*, and *antonymy*. We found that the *antonymy* relationship appears often in CCs (e.g., *brother and sister*, *man and woman*, and *boy and girl*) and shows a strong tie between two nouns in the coordinate structure like (14).

- (12) (cheerful (boys and girls))

Another way of measuring semantic similarity is to use the mutual information (MI) (Church and Hanks, 1990). It may be taken from the co-occurrence relations using two corpora: the EDR English Corpus and Brown Corpus³. We define the similarity between two words w_1 and w_2 when appearing with an adjacent word w . In an example,

- (13) *new books and hamburgers*

where w_1 , w_2 , and w are *book*, *hamburger* and *new*, respectively. We can measure two similarities, the *left side similarity* and *right side similarity*. w comes to the left of w_1 and w_2 in the former like *young* in (14) and it comes to the right side of w_1 and w_2 in the latter like *investigation* in (15).

- (14) *young* nations and superpowers
 (15) finance and market *investigation*

The *left side similarity* and the *right side similarity* are defined as:

$$sim_L(w_1, w_2, w) = \frac{\min(I(w, w_1), I(w, w_2))}{\max(I(w, w_1), I(w, w_2))} \quad (1)$$

$$sim_R(w_1, w_2, w) = \frac{\min(I(w_1, w), I(w_2, w))}{\max(I(w_1, w), I(w_2, w))} \quad (2)$$

Here, for two words x and y , $I(x, y)$ is MI between x and y , where x is on the left side of y , and $I(y, x)$ is MI between y and x , where x is on the right side of y .

If w appears in either side of w_1 , or in either side of w_2 , we can define *two-sided similarity* as:

$$sim(w_1, w_2, w) = \frac{sim_L(w_1, w_2, w) + sim_R(w_1, w_2, w)}{2} \quad (3)$$

Here, w may be a verb, a noun, an adjective, or a phrase like *agree with*.

³EDR English Corpus, compiled by Japan Electronic Dictionary Research Institute, Ltd., contains 160,000 sentences with annotated morphological, syntactic and semantic information. The Brown Corpus was compiled in the early 1960s at Brown University, USA, under the direction of W. Nelson Francis and Henry Kucera. It contains 500 text samples representing 15 categories of American English texts printed in 1961.

Often the word pairs from w , w_1 and w_2 are unobserved in the corpus being used. This problem is known as the sparse data problem. There are a number of methods for dealing with this problem (e.g., Resnik, 1993; Dagan et al., 1995). Our solution in this paper is to use a synonym set (S) in the WordNet to substituting for the sparse word w . The above formulae in this case becomes:

$$sim_L(w_1, w_2, w) = \frac{\sum_{c \in S} \min(I(c, w_1), I(c, w_2))}{\sum_{c \in S} \max(I(c, w_1), I(c, w_2))} \quad (4)$$

$$sim_R(w_1, w_2, w) = \frac{\sum_{c \in S} \min(I(w_1, c), I(w_2, c))}{\sum_{c \in S} \max(I(w_1, c), I(w_2, c))} \quad (5)$$

here c stands for a synonym set related to the word w .

w_1 and w_2 may be replaced by their synonym sets S1 and S2 in the WordNet. In this case, the similarity between them is estimated by:

$$sim_L(w_1, w_2, w) = \frac{\max_{c_1, c_2} \sum_{c \in S} \min(I(c, c_1), I(c, c_2))}{\sum_{c \in S} \max(I(c, c_1), I(c, c_2))} \quad (6)$$

$$sim_R(w_1, w_2, w) = \frac{\max_{c_1, c_2} \sum_{c \in S} \min(I(c_1, c), I(c_2, c))}{\sum_{c \in S} \max(I(c_1, c), I(c_2, c))} \quad (7)$$

The values produced by (6) and (7) have an intuitive interpretation. Each of them denotes a maximum similarity between synonyms of w_1 and synonyms of w_2 with synonyms of w .

3.3 Selectional Restrictions as Constraints

Selectional restrictions (SRs) are often defined on an empirical basis using AI (artificial intelligence) techniques. Recently several research efforts have turned to corpus-based methods to define and acquire them. Along the line, we search a new approach for finding SRs using on-line corpora and a lexical database.

We view SR as negative information, a constraint between two words. The information on SRs may be computed from corpora: for a particular adjective and a particular noun, we try to find 'similar' words and then check if they co-occur in the corpora. If no co-occurrences are observed in the corpora, we assume that there is a SR between the adjective and the noun.

A problem we encounter here is how to find 'similar' words. It has been proven by an experiment that similar words generated by corpus-based clustering methods does not work well (Grefenstette 1993). We choose to acquire similar words from a taxonomy (the WordNet).

It seems to be reasonable for a noun to use its direct hypernym (father node in IS-A hierarchy in the WordNet) and hyponyms of the hypernym (the siblings of the noun) as similar words. But it does not work for an adjective. A compromising measure for an adjective is to use synonyms rather than hypernyms as similar words, e.g., {pure, unmixed, undiluted} for *pure*.

Thus, if n_2 or its similar words co-occur adj or its synonyms (in the WordNet) in the corpora (the EDR English corpus and the Brown corpus), we may conclude that the CC has the structure (adj (n_1 and n_2)). Otherwise, it would be ((adj n_1) and n_2) as there is a SR between adj and n_2 . Take (16) for an example,

(16) ((Fresh air) and sunshine) bring me health and feelings of joy.

We are sure that *fresh air and sunshine* has the structure ((*fresh air*) and *sunshine*) as no co-occurrence between the synonyms of *fresh* and the similar words of *sunshine* is found.

4 Disambiguation for Structure of Coordinate Constructions

We have devised disambiguation algorithms based on what we have described in section 3. Algorithm 1 below is for CCs in the form adj+n1+and+n2 and algorithm 2 is for CCs in the form n1+and+n2+n3. Algorithm 3 describes the process of decomposing a normal CC into CCs in the forms adj+n1+and+n2 and n1+and+n2+n3, and determining its inner syntactic structure.

Algorithm 1: Disambiguation of CCs in the form adj+n1+and+n2

1. Check adj+n2 in the corpora. If it is observed, produce (adj (n1 and n2)).
2. Otherwise, if there is a synonym of adj and a synonym of n2 and if they co-occur in the corpora, produce (adj (n1 and n2)).
3. Otherwise, create a class which includes n2, its parent (hypernym), and its siblings (the hyponyms of the hypernym) in IS-A hierarchies of the WordNet; If no member in this class co-occurs with adj and any of its synonyms, produce ((adj n1) and n2).
4. Otherwise, produce (adj (n1 and n2)) as a statistics-based default.

Algorithm 2: Disambiguation of CCs in the form n1+and+n2+n3

1. If all of the three nouns are capitalized, produce ((n1 and n2) n3).
2. Otherwise,
 - (a) if n1 and n2 match in number and n1 and n3 do not, produce ((n1 and n2) n3).
 - (b) if n1 and n3 match in number and n1 and n2 do not, produce (n1 and (n2 n3)).
3. Otherwise,
 - (a) if n1 is the antonym of n2, produce ((n1 and n2) n3);
 - (b) if n1 is the antonym of n3, produce (n1 and (n2 n3)).
4. Otherwise,
 - (a) if $\text{sim}_R(S1,S2,S3) > \text{sim}_L(S2,S3,S1)$ ⁴, produce ((n1 and n2) n3).
 - (b) if $\text{sim}_R(S1,S2,S3) < \text{sim}_L(S2,S3,S1)$, produce (n1 and (n2 n3)).
5. Otherwise, produce ((n1 and n2) n3).

Experimental Results To evaluate the performance of the disambiguation algorithm, we randomly selected two sets of 300 coordinate structures of the form adj+n1+and+n2 and n1+and+n2+n3 from on-line CNN news using the method proposed by Mela and Fouqueré (1996) and ran the algorithms on a computer.

The result for disambiguating CCs of adj+n1+and+n2 is shown in Table 1 and that for n1+and+n2+n3 in Table 2.

Table 1: Experimental Results on Testing adj+n1+and+n2

Step	Recall (%)	Number	Accuracy
(1) directly observed pattern	38.0	114	100.0%
(2) indirectly observed pattern	40.3	75	86.7%
(3) selectional restriction	46.8	52	86.5%
(4) default	100.0	59	66.7%
Total	100.0	300	87.7%

Table 2: Experimental Results on Testing n1+and+n2+n3

Step	Recall (%)	Number	Accuracy
(1) orthographic form	3.7	11	100.0%
(2) antonym	8.3	24	95.2%
(3) similarity in form	38.8	104	91.4%
(4) semantic similarity	100	159	79.9%
Total	100.0	300	85.3%

The algorithms have adopted a back-off form to integrate different cues in the disambiguation process and the reliable cues with certainty are used first to achieve a better overall performance. As we can see from the results, the cues (orthographic forms, syntactic constraints, antonymy relation, observed patterns) with high success rates have comparatively low recall rates (from 3.7% to 40.3%); other cues, such as selectional restriction and semantic similarity, on the other hand, have comparatively high recall rates.

Evaluation Table 3 shows the results of the performance achieved by our method and those by others for the structure of n1+and+n2+n3. All these methods use the same data. Here, (1) shows the result obtained from attaching the modifier to the nearest head (Kimball, 1973), i.e., (n1

⁴S1 is a set which consists of n1 and its synonyms extracted from the WordNet. Similarly, S2 is a set including n2 and its synonyms, and S3 is a set including n3 and its synonyms.

and ($n_2 n_3$); (2) shows the result of the method proposed by Resnik (1993) in which class-based similarity and a measure of noun-noun modification estimated from corpora are used in resolving ambiguous coordinations; (3) shows the performance of our method; and (4) is the performance of human judgment by three native speakers who were just presented the words of n_1 , coordinator, n_2 and n_3 without surrounding contexts.

Table 3: Comparison with Other Work for Determining $n_1+and+n_2+n_3$

Method	Success Rate
(1) Closest attachment	65.3%
(2) Resnik's method	80.7%
(3) Our method	85.3%
(4) Average human	91.7%

The lower bound and the upper bound on the performance of our method seem to be 65.3% scored by the simple heuristics of closest attachment (1) and 91.3% by human beings (4). We can clearly see here that the performance of our method (3) is better than those of (1) and (2), and is close to that of human beings.

Table 4 shows the results achieved by our method and those by others for the structure of $adj+n_1+and+n_2$. Here, each row of (1), (3), and (4) is analogous to that in Table 3. (2) shows the result from estimating the strength of association between adj and n_2 using the maximum mutual information over their classes (Resnik, 1993; Alves, 1996).

Table 4: Comparison with Other Work for Determining $adj+n_1+and+n_2$

Method	Success Rate
(1) Closest attachment	64.7%
(2) Maximum MI	82.3%
(3) Our method	87.7%
(4) Average human	93.3%

The performance by the closest attachment (1) is so poor that it would be unusable in any real applications. The method (2) using maximum MI performed better. But again ours is much better.

This is not to say our method is perfect, however. Selectional restrictions do not work well in idiomatic or fixed expressions (e.g., *green peace*) or when the adjective has multiple senses. Take (17), for instance,

(17) I bought some *soft balls and drinks* in a drugstore.

The algorithm 1 produces the parse (*soft (balls and drinks)*), but the correct one should be (*(soft balls) and drinks*).

The judgment made by semantic similarity succeeded in about 80% of the cases (Table 2), but often failed when the co-occurrences are low and especially when the words involved are polysemous. We see that we need to use a larger corpus to overcome this problem.

5 Discussion

Applications to Complex Coordinations and Nomial Compounds The method presented in this paper can be directly used for resolving complicated cases of coordinate structures (e.g., *freshman training and personal management system*). The coordinate structure of $adj+n_{left_1}+\dots+n_{left_l}+and+n_{right_1}+\dots+n_{right_k}$ can be reduced to CCs of $adj+n_1+and+n_2$, $adj+n_2+and+n_3, \dots, adj+n_{k-1}+and+n_k$. So we can apply algorithm 1 to disambiguate these constructions and integrate them to acquire the overall structure of the CC.

The coordinate structure of $n_{left_1}+\dots+n_{left_l}+and+n_{right_1}+\dots+n_{right_k}$, on the other hand, can be quite complex. Theoretically, l or k may be quite a big number. We found that in the most of the cases (>99.3%), l is no greater than 2 and k is no greater than 3 in real texts, however. For a complex CC in form of $n_1+n_2+and+n_3+n_4+n_5$, for instance, we can use

semantic similarity defined in formulae (6) and (7) in Section 3 to find the word in the right hand side of the coordinator that is most similar to n_2 . Suppose n_2 is similar to n_4 , we then check whether n_1 co-occurs with n_3 in pattern $n_1 n_3$ in the corpora, if the answer is yes, then produce $(n_1 (n_2 \text{ and } (n_3 n_4)) n_5)$, otherwise produce $((n_1 n_2) \text{ and } (n_3 n_4)) n_5$. Using this method, the analysis for CC *freshman training and personal management system* is $((\text{freshmen training}) \text{ and } (\text{personal management})) \text{ system}$; while for *food handling and storage procedure* the result turns to be $((\text{food handling and storage}) \text{ procedure})$.

The method can also be applied to analyzing structures of nominal compounds. Take the phrase *novice song bird feeder kit*, for instance. Selectional restriction may play an important role in judging which constituent the adjective *novice* refers to in the candidates *song, bird, feeder, kit* and their combinations. Co-occurrence relation, on the other hand, is crucial to determining the structure of a nominal compound like *song bird feeder*. We may collect statistics to see if *song bird* is observed more often than *song feeder*.

Conclusion Resolving ambiguities in coordinate structure has a great importance for its applications to text understanding and machine translation. It is crucial to information retrieval, too (e.g., internet information retrieval). Given "training system" as a conjoined retrieval condition, for example, the phrase *freshman training and personal management system* in the target text can be retrieved using the method presented in this paper, but it can hardly be found with the retrieval techniques available so far.

The disambiguation method proposed is scalable since it does not depend on any handcrafted rules. It is strong at the data sparseness problem also as we use co-occurrences between semantic classes, rather than words, extracted from a lexical database.

The disambiguation experiment has proven that our method for disambiguating syntactic structures is valid, effective, and useful in practical applications. The performance of our method is significantly better than those of other work. We think that the performance can be improved further by using a larger corpus that contributes to the precision for estimating semantic similarities and/or selectional restrictions.

References

- Alves, E. (1996). "The Selection of the Most Probable Dependency Structure in Japanese Using Mutual Information." *Proceedings of the 34th Annual Meeting of ACL*, 372-374.
- Bayer, S. (1996). "The Coordination of Unlike Categories." *Language*, 72(3):579-616.
- Church, K. W. and Hanks, P. (1990). "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics*, 16(1):22-29.
- Cooper, R. P. (1991). "Coordination in Unification-based Grammars." *Proceedings of the 29th Annual Meeting of ACL*, 167-172.
- Dagan, I.; Marcus, S.; and Markovitch S. (1995). "Contextual Word Similarity and Estimation from Sparse Data." *Computer Speech and Language*, 9:123-152.
- Grefenstette, G. (1993). "Evaluation Techniques for Automatic Semantic Extraction: Comparing Syntactic and Window-based Approaches." Technical Report, Department of Computer Science, University of Pittsburgh.
- Hindle, D. and Rooth, M. (1993). "Structural Ambiguity and Lexical Relations." *Computational Linguistics*, 19(1):103-120.
- Kimball, J. (1973). "Seven Principles of Surface Structure Parsing in Natural Language." *Cognition*, 2:15-47.
- Kozima, H. and Furugori, T. (1994). "Segmenting Narrative Text into Coherent Scenes." *Literary and Linguistic Computing*, 9(1):13-19.
- Kurohashi, S. and Nagao, M. (1994). "A Syntactic Analysis Method of Long Japanese Sentences Based on the Detection of Conjunctive Structures." *Computational Linguistics*, 20(4):507-534.
- Mela, A. and Fouqueré, C. (1996). "Coordination as a Direct Process." *Proceedings of the 34th ACL Meeting*, 124-130.
- Miller, G. (1990). "WordNet: An On-Line Lexical Database." *International Journal of Lexicography*, 3(4) (special issue).

- Paritong, M. (1992). "Constituent Coordination in HPSG." *KONVENS 92*, 228-237, Springer Verlag.
- Resnik, P. S. (1993). "Selection and Information: A Class-Based Approach to Lexical Relationships." Doctoral Dissertation, University of Pennsylvania, Philadelphia, P.A.
- Sleater, D. and Temperly, D. (1991). "Parsing English with a Link Grammar." Carnegie Mellon University Computer Science technical report CMU-CS-91-196.
- Wu, H. and Furugori, T. (1996). "A Hybrid Disambiguation Model for Prepositional Phrase Attachment." *Literary and Linguistic Computing*, 11(4), 187-192.
- Wu, H. (1997). "A Hybrid Model for Resolving Syntactic Ambiguities in Natural Language Processing." Doctoral Dissertation, the University of Electro-Communications, Tokyo, Japan.