Automatic Construction of a Concept Hierarchy from Coordinated Noun Phrases

Yongkyoon No*

Chungnam National University

Yongkyoon No. 2007. Automatic Construction of a Concept Hierarchy from Coordinated Noun Phrases. *Language and Information 11.1*, 39–52. Noun phrase coordination is an extremely productive phenomenon. Based on an observation that conjuncts tend to denote semantically related concepts, we collect four hundred thousand pairs of conjuncts from the British National Corpus, in an attempt to build an is-a hierarchy of English noun concepts. The modifiedness patterns of the two words in these pairs point to three distinct semantic relations: sibling, cousin, and ancestor-or-ancestor's-sibling. The process of finding them and how these pairs are used to motivate groups of quasi-synonyms and then to locate the hypernyms are discussed. (**Chungnam National University**)

Key words: is-a hierarchy, concepts, English nouns, coordination, modifiedness

1. Introduction

A lexical semantic hierarchy has proved useful in natural language processing. When, as is too often the case, a database of worldly facts is practically unobtainable, a concept hierarchy, like WordNet (Fellbaum, 1998), can serve as the next-best solution to the problem of disambiguation and inference.

Actual construction of a hierarchy is not an easy task. Languages typically contain hundreds of thousands words. Even if compounds and proper nouns are excluded, the number of nouns in a typical language is well over thirty thousand.[1]

* Thanks are due to three anonymous referees of *Language and Information*, who went through earlier versions of this paper, pinpointed inadequacies to the author, suggested ways of achieving some of the subgoals as well as stylistic patches, and, in general, contributed to sharpening our thoughts. The author can be contacted at Department of Linguistics, Chungnam National University, 220 Koong dong, Yuseong gu, Taejeon 305-764. E-mail: yno@linguist.cnu.ac.kr

[1] WordNet 1.4 has 57,000 nouns (Miller, 1993), and *Collins English Dictionary* has 43,636 nouns (Fellbaum, 1993). The abridged *Yonsei Dictionary of Korean* has 30,774 nouns (indirectly from now defunct *http://clid.yonsei.ac.kr:8000/dic/faq.html*).

Considering the slippery semantic relations between a billion or so pairs of words[2] would take a team of qualified researchers ten years.

Thus, there is enough motivation to an automatic construction of a concept hierarchy. For further motivation and the current state of the art, see Cimiano, Völker, and Studer (2006). In this paper, we describe an algorithm for acquiring pairs of semantically close nouns from a very large corpus. How these pairs can be used to populate vertices of a concept hierarchy is discussed. We assess difficulties involved in acquiring the right pairs from large corpora. Firstly, we begin by putting forth our assumptions regarding the language system and the notion of concept hierarchy as a mathematical object.

2. Assumptions about coordination and concept hierarchies

We gather a large number of semantically related nouns through analyses of coordination constructions. This move is motivated by the following assumptions:

1. The meaning of a noun phrase is most crucially a function of the meaning of its head noun.

2. Conjuncts are semantically similar, or close, to each other.

3. The way in which the head noun of one conjunct is related to that of the other is dependent on the structure of both conjuncts..

We take the first to be noncontroversial, in line with, among many others, Zwicky (1985). What is not all that evident is the second assumption. Let us, at this moment, note that the property of being similar is a vague notion; it is hard to refute this assumption. Since Brown et al. (1992), semantic similarity between two words is usually computed from corpora by statistical means. Finch (1993), Resnik (1995), Lin (1998), and Caraballo (2000) all succeed in automatically collecting groups of nouns from corpora that are semantically similar. Lin (1998), for instance, gathers millions of $\langle$verb, rel-name, noun$\rangle$ triples and applies informationtheoretically defined similarity measures to find groups of nouns that are semantically similar to each given noun. To these researchers at least, similar words are similar simply by virtue of occurring in similar contexts.

If two nouns occur as objects of the same group of verbs with similar frequency, and if they occur as subjects of another same group of verbs, and if they occur as objects of a same group of prepositions, they can hardly be distinct in meaning. This is the intuition behind the statistically oriented research. However, the community of statistical research has not paid much attention to the very fact that the construction of coordination itself presents pairs of words that are semantically similar to each other. As coordination involves two categories of the same sort, and a coordinated phrase stands in a substitution and entailment relationship to its members, the two members share the same lexicogrammatical relations wher-

[2] $30000^2 \approx$ one billion

ever the coordinated phrase consisting of them occur throughout the corpus.[3] In other words, two members of a coordinated phrase will contribute the same triples in Lin's (1998) and others' calculations. Thus, our second assumption above is motivated.

Those readers who are not convinced may want to note that, while the first two sentences of (1) sound acceptable, the last two are odd.[4] This oddity is taken to prove that *kindness* is semantically unsimilar to *sandwich*, while *politeness* and *kindness*, on the one hand, and *biscuit* and *sandwich*, on the other, are similar to each other.

(1) a. There were plates of biscuits and sandwiches and little slices of sausage. (B0U)

b. I must take no notice of their politeness or kindness which was designed to trap me into giving information. (B0U)

c. There were plates of kindness and sandwiches and little slices of sausage.

d. I must take no notice of their sandwiches or kindness which was designed to trap me into giving information.

There are infinite number of ways a concept hierarchy can be built. In order to constrain this space of possibility somewhat, we assume:

1. A concept hierarchy is a directed acyclic graph whose vertices contain a distinct concept.

2. While some concepts are lexicalized, others are not.

3. Vertices which do not contain a lexicalized concept are allowed, but only as an ancestor of one which does.

4. The longest path of a concept hierarchy consists of a designated number of edges.

These assumptions are harmless, and their usefulness will be proved only by the usefulness of the resulting hierarchy. Unlike Princeton WordNet (Fellbaum, 1998), where some synonym groups are removed from the most general concept by as many as twelve edges, we aim at a graph whose longest edges are of length five or smaller. Li and Abe (1998), for instance, use the upper portions of WordNet's noun hierarchy in determinging acceptable classes of verb arguments and prepositional phrase attachment. Rosario, Hearst, and Fillmore (2002) follow suite, when they show that instances of spurious polysemy disappear from the thesaurus of medical terms, i.e. MeSH, as low level distinctions are abandoned.

[3] If *I saw a book and two pictures* is grammatical, then both *I saw a book* and *I saw two pictures* are grammatical. If the former is true, both of the latter two are true. The former would be false if either of the latter two is false.

[4] Example sentences are drawn from the British National Corpus and accompanied with the names of the files in which they occur.

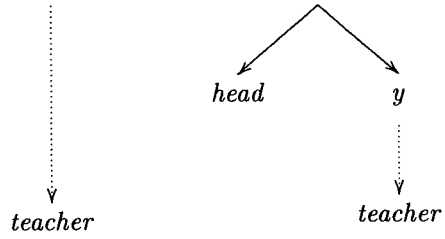
A crucial assumption in addition to the above ones, and one which seems to be entirely of our own, concerns a syntactic property of conjuncts relative to each other and its reflection on the head nouns' semantic relationship.

> Two nouns, say N_1 and N_2, stand in a semantic relation that depends systematically on whether they occur with modifiers in a conjunction.
>
> i. If N_1 is modified and N_2 is not, and they occur as heads of conjuncts in a conjunction, N_1 tends to denote a concept which either subsumes N_2's denotation or is a semantic sibling of an ancestor of the concept denoted by N_2.
> ii. If neither conjuncts are modified, they tend to denote conceptual siblings.
> iii. If both conjuncts are modified, they tend to denote concepts that are subsumed by a remote predecessor in the graph and are of equal level of specificity.

The first sub-assumption, when applied to a sentence like (2), yields partial graphs in (3). (Unlike a solid arrow, which represents a direct hypernymy relation, a dotted one represents an indirect hypernymy.)

(2) ... schools are going to be excellent if the teachers and heads of faculty in those schools are excellent, and we've got lots of those in Banbury. (KRK)

(3) a. *head* b. *x*



The partial graph in (3a) is simply wrong: a teacher is not a kind of head. At present, however, we have no way of automatically judging it to be an incorrect representation. We will say. for this reason, that *head* stands in a relationship called ancestor-or-ancestor's-sibling to *teacher*. More discussion follows in Section 4. The partial graph (3b), however, seems to be correct: a head is a more general concept than a teacher.[5]

The reason why the modifiedness pattern $\langle -w_1, +w_2 \rangle$ is associated with a hypernymy at all, is that there are many instances of such an association.[6]

[5] WordNet has person–adult–professional–educator–teacher and person–leader–head.

[6] A plus sign before a word indicates that it is modified in the smallest constituent that contains both the word itself and the conjunction. A minus sign indicates that it is not modified. Modifiers of the noun includes adjectives, nouns, prepositional phrases, and relative clauses. Specifiers are not taken to be modifying a noun.

(4) There's a fitness room, sauna and a solarium. (ED1)

(5) ... which conveys meaning to a group of pupils, the teacher or another known adult. (ANS)

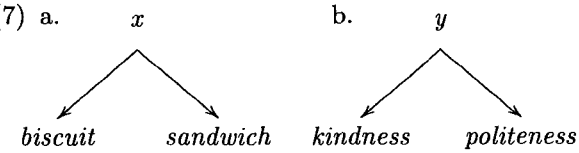
(6) ... these also involved the resentment of building workers and weavers at being displaced from employment by the cheaper Irish. (HXC)

Sentences (4), (5), and (6) contain coordinated noun phrases with the contrasting modifiedness specifications on the head nouns. A solarium is a kind of room; a teacher is a kind of adult; a weaver is a kind of worker.

It is worth noting that the particular "lexico-syntactic" structure that is known, since Hearst (1992), to be a reliable indicator of a hypernym – hyponym relation, namely that of "N and other N[plural]" is taken as a structure which yields a pair related this way if, and only if, the first noun is not modified. This pattern of Hearst's is a special instance of $\langle -w_1, +w_2 \rangle$ and naturally supports the first sub-assumption.

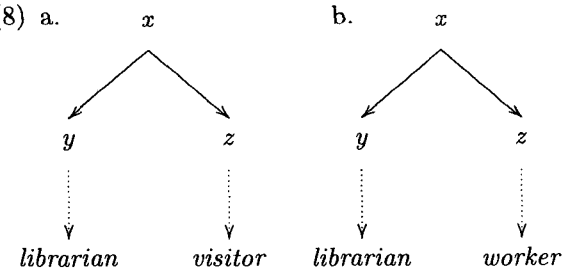
The second sub-assumption yields such partial graphs as (7), when applied to sentences a and b in (1).

(7) a. b.



In a similar way, partial graphs in (8) are motivated by the last sub-assumption when it is applied to a coordinated noun phrase in which conjuncts both contain modified head nouns.

(8) a. b.



(9) Beyond these, ..., stand librarians who ..., social workers who ..., and health visitors who .... (HPX)

This relationship will be referred to as cousin, in what follows.[7]

[7] More strict adherence to geneology would suggest, in addition, second-cousin, third-cousin, etc. All relations that might be termed "n-th cousin-once-removed" or the like are excluded from "cousins", since they are covered by our aoasib.

3. Gathering pairs of head nouns from conjoined NP's

We gather pairs of nouns and their individual modifiedness specifications from a corpus. The process and the results follow.

3.1 The process

Using the British National Corpus, the version called BNC World, we identify all sentences that contain a conjunction. It is evident that conjoined noun phrases are to be found in a subset of these sentences. Since BNC World is not a syntactically parsed corpus, we run a parser on these sentences to get their syntactic structures. [8] Features of this parser have been described in Klein and Manning (2003).

The output of the parser is then converted into an XML format, in order for it to be analyzed efficiently with the help of class libraries like *javax.xml.xpath*, *javax.xml.parsers*, and *org.w3c.dom*. It is on the XML files that conjoined noun phrases are identified and head nouns of conjuncts are paired with each other, with their modifiedness specifications.

Each document contributes a set of pairs to the main storage of pairs. When a pair is to be contributed that is already in the storage, its frequency count is incremented. Thus, the nodes that comprise the main storage are all ordered pairs of the form:

$\langle\langle \text{noun}_1, \text{modifiedness}_1, \text{noun}_2, \text{modifiedness}_2\rangle$, frequency_of_this_quadruple $\rangle$

3.2 Results

3.2.1 The quadruples. We glean slightly over four hundred thousand quadruples. Each quadruple contains two nouns and information about their respective modifiedness. Table 1 shows how many quadruples of different modifiedness patterns are found at each frequency level. [9]

Roughly half of the pairs, i.e. 205,481, are those whose nouns are assumed to stand in a sibling relation to each other. Of the remaining half, the pairs one of whose nouns is assumed to stand in an ancestor-or-ancestor's-sibling relationship to the other outnumber the ones whose nouns are assumed to stand in a cousin relationship to each other: 127,702 vs 70,625. However, the ratio of sibling-pairs out of all pairs goes up with the frequency level. At frequency level of five or above, 73.4% of 7,774 pairs are of a sibling relationship; at that of ten or above, 79.8% of 2,049 pairs and at twenty or above, 83.2% of 927 pairs, respectively, are of a sibling-relationship. The types of nouns involved in one or more of the quadruples are 36,950. This means that each noun type is associated with approximately eleven quadruples on the average.

---

[8] Of the 6,051,206 sentences in the corpus, 835,896 contained a conjunction. It takes thirty two days for the parser to come up with a parse for all the sentences, on a personal computer with a 3 GHz processor and 1 GBytes of memory.

[9] The most frequent pair of nouns turns out to be *man* and *woman*. This pair occurs in at least 615 places. Of these, 440 occurrences indicate that the two words denote concepts in a sibling relationship; 100 indicate that the concept denoted by *woman* would be a descendant of, or a descendant of a sibling of, the one denoted by *man*; 32 indicate the opposite state of affairs; the remaining 43 indicate that they denote concepts in a cousin relationship.

freq	Number of pairs in the semantic relation						total
	sibling	sum(n_1)	cousin	sum(n_2)	aoasib	sum(n_3)	($n_1+n_2+n_3$)
≥20		459		31		45	535
≥15		772		60		95	927
≥10		1636		171		242	2049
9	379	2015	56	227	66	308	2550
8	461	2476	64	291	94	402	3169
7	659	3135	109	400	147	549	4084
6	1003	4138	182	582	215	764	5484
5	1574	5712	303	885	413	1177	7774
4	2823	8535	633	1518	862	2039	12092
3	6079	14614	1468	2986	2180	4219	21819
2	18920	33534	5579	8565	8831	13050	55149
1	171947	205481	62060	70625	114652	127702	403808

[Table 1] Number of pairs at each frequency level and modifiedness pattern

3.2.2 The equivalence classes of quadruples. The relationship between the syntactic notions of patterns of modifiedness and the semantic notions of siblinghood and/or ancestor-or-ancestor's-sibling relation, is one of a tendency rather than of an exceptionless rule. This is most readily proved by the existence of quadruples that involve the same two nouns with different modifiedness specifications. Two such cases are given below.

(10)

−friend −wife 12	−comfort −warmth 8
−friend +wife 1	−comfort +warmth 1
+friend −wife 3	+comfort −warmth 1
+friend +wife 2	

In order to build an is-a hierarchy from the seemingly disorderly quadruples we obtain, an attempt at normalizing them can be made. Stronger tendencies can be identified of each pair of nouns by considering which modifiedness pattern of this pair is especially frequent.

For each equivalence class of quadruples, where the equivalence relation is 'has the same nouns', we apply the following procedure and choose only one modifiedness specification out of the two, three, or four.

1. If the most frequent pattern is *ancestor-or-ancestor's-sibling*(w_1, w_2),

   (a) suggest it as the pattern of the equivalence class if *ancestor-or-ancestor's-sibling*(w_2, w_1) is not in the class and the sum of the frequencies *sibling*(w_1, w_2) and *cousin*(w_1, w_2) is smaller than the frequency of the aoasib pattern.

freq	# of equivalence classes in the semantic relation						total
	sibling	sum(n_1)	cousin	sum(n_2)	aoasib	sum(n_3)	($n_1+n_2+n_3$)
≥20		170		40		1	211
≥15		303		78		6	387
≥10		684		196		32	912
9	189	873	60	256	13	45	1174
8	220	1093	70	326	20	65	1484
7	372	1465	109	435	21	86	1986
6	547	2012	172	607	55	141	2760
5	940	2952	303	910	101	242	4104
4	1827	4779	620	1530	256	498	6807
3	4340	9119	1478	3008	752	1250	13377
2	15273	24392	5810	8818	4308	5558	38768
1	164036	188428	61908	70726	83364	88922	348076

[Table 2] Number of equivalence classes and their dominant semantic relationships

(b) Otherwise, suggest *cousin* instead.

2. If the most frequent pattern is *cousin*,

   (a) if the frequency of the *sibling* pattern is smaller than or equal to the sum of frequencies of all other patterns divided by two, suggest *cousin* as the pattern of the equivalence class.
   (b) otherwise, suggest *sibling* instead.

3. Otherwise, suggest *sibling* as the pattern of the equivalence class.

An application of this procedure weeds out approximately fifty five thousand quadruples. Table 2 shows how the three semantic relations are distributed in the 348,076 equivalence classes at different frequency levels. (Here and hereafter, aoasib stands for "ancestor-or-ancestor's-sibling.")

4. Building a hierarchy from pairs

We aim at building a concept hierarchy with the 348 thousand equivalence classes. The overall procedure of this enterprise can be divided into three steps: (i) Form groups of words which would be direct hyponyms of an unlexicalized concept; (ii) Give further hypernym-hyponym relationships to these groups, creating more unlexicalized concepts as necessary; (iii) Check the resulting homogeneous hierarchy against known restrictions about semantic relations between words and modify parts if necessary.

4.1 How to group quasi-synonyms

The first step can be achieved with the help solely of the sibling classes. If quasi-synonyms are gathered into groups, then a partial graph can be formed immediately. Subgraphs of vertices in which an edge is drawn from a designated concept to all other concepts can be formed immediately. The subgraphs are connected to no other subgraphs yet. It is in the next step that these local graphs are incrementally connected to other ones, ultimately forming a totally connected graph. This step makes crucial use of our *aoasib* classes. If w_1 is of a higher-generation than w_2, and if the former is still relatively close to w_2, the subgraphs to which these two words belong can be connected: w_1's immediate predecessor (or next-to-immediate predecessor) comes to be followed by an immediate predecessor of w_2's immediate predecessor. The last task, namely, (iii) above, can be helped greatly by our equivalence classes whose dominant relationship is one of *is-a-cousin*. Check, for each vertex, whether one of its close relatives that are closer than a cousin, is, in fact, required to be a cousin. If there is such a pair of vertices, reconnect nearby vertices in such a way that the requirement is met and the change remains unsubstantial.

Here, we will not have the space in which to discuss all these issues in detail. We will, however, describe the first step above in some detail and give a portion of our result.

Due to polysemy, and, more important, to ambiguity, of words, the sibling relation we use is not transitive: Given $[\text{sibling}(w_1, w_2) \wedge \text{sibling}(w_2, w_3)]$, it does not follow that $\text{sibling}(w_1, w_3)$. [10] We need a way of grouping words from pairs of siblings which share one word: from the set $\{\text{sibling}(w, x_1), \text{sibling}(w, x_2), \ldots \text{sibling}(w, x_n)\}$, get the sets $\{w, \ldots x_i, \ldots\}$, $\{w, \ldots x_j, \ldots\}$, etc., separately, if the words x_i and x_j are close only to one of the meanings w has. This process of selective grouping of siblings requires careful programming.

We assume that no two words are ambiguous in an exactly parallel fashion. If w is ambiguous (or, polysemous for that matter) in two or more ways such that it denotes either of concepts c_1 and c_2, there is assumed to be no other word x such that x also denotes either of the two concepts. On this assumption, we require there to be at least two words in the basket which have the candidate word as a sibling before the candidate is added to the basket. The initial basket contains two words from a sibling pair. The third word must be siblings of both words that are already in the basket. Further additions to the basket can be subject to a more strict condition.

The total number of sibling groups and their average sizes depend on the condition imposed on additions and the frequency threshold of the seed pairs used. When the seeds are of frequency at or above 2 and the condition on additions is quite strict, the procedure yields 20,781 groups. Of these, 87 have six or more members, of which twenty are shown in Table 3, 119 have five members, 410 have four, and 1,836 have three members. 18,329 are sets with a modicum of two members. The less strict the condition on additions is, the greater the average size of the groups and the smaller the number of groups with just two members.

[10] An example of such pairs is: {bank, ditch} and {bank, shop}

[ability_background] ability background interest experience temperament talent

[advice_education] advice education experience skill work idea

[alveolus_canker] alveolus canker pimple ventricle hole pustule

[anxiety_confusion] anxiety confusion fear despair disgust excitement

[architecture_building] architecture building furniture painting theater church design fashion

[arsenic_cadmium] arsenic cadmium mercury silver copper iron lead

[athletics_badminton] athletics badminton football basketball boxing cricket

[beer_cider] beer cider spirit cigarette gin lager

[business_commerce] business commerce law government industry health research training hospital

[care_equipment] care equipment labor management worker food

[color_fabric] color fabric shape habit pattern scent

[culture_economy] culture economy history society religion service

[doctor_employer] doctor employer public school hospital parent

[education_finance] education finance maintenance service housing infrastructure practice rehabilitation sale

[engineer_journalist] engineer journalist teacher judge nurse politician

[ex-services_journalist] ex-services journalist lawyer youth police teacher

[family_guest] family guest staff planning prisoner relative

[film_metal] film music news radio story play print

[information_planning] information planning staff training publication supervision system

[music_name] music name sex poetry politics rock

[Table 3] Sibling groups with six or more members

4.2 Evaluation of the sibling groups

We evaluated the results by comparing English speakers' intuitions of in-group similarity of selected groups yielded by the procedure in section 4 to those of similarity of randomly chosen words that share a hypernym. Specifically, we chose ten groups of ours, ten groups of random words that are hyponyms of synonym groups at the level of four and five, respectively, of the WordNet hierarchy.

Thirty groups of words, each numbered 1 through 30, were presented to our subjects, who were asked to indicate, for each group, whether they think each word

	Pairs of conjuncts	WordNet (random 4)	WordNet (random 5)
Related	83	65	92
Unrelated	17	35	8

[Table 4] Responses of native speakers of English to three kinds of word groups

in the group denotes a concept that is close to the concepts the other words in the group denote and that the relations between any two of the concepts are the same in the given group.[11] The subjects were asked to check one of the two options: (a) Similar (b) Not similar.

The result, in Table 4, shows that groups that have been collected in the way described in this paper fare better than ones of randomly selected words from WordNet concepts at level 4. Random collections of words associated with WordNet concepts at level 5, however, seem closer than ours on the average.

4.3 How to get hypernym-hyponym relations

The many quasi-synonym groups formed in our first phase are to be connected to other groups. The 38,768 equivalence classes with aoasib as the dominant relationship play a crucial role in this process.[12]

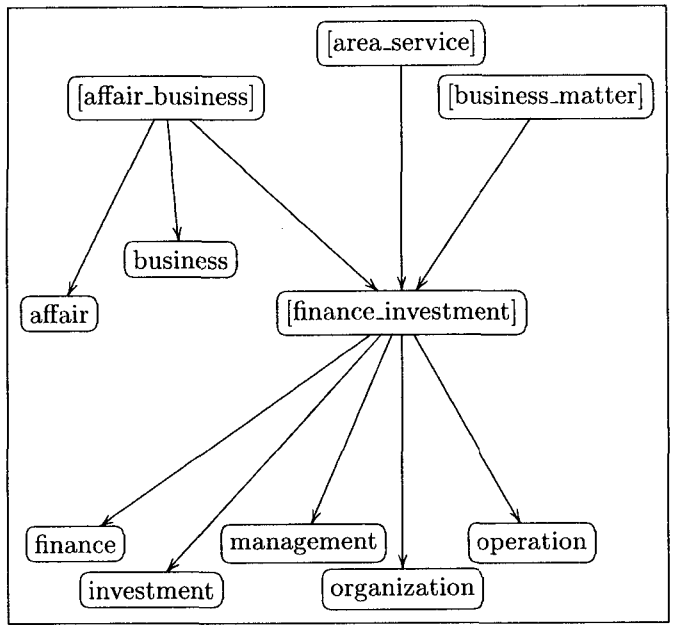
For each concept c, identify the set of words shared by c's immediate descendants' aoasibs. If there is a nonempty set of shared words, $\mathcal{A}$, locate all vertices that are immediate hypernyms of nonsingleton subsets of $\mathcal{A}$. Connect each of these vertices to c. As a concrete example, let us try to locate vertices corresponding to the immediate hypernyms of the nonlexicalized concept [finance_investment]. This concept has, as its direct hyponyms, *finance, investment, research, training, management, organization, technology, operation, planning, safety, staff, staffing, structure*, and *teaching*. Many of these hyponyms have, as their aoasibs, *affair, area, business, matter*, and *service*. Among the nonsingleton subsets of this set are {affair, business}, {area, service}, and {business, matter}. we locate the direct hypernyms of all members of each of the subsets and make the concept under focus, [finance_investment], a direct hyponym of theirs. Figure 1 shows these connections.

The problem of finding hypernym-hyponym pairs from a collection of aoasib, cousin, and sibling relations can be recast as a constraint resolution problem. The following Prolog program would do the job. (Note that the first clause relies on our assumption that the relationships gleaned from any finite corpora are a proper subpart of those gleanable from the language.)

```
ancestor(x,y) :-
  aoasib(x,y), sib(y,w), aoasib(x,w).
mother(x,y) :-
  ancestor(x,y), sib(x,w),
  sib(y,z), aoasib(w,z), ~ancestor(w,z).
```

[11] The subjects are ten students, all native speakers of American English, from a university in the States.

[12] We are confining ourselves to tuples of frequency count 2 or above.



[Figure 1] The part of the hierarchy showing that *finance* is a hyponym of *affair*'s hypernym

Dowling and Gallier (1984) show that the time required for testing satisfiability of Horn clauses is $\mathcal{O}(N)$, where N is the number of atomic sentences in the clause. This means that, for each pair $\langle w_1, w_2 \rangle$ of concepts, it takes a constant time to decide whether w_1 is an ancestor of w_2. The same holds for a decision as to whether w_1 is a[sic.] mother of w_2.

5. Conclusion

Pairs of conjuncts can be a great source for an automatic construction of an is-a hierarchy. The methods described in this paper can be used to simplify the construction of an ontological database. While we concentrated on the English nouns, essentially the same approach can be taken to other parts of speech in English. Of course, we believe all natural languages have constructions that serve the same functions as the coordination constructions in English. The level of success in the construction of an is-a hierarchy for the lexicon of another language, hopefully, will be similar to that for English nouns.

If we are on the right track, the difficulty associated with "word clustering" algorithms, which are based on cooccurrence patterns, can simply be avoided. Apart from the independently available syntax parser, we do not need anything that would burden the computer with $\mathcal{O}(V^3)$ time complexity and $|V|^2/4$ space requirement, in order to acquire an is-a hierarchy.

In a more theoretical vein, this work naturally leads us to the generaliza-

tion: pairs of words in any of all interesting semantic relationships tend to occur in the same sentence. Adjectival antonymy is seen as one such relation (Charles and Miller, 1989); hypernymy-hyponymy is another (Hearst, 1992). We've added siblinghood, cousinhood, and the disjunctive aoasib to this repertoire. Now it is extremely likely that all semantic relations between a pair of words exhibit themselves in the same sentence far more frequently than by chance.

<References>

Brown, Peter F., Vincent J. Della Piertra, Peter V. deSouza, Jenifer C. Lai, and Robert L. Mercer. 1992. Class-based *n*-gram models of natural language. *Computational Linguistics*, 18.4: 467–479.

Caraballo, Sharon A. 2000. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics.*

Charles, Walter G. and George A. Miller. 1989. Contexts of antonymous adjectives. *Applied Psycholinguistics*, 10.3: 357–75.

Cimiano, Philipp, Johanna Völker, and Rudi Studer. 2006. Ontologies on demand? – A description of the state-of-the-art, applications, challenges and trends for ontology-learning from text. *Information, Wissenschaft und Praxis*, 57.6-7: 315–320.

Dowling, W. and J. Gallier. 1984. Linear time algorithms for testing the satisfiability of propositional Horn formulae. *Journal of Logic Programming*, 3: 267–284.

Fellbaum, Christiane. 1993. English verbs as a semantic net. In: Five Papers on WordNet.

Fellbaum, Christiane, editor. 1998. *WordNet: An Electronic Lexical Database.* MIT Press.

Finch, S. 1993. *Finding Structure in Language.* Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Hearst, Marti A. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics.*

Klein, Dan and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics.*

Li, Hang and Naoki Abe. 1998. Generalizing case frames using a thesaurus and the MDL Principle. *Computational Linguistics*, 24.2: 217–244.

Lin, Dekang. 1998. Automatic retrieval and clustering of similar words. In *COLING-ACL '98*, pp. 768–773.

Miller, George A. 1993. Nouns in WordNet: a lexical inheritance system. In: Five Papers on WordNet.

Resnik, P. 1995. Using information content to evaluate semantic similarity. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 448–453.

Rosario, Barbara, Marti A. Hearst, and Charles Fillmore. 2002. The descent of Hierarchy, and selection in relational semantics. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 247–254.

Zwicky, Arnold M. 1985. Heads. *Journal of Linguistics*, 21.1: 1–29.

⟨**Web sites**⟩
http://www.natcorp.ox.ac.uk British National Corpus
http://nlp.stanford.edu/software/lex-parser.shtml The Stanford Natural Language Processing Group

Submitted on: April 23, 2007
Accepted on: May 21, 2007