Exploiting Term Relations for Semantic Hierarchy Construction

Cliff Joslyn, Patrick Paulson
Knowledge Systems Group
Pacific Northwest National Laboratory
PO Box 999, MS K7-28
Richland, WA, 99352, USA
{cjoslyn,patrick.paulson}@pnl.gov

Karin Verspoor
Center for Computational Pharmacology
University of Colorado Denver
PO Box 6511, MS 8303
Aurora, CO 80045-0511 USA
karin.verspoor@ucdenver.edu

Abstract

This paper presents a method to induce semantic taxonomies by applying the lattice theoretical technology of Formal Concept Analysis to relations of predicates extracted from a natural language corpus. Our initial research results are in support of a future overall methodology for the semi-automatic construction of semantic hierarchies from term relations extracted from text. We describe our formal method for hierarchy construction, selection and processing of a test corpus for extracting verb-noun pairs from natural language, measurement and filtering of the resulting verb-noun term matrices for density optimization, and look at the resulting semantic hierarchies produced by FCA.

1. Introduction

The central mechanisms to represent organized knowledge, and the cores of real-world ontologies, are **semantic hierarchies:** mathematically ordered typing structures relating terms or concepts of a domain within semantic relations like subsumption (taxonomic, "is-a", generality relations) and composition (meronomic, "has-part", part-whole relations). The vast increase of computational information in science has resulted in a proliferation of ontologies of increasing size, in areas ranging from bioscience to the military and intelligence communities to the World Wide Web. Successful ontologies like the GO (http://www.geneontology.org) and Wordnet (http://wordnet.princeton.edu) have grown to be very large, with over 20K and 150K nodes respectively.

The automated creation and management of large ontologies is thus a critical need for the knowledge sciences [13]. Source information for ontology induction could be user elicitation, structured data, or other sources, but one prominent and obvious source is text. Ideally, lexical and linguistic information can be gleaned from texts in a way to automatically or semi-automatically determine subsumptive and meronomic relations among terms and concepts.

Research in ontology construction from text draws on a number of strategies. Following the **distributional similarity hypothesis** [7, 9] terms are semantically similar if they share linguistic contexts like co-occurrence with the same terms, or occur in the same syntactic relation to a term. Thus agglomerative clustering [2] of terms can be accomplished based on using similarity of term usage vectors capturing the linguistic contexts of the term. Alternatively, particular phrasal structures can be identified that stereotypically distinguish specific from general terms [8], and can be captured by schemata or templates such as "X, Y, and other Z" (e.g. "Cats, dogs, and other animals") [12].

Finally, methods from **order** and **lattice theory** [1], specifically **Formal Concept Analysis** (**FCA**) [5, 6], can be employed to discover inherent relations between objects described through a set of attributes. FCA is a discrete math method for representing the set of sub-relations present in an overall binary relation (say, among lexical terms) in the context of their hierarchical structure, and thus yields natural semantic hierarchies from term relations.

While FCA is being increasingly recognized as a canonical tool for representing relational information in hierarchical structures and automatically inducing taxonomic ontologies from text [3], existing methods are still limited:

- There is no attention to the dependence of the quality of the induced ontology on the linguistic properties of the text, e.g. the impact of lexical variation (different words expressing similar meanings), or how much redundancy and data density is required to produce a non-trivial concept lattice.
- Focus is almost always on relations among words of the same part of speech, ignoring relations that cross parts of speech and thus the critical semantic relations such as constraints on the classes of nouns which are valid arguments to verbs.

978-0-7695-3279-0/08 $25.00 © 2008 IEEE
DOI 10.1109/ICSC.2008.68

	Dog	Goat	Horse	Sheep
Name	√		√	√
Collar	√	√		√
Milk		√	√	√
Pat	√	√		√

Table 1. A formal context K.

- There is almost exclusive focus on unary predicates (properties of objects), with higher order n-ary predicates expressing critical semantic relations among multiple objects collapsed to the simpler form.

Our approach extends the intuition of the distributional similarity hypothesis to assert a **Distributional Generality Hypothesis**, that *semantic generality of terms, as represented by hierarchical relations among them, can be determined from the analysis of their shared linguistic contexts*. In particular, to the extent that the more general a term is, the more contexts it can be used in, then we aim to show that the distribution of nouns with respect to the verbs they are arguments of, and *vice versa*, is an important source of information about term generality that can be used to construct a semantic hierarchy of relations among terms.

This paper reports on some initial research results supporting a future overall methodology towards the semi-automatic construction of semantic hierarchies from term relations extracted from text. Specifically, we describe our formal method for hierarchy construction, selection and processing of a test corpus for extracting verb-noun pairs from natural language, measurement and filtering of the resulting verb-noun term matrices for density optimization, and finally look at the resulting semantic hierarchies produced by FCA.

2 Concept Lattices and Semantic Hierarchies

Our method for initial construction of semantic hierarchies is based on FCA, which we now very briefly describe, see [5, 6] for complete details.

FCA casts Boolean matrices as **formal contexts** $K := \langle G, M, I \rangle$, where G is a finite set of objects, M is a finite set of attributes of those objects, and $I \subseteq G \times M$ is an incidence relation, so that for $g \in G, m \in M, \langle g, m \rangle \in I$ means that object g has attribute m. An example formal context K is shown in Table 1.

Given a formal context K, FCA automatically constructs all possible permutations of rows and columns to identify "maximal rectangles" of checkboxes, and then organizes these into a semantic hierarchy, a **concept lattice** $\mathcal{L} = \langle P, \leq \rangle$, where $P \subseteq 2^G \times 2^M$ is a set of **concepts** generated by I, so that for $A \subseteq G, B \subseteq M, \langle A, B \rangle \in P \iff A = \{g \in G : \forall m \in B, \langle g, m \rangle \in I\}$ and $B = \{m \in M : \forall g \in A, \langle g, m \rangle \in I\}$; that is, each concept is a pairing of all objects $A \subseteq G$ which have exactly a certain set of properties $B \subseteq M$, and (importantly) *vice versa*. The order $\leq$ on the concepts in P is then $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ iff $A_1 \subseteq A_2$. Necessarily $A_1 \subseteq A_2 \iff B_1 \supseteq B_2$.

Fig. 1 shows the concept lattice $\mathcal{L}$ for a formal context K shown in Table 1. $\mathcal{L}$ (also called a "Galois lattice") represents all the embedded sub-relations present in the original relation I, and thus an empirically derived taxonomy, dually cataloging both the collections of objects which have certain attributes and collections of attributes which hold for certain objects.

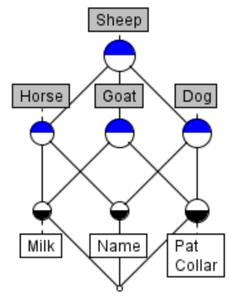

Figure 1. Concept lattice $\mathcal{L} = \langle P, \leq \rangle$ of the formal context K.

For any real matrix $A_{N \times M} = [a_{i,j}]_{i=1}^{N}{}_{j=1}^{M}$, let $\|A\| := \sum_{i=1}^{N} \sum_{j=1}^{M} a_{i,j}$. Then for a Boolean matrix $B_{N \times M}$ with $b_{i,j} \in \{0, 1\}$, let $D(B) := \frac{\|B\|}{NM}$ be the **density** of B. Then note the sensitivity of $\mathcal{L}$ on the density $D(K)$. For example, consider square matrices with $N = M$, and with no non-empty rows or columns. Then $D(K)$ is minimal at $1/N$ if the rows and columns of K can be permuted to make a diagonal matrix. But the resulting lattice $\mathcal{L}$ has no structure, just a set of disconnected concepts. But if K is full so that $D(K) = 1$ is maximal, the lattice has a single node. Thus we generally desire intermediate densities $D(K)$ to produce optimal lattices $\mathcal{L}$.

Specific object and attribute semantic hierarchies are implied by a concept lattice. Observe that each object $g \in G$ and attribute $m \in M$ has a unique concept node denoted $\gamma(g), \mu(m)$ respectively. Taken separately, these form distinct object and attribute hierarchies according to a **conceptual pre-order** [10], where $m_1 \leq m_2 := \mu(m_1) \leq \mu(m_2)$, while dually $g_1 \leq g_2 := \gamma(g_1) \geq \gamma(g_2)$, as shown in Fig. 2 for our example.

3 Method

Parse a corpus to produce a list $V = \{v_i\}_{i=1}^{n}$ of verbs and $O = \{o_j\}_{j=1}^{m}$ of nouns. Let c_i be the count (num-

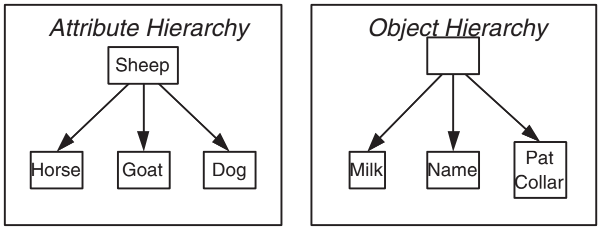

Figure 2. Term hierarchies determined by the conceptual pre-order of $\mathcal{L}$.

	c_j	76	52	43	41
c_i	v_i/o_j	Dog	Goat	Horse	Sheep
82	Buy	26	23	17	16
38	Name	14		19	5
37	Collar	30	4		3
30	Milk		15	7	8
25	Pat	6	10		9

	j	1	2	3	4
i	v_i/o_j	Dog	Goat	Horse	Sheep
1	Buy	2	3	5	6
2	Name	8		4	14
3	Collar	1	15		16
4	Milk		7	12	11
5	Pat	13	9		10

Table 2. (Top) Noun, verb, and pair counts c_j, c_i, and c_k. (Bottom) Noun, verb, and pair ranks j, i, and k.

ber of appearances in the corpus) of v_i and c_j be the count of o_j. Let V and O be sorted down so that v_1 and o_1 are the most frequently occurring, and i and j are the verb and noun **ranks**. Further assume that the corpus is also parsed to produce a list $P = \{p_k\}_{k=1}^{l}, p_k = \langle v_i, o_j \rangle$ of unique noun-verb pairs appearing. Note that P is a binary relation $P \subseteq V \times O$, and $l \leq nm$. Let c_k be the count of the pair p_k, and the pairs P also be sorted down by count producing the pair ranks k.

There is a mapping $k \mapsto \langle i, j \rangle$, defined as a vector function $f \colon \mathbb{N}_l \to \mathbb{N}_n \times \mathbb{N}_m$ (where for any natural number $n, \mathbb{N}_n := \{1, 2, \ldots, n\}$), so that $f(k) := \langle i, j \rangle$. Further say $f_1(k) := i, f_2(k) := j$, so that k, $f_1(k)$, and $f_2(k)$ are now the pair, verb, and noun ranks.

As an example, consider Table 2 showing on the top pair counts for a set of sorted noun and verb counts, and in the bottom the same information for ranks. We have $l = 16 < (n = 5) \times (m = 4) = 20$. Table 3 then shows the mapping

k	$f_1(k) = i$	$f_2(k) = j$
1	3	1
2	1	1
3	1	2
4	2	3
5	1	3
6	1	4
7	4	2
8	2	1
9	5	2
10	5	4
11	4	4
12	4	3
13	5	1
14	2	4
15	3	2
16	3	4

Table 3. Noun and verb ranks as a function of pair ranks.

$k \mapsto \langle i, j \rangle$.

Any subrelation $\rho \subseteq P$ of the pairs determines a projection onto the sets of nouns, verbs, and pairs appearing in ρ. So given such a $\rho \subseteq P$, let

$$\nu := \{i \in \mathbb{N}_n : \exists p_k \in \rho, f_1(k) = i\},$$

$$\mu := \{j \in \mathbb{N}_m : \exists p_k \in \rho, f_2(k) = j\},$$

$$\Lambda := \{k \in \mathbb{N}_l : \exists p_k \in \rho\}$$

be the sets of indices of the verbs, nouns, and pairs, respectively, appearing in ρ. Casting P and ρ as matrices, then ν is the set of row indices, μ the column indices.

Let $1 \leq N_* \leq N^* \leq n, 1 \leq M_* \leq M^* \leq m$ be upper and lower cutoffs on the ranks of verbs and nouns being considered. $N' := N^* - N_* + 1 \leq N$ and $M' := M^* - M_* + 1 \leq M$ are the total number of verbs and nouns respectively, and denote $\bar{N} = [N_*, N^*], \bar{M} = [M_*, M^*]$. Then construct the **count matrix** $A_{N' \times M'}$ with rows $v_{i'}, 1 \leq i' \leq N'$ for verbs and columns $o_{j'}, 1 \leq j' \leq M'$ for nouns, letting

$$a_{i,j} := \begin{cases} c_k, & \exists p_k = \langle v_i, o_j \rangle \\ 0, & \nexists p_k = \langle v_i, o_j \rangle \end{cases} .$$

Note that i', j' are the indices in the matrix A constrained by the selection of $\bar{N}, \bar{M}$ to reference some absolute ranks i, j. It follows that

$$\nu = \{N_*, N_* + 1, \ldots, N^* - 1, N^*\},$$

$$\mu = \{M_*, M_* + 1, \ldots, M^* - 1, M^*\}.$$

		j'	1	2	3	4
		j	1	2	3	4
i'	i	v_i/o_j	Dog	Goat	Horse	Sheep
1	2	Name	8		4	14
2	3	Collar	1	15		16
3	4	Milk		7	12	11
4	5	Pat	13	9		10

Table 4. Rank matrix $A_{N' \times M'}$.

The count matrix A generates the Boolean matrix $B_{N' \times M'}$ by

$$b_{i,j} := \begin{cases} 1, & a_{i,j} \neq 0 \\ 0, & a_{i,j} = 0 \end{cases} .$$

To ease notation, we will say $p_k \in B$ to mean that $\exists i \in \nu$ and $\exists j \in \mu$ such that $p_k = \langle v_i, o_j \rangle$; that is, for the verb v_i and noun o_j, the pair $p_k = \langle v_i, o_j \rangle$ appears in the corpus.

To continue our example, let $\bar{N} = [2,5], \bar{M} = [1,4]$, so that $N' = M' = 4$, yielding A as shown in Table 4, and B as the formal context K back in Table 1. We then have $D(B) = 75\%, \Lambda = \{1,4,7,8,9,10,11,12,13,14,15,16\} \subseteq \mathbb{N}_{16}$.

As can be seen in Table 4, in this formulation, whereas the verb indices i and noun indices j are constrained so that $i \in \nu, j \in \mu$, the pair indices k for which $p_k \in B$ can be distributed anywhere in $\mathbb{N}_l$, so that $\Lambda \subseteq \mathbb{N}_l$ is some arbitrary subset. One might imagine that the $k \in \Lambda$ are relatively low and contiguous in $\mathbb{N}_l$, with $\max_{k \in \Lambda} \sim \|B\|$. In our test corpus, we have seen in practice that this is not the case at all.

So given a corpus, a relation extraction method, and a selection of noun and verb ranks $\bar{N}, \bar{M}$, resulting densities $D(B)$ may or may not be appropriate to generate meaningful concept lattices. To the extent that B is dense, the ranks k of pairs might be expected to be distributed according to $f_1(k) \times f_2(k)$, the product of the noun and verb ranks, as is the case in Table 2. This observation brings us to consider some filtering methods which constrain Λ to some subset $\Lambda' \subseteq \Lambda$ related to the products of these ranks:

1. **No Restriction:** Let $\Lambda' = \Lambda$.
2. **Max Row/Column Rank:** Let $\Lambda' = \{k \leq \max(N^*, M^*)\}$.
3. **Max Row/Column Rank Product:** Let $\Lambda' = \{k \leq N^* M^*\}$.
4. **Row/Column Rank Product:** Let $\Lambda' = \{k \leq f_1(k) f_2(k)\}$.
5. **Row/Column Rank Max Squared:** Let $\Lambda' = \{k \leq [\max(f_1(k), f_2(k))]^2\}$.

	Dog	Goat	Horse	Sheep
Name			√	
Collar	√			
Milk				
Pat				
	Dog	**Goat**	**Horse**	**Sheep**
Name			√	
Collar	√			
Milk		√	√	√
Pat		√		√
	Dog	**Goat**	**Horse**	**Sheep**
Name			√	√
Collar	√			√
Milk		√	√	√
Pat	√	√		√

Table 5. B produced from Λ filtering methods 2 (top), 4 (center), and 5 (bottom).

6. **Max Row/Column Rank Max Squared:** Let $\Lambda' = \{k \leq [\max(N^*, M^*)]^2\}$.

Continuing our example, only methods 2, 4, and 5 above produce a $\Lambda' \subset \Lambda$ (this is not true in general), and these are shown in Table 5. Fig. 3 shows the resulting concept lattices for Λ' determined by methods 4 (left) and 5 (right).

4 Test Corpus Selection and Processing

Optimal density attainment for the FCA methodology for semantic hierarchy calculation depends on relational redundancy in the input matrices, that is, row terms which have a moderate amount of overlap with column terms. Thus our goal in a test corpus selection is that it be neither so small that terms only pair individually, nor so complex that either syntactic structure or shared linguistic context cannot be identified.

We selected the TASA corpus[1] [14], a carefully constructed corpus of samples of texts used in schools and colleges in the United States. This corpus was originally collected to estimate the frequency of words encountered by school-age American children. It consists of 37,651 documents, just over 12 million word tokens, and 92,409 word types (cf. the British National Corpus with over 315,000 word types).

The rationale behind this selection is that (a) the size of the vocabulary is smaller than in generic English text, (b) the size of the corpus is adequate to contain adequate variation in the usage contexts of individual terms, and (c) the grammatical complexity is likely to be simpler than in other

[1] `http://www.tasaliteracy.com/wfg/wfg-main.html`

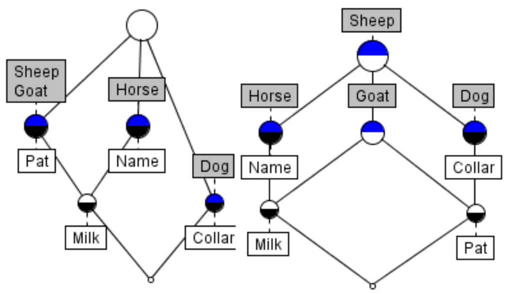

Figure 3. Concept lattices produced from filtering methods 4 (left) and 5 (right).

Rank	o_j	c_j	v_i	c_j	p_k		c_k
1	person	376115	be	294904	be	thing	43258
2	thing	181544	have	96097	be	person	27307
3	people	141553	say	47230	have	person	16851
4	time	9311	make	39502	say	person	13685
5	way	8501	see	34007	see	person	13008
6	day	6781	take	31198	know	person	11561
7	part	6653	do	31077	think	person	8975
8	some	6049	go	26222	go	person	8596
9	water	5484	get	24511	be	people	8499
10	something	5328	know	22593	do	person	7949
11	child	5028	come	22333	have	people	7794
12	mother	4990	use	20814	want	person	7697
13	kind	4856	find	19910	tell	person	7569
14	word	4849	give	19729	do	thing	7296
15	food	4749	call	19157	get	person	6741
16	all	4734	tell	18280	find	person	5234
17	place	4545	want	15762	take	person	5187
18	father	4324	think	15296	look	person	4969
19	boy	4064	look	15060	make	person	4716
20	problem	3928	need	11242	come	person	4226

Table 6. Top 20 nouns (left), verbs (center), and pairs (right) in test corpus.

corpora, due to its orientation to younger readers.

The impact of these corpus characteristics for this experiment is that we can expect reliable parse analysis of the text, and we can expect that an individual verb will co-occur with a variety of nouns while still containing sufficient lexical redundancy to provide evidence of the usage patterns of specific terms.

To process the text, we employed the Stanford Parser [11], a program that analyzes the grammatical structure of a sentence. This parser includes functionality to produce the dependency structure of the sentence, typed by grammatical role. So, for instance, for sentence (1)

```
(1) He collects small crocodiles from
the wild.
```

the dependency relations identified by the parser are:

```
nsubj(collect-2, he-1)
amod(crocodile-4, small-3)
dobj(collect-2, crocodile-4)
det(wild-7, the-6)
prep_from(collect-2, wild-7)
```

where the relational predicate `nsubj` stands for nominal subject, `amod` for adjectival modifier, `dobj` for direct object, `det` for determiner, and `prep_from` for a "from" prepositional relation. The numbers `-d` attached to each word represent the position of the word in the sentence. We remove these positional markers, and only keep the `nsubj` and `dobj` relations as we wish to focus on verb-noun pairs. Thus for sentence (1) we extract two (verb, noun, relation type) triples:

```
(collect, he, subject)
(collect, crocodile, object)
```

We parse the complete TASA corpus using the Stanford Parser, post-process each parsed sentence in the corpus as indicated above, and count the number of occurrences of each (verb, noun, relation type) triple in the corpus. We store each unique triple and its frequency count in a database that will then form the input to the FCA analysis.

For our basic analysis, we ignore the relation type and merge subject and object occurrence counts together prior to analysis. Additionally, we merge pronoun references as the semantic distinctions between them are not important for the purpose of this analysis. Specifically, we map all occurrences of:

- The words "he", "you", "I", "she", "one", "who", and "man" to "person";
- The words "they" and "we" to "people"; and
- The words "it", "that", "this", "what", and "which" to "thing".

This transformation will result in a somewhat denser matrix around particularly common words, while retaining the semantics implied by the pronouns.

The natural language processing described above results in a database of n =43,702 nouns, m =18,841 verbs, and l =541,146 noun-verb pairs. Tables of the top 20 nouns, verbs, and pairs is shown in Table 6. The total matrix is thus very sparse, filling only .07% of the approximately 823 million available cells.

5 Contexts and Lattices

We now show a partial analysis of some of the formal contexts and concept lattices available within this framework. In particular, in this paper we focus only on the top 50 square sub-matrices $\rho \subseteq P$ for $\bar{N} = \bar{M} = [1, N^*] =$

$[1, M^*], N^* = M^* \in \mathbb{N}_{50}$. Density calculations for these contexts are shown in Fig. 4. The top curve shows method 1 (no filtering, $\Lambda' = \Lambda$), revealing that the upper-left corner of the matrix with low i and j is very dense, dropping to only 94.1% density for $N^* = M^* = 50$. The results for methods 3 and 6 are identical to each other in this case, since these would vary only for rectangular matrices.

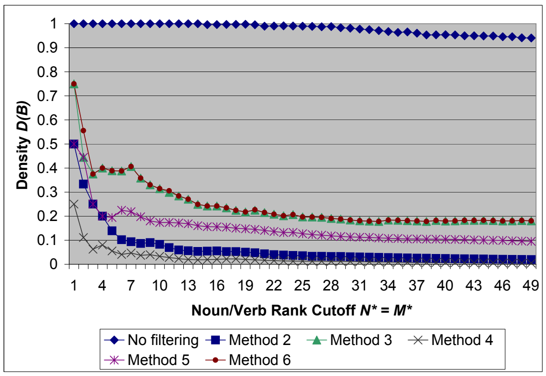

Figure 4. Densities for square matrices B with $\bar{N} = \bar{M} = [1, N^*] = [1, M^*]$.

Space precludes a very detailed discussion of the density results, but inspection reveals that methods 2 and 4 produce matrices and lattices which are too sparse, while those from methods 3 and 6 are too dense. While in practice the criteria for optimal density requires more analysis, the row/column rank max squared approach (method 5), where $k \leq [\max(f_1(k), f_2(k))]^2 = \max(i, j)^2$, produces quite good results.

The main example considered is shown in Fig. 5, for the square context determined by $\bar{N} = \bar{M} = [1, 37]$ and filtering method 5. The resulting verb and noun taxonomies are shown in Fig. 6. It reflects all of the the dual subsumption relations present among the 37 nouns and verbs considered.

The most important factor to be taken into account when interpreting the resulting lattice is the extensional nature of the subsumption relations represented in the context K. For example, in the upper left of Fig. 5, "come" and "go" are asserted to be more general verbs than "get". Tracing downwards from "get" and listing the nouns encountered, we can see that the noun set, called the **extension**, is { home, people, person }, while that for "come" and "go" is { home, people, person, thing }. In other words, "come" and "go" are paired with all of the verbs that "get" is, plus some additional ones.

This expresses the semantics of class inclusion by extension: a class (a verb, in our case) is more general than another if it holds for strictly more objects. And dually, an object (a noun, in our case) is more general if it holds for more verbs (if its intension is larger). This is the inherent Leibnizian semantics of FCA, reflecting the generality of objects and attributes in terms of dual inclusions of extensions and intensions. While other semantics of subsumption are available and sometimes desirable in knowledge systems, this semantics is also fundamental and necessary.

As another example within the lattices available, Fig. 7 shows the lattice for $N^* = M^* = 16$ on the left, and $N^* = M^* = 17$ on the right, both with filtering method 5. Moving from 16 to 17 involved bringing in the noun o_{17} ="place" and the verb v_{17} ="want", and the new pairs shown in Table 7. "Want" introduces no new information, adding in to the large block of verbs associated with the noun set $o_{1-3} = \{$ person, people thing $\}$. But note the introduction of "place" as a new object in the lattice, a specialization of "thing", and how the additional pairing of "take" now with "place" p_{76} removes it from being a sub-class of the "tell" block to being its own "top-level" verb. These are the kinds of changes introduced both from changing matrix size, and from changing filtering methods for the same underlying corpus and extraction methodology.

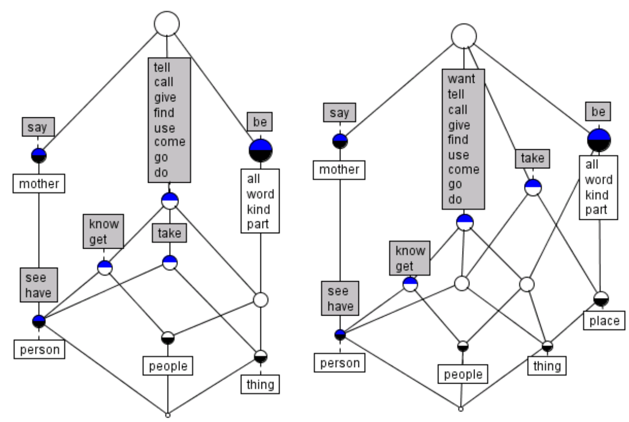

Figure 7. Lattice for $N^* = M^* = 16$ (left), and $N^* = M^* = 17$ (right), both filtering method 5.

6 Discussion and Future Work

The work presented here represents a very first report of our attempt to make a systematic analysis of the properties of semantic hierarchies produced by a pure FCA approach in the context of paired term extraction from a corpus. It is obvious that a more thorough analysis of the properties

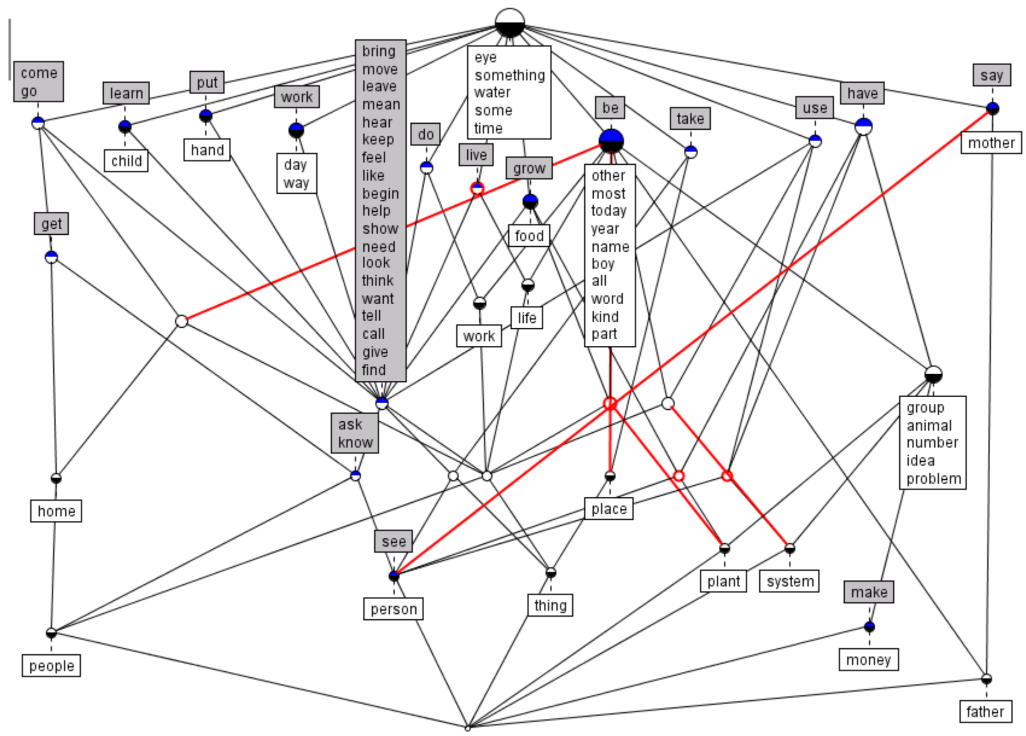

Figure 5. Concept lattice for $\bar{N} = \bar{M} = [1, 37]$, and filtering method 5.

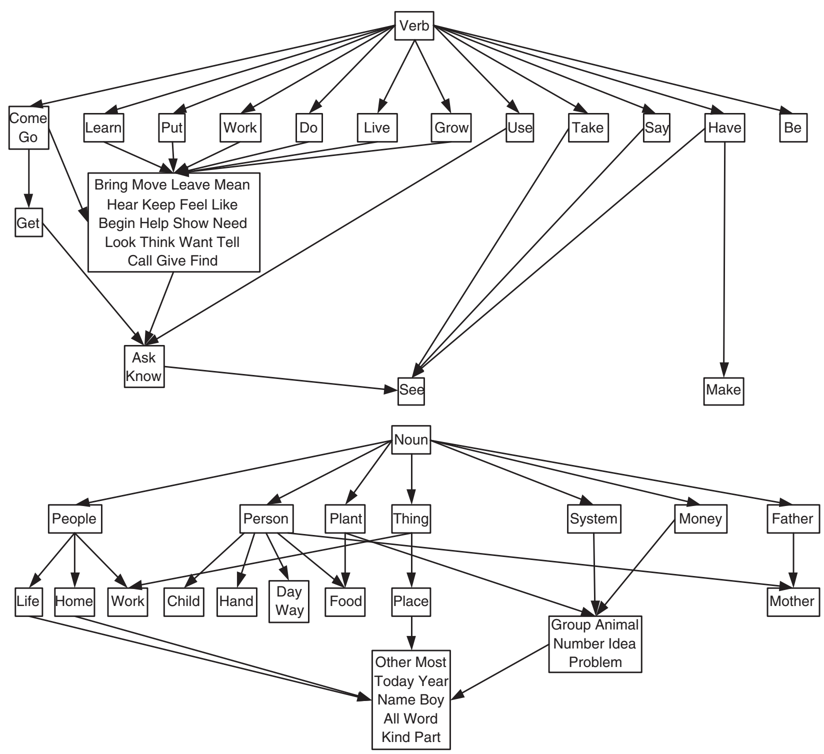

Figure 6. (Top) Verb taxonomy. (Bottom) Noun taxonomy.

k	j	o_j	i	v_i
12	1	Person	17	Want
49	3	Thing	17	Want
76	17	Place	6	Take
87	17	Place	1	Be
182	2	Thing	17	Want

Table 7. Pairs introduced when moving from $N^* = M^* = 16$ **to** $N^* = M^* = 17$**.**

of the resulting contexts and lattices, as a function both of context size and filtering approaches, is required.

Additionally, exploration of non-square contexts and contexts which are not strictly in the upper-left of the main matrix (that is, for which $N_*, M_* \neq 1$) are also necessary. In particular, density falls off dramatically as N_*, M_* grow, and there will likely be very interesting semantic structure represented by these noun-verb combinations away from the most common.

The general issue of sensitivity to corpus selection and NLP techniques employed is also quite a large one. In particular, in this analysis it can be seen that both `nsubj` (nominal subject) and `dobj` (direct object) relations are present in our contexts. Separate analyses for each of these verb-noun pairings is also in order, as well as noun-adjective pairs.

The kinds of extensional, subsumptive hierarchical relations produced in FCA need not necessarily map well to our linguistic senses of generality, or the semantics of "is-a-kind-of". That said, it is clear from our analysis that in fact, more general nouns and verbs do rise to the top of their respective hierarchies. For example, while it is not true in context 37, by the time we reach context 50, for filtering method 5, "be" is at the top of the verb hierarchy, while from the beginning our pronominal nouns "person", "people" and "thing" dominate the top of the noun hierarchy. These are all expected.

Finally, it is clear that for all the formal strengths provided by this methodology, evaluation and interpretation are central to its success and furtherance. We intend to bring this work fully into the context of the ongoing efforts within the ontology community to compare semantic structures against each other for validation [4], for example by developing valid measures to compare our generated taxonomies against standards semantic taxonomies.

Acknowledgements

The TASA corpus was obtained courtesy of Tom Landauer of the University of Colorado and Touchstone Applied Science Associates.

References

[1] Birkhoff, Garrett: (1967) *Lattice Theory*, Am. Math. Soc., Providence RI, 3rd edition

[2] Cimiano, Phillipp; Hotho, Andreas; and Staab, Steffen: (2004) "Comparing Conceptual, Divisive, and Agglomerative Clustering for Learning Taxonomies from Text", in: *Proc. ECAI 2004*, ed. RL de Mantaras et al., pp. 435-439, IOS Press

[3] Cimiano, Phillipp; Hotho, Andreas; and Staab, Steffen: (2005) "Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis", *J. Artificial Intelligence Research*, v. **24**, pp. 305-339

[4] Euzenat, Jérôme and Shvaiko, P: (2007) *Ontology Matching*, Springer-Verlag, Hiedelberg

[5] Ganter, Bernhard; Stumme, Gerd; and Wille, Rudolf, eds.: (2005) *Formal Concept Analysis: Foundations and Applications*, Springer-Verlag

[6] Ganter, Bernhard and Wille, Rudolf: (1999) *Formal Concept Analysis*, Springer-Verlag

[7] Harris, Z: (1968) *Mathematical Structures of Language*, Wiley

[8] M Hearst: (1992) "Automatic acquisition of hyponyms from large text corpora", in: *Proc. COLING, pp. 539-545*

[9] D Hindle: (1990) "Noun Classification from Predicate-Argument Structures", in: *Proc. ACL, pp. 268-275*

[10] Kaiser, Tim; Schmidt, Stefan; and Joslyn, Cliff: (2007) "Concept Lattice Representations of Annotated Taxonomies", *Int. J. on the Foundations of Computer Science*, in press

[11] Dan Klein and Christopher D. Manning. (2003) "Accurate Unlexicalized Parsing", *Proc. 41st Annual Meeting of the Association for Computational Linguistics*, pp. 423-430

[12] Shamsfard, Mehrnoush and Barforoush, AA: (2004) "Learning Ontologies from Natural Language Texts", *Int. J. Human-Computer Studies*, v. **60**, pp. 17-63

[13] Tijerino, Yuri A; Embley, David W; Lonsdale, Deryle W; and Ding, Y et al.: (2005) "Towards Ontology Generation from Tables", *World Wide Web: Internet and Web Information Systems*, v. **8**, pp. 261-285

[14] Zeno, S., Ivens, S., Millard, R., and Duvvuri, R. (1995) *The educator's word frequency guide*, Touchstone Applied Science Associates (TASA), Inc.