Context-theoretic Semantics for Natural Language: an Algebraic Framework

Daoud Clarke

September 2007

Chapter 6

Taxonomies and Vector Lattices

A crucial feature that we require of the context-theoretic framework is that we are able to make use of logical representations of meaning within the framework. Ontologies form an important part of many systems that deal with logical representations of natural language, thus it is important to examine the relationship between ontological representations of meaning and vector based ones. In this chapter we show how an important part of an ontology, a taxonomy, can be represented in terms of vectors in a vector lattice, by means of vector lattice completions, a concept that we define. The ideas presented in this chapter marry the vector-based representations of meaning with the ontological ones by considering both from the unifying perspective of vector lattice theory.

The constructions presented here may have several practical benefits:

- The constructions provide a link with statistical representations of meaning such as latent semantic analysis (Deerwester et al., 1990) and distributional similarity measures (Lee, 1999). Eventually the ideas presented here in combination with such techniques may lead to new methods of automatic ontology construction. For example, by relating semantic analysis vectors to the taxonomy vectors it may be possible to place a new concept in the vector space of the taxonomy based on its latent semantic analysis vector.
- The vector-based representation of a taxonomy can be used to build context theories that make use of the taxonomy whilst remaining entirely vector-based, allowing the use of techniques to combine vectors such as tensor or free products, discussed in the next chapter. These could lead to new, more robust representations of natural language semantics.
- Vector spaces give us a lot of flexibility: vectors can be scaled, rotated, translated, and the dimensionality of the vector space can be reduced. These properties may lead to new techniques for the efficient representation of meaning. For example,

there may be a way to use a dimensionality reduction to efficiently represent a taxonomy in terms of vectors.

Perhaps more importantly though, the subject of this chapter is the nature of meaning itself, and the techniques we present here show that the vector lattice representations of meaning generalise ontological ones. In addition to the lattice structure of ontologies, however, vector lattices allow a more subtle description of meaning that allows the quantification of nearness of meaning that cannot be described fully in the lattice structure of ontologies. Part of the success of techniques such as latent semantic analysis is due to the ability to quantify nearness of meaning in this way; it thus seems only natural that vector lattices will eventually become the structure of choice for representing meaning in natural language.

The contributions of this chapter are as follows:

- We give a definition for a vector lattice completion as a way of representing a taxonomy in terms of a vector lattice.
- We describe several vector lattice completions with different properties:
 - A probabilistic completion (see Section 6.1.2) allows the incorporation of the "probability of a concept" into the vector-based description.
 - We describe a distance preserving completion (see Section 6.1.3) in which the distance between vectors in the vector lattice representation is the same as the distance in the ontology using a measure of Jiang and Conrath (1998).
 - The vector space representation typically uses a large number of dimensions. In Section 6.1.4 we describe a vector lattice completion that in many cases uses a smaller number of dimensions than the probabilistic completion, and discuss its application to two real world ontologies.
- The constructions we present allow the description of concepts; in Section 6.2 we discuss the representation of words which may be ambiguous, requiring the representations of the individual senses of a word to be combined.

6.1 Taxonomies

Ontologies describe relationships between concepts. They are considered to be of importance in a wide range of areas within artificial intelligence and computational linguistics. For example, WordNet (Fellbaum, 1989) is an ontology that describes relations between word senses.

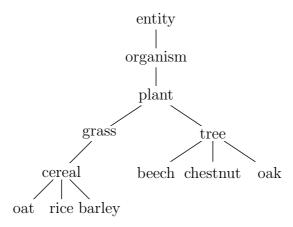


Figure 6.1: A small example taxonomy extracted from WordNet (Fellbaum, 1989).

Arguably the most important relation described in an ontology is the **is-a** relation (also called subsumption), which describes inclusion between classes of objects. When applied to meanings of words, the relation is called *hypernymy*. For example, a *tree* is a type of *plant* (the concept *plant* subsumes *tree*), thus the word "plant" is a hypernym of "tree". The converse relationship between words is called hyponymy, so "tree" is a hyponym of "plant". A system of classification that only deals with the **is-a** relation is referred to as a *taxonomy*. An example taxonomy is shown in figure 6.1, with the most general concept at the top, and the most specific concepts at the bottom.

The **is-a** relation is in general a partial ordering, since

- it is always the case that an a is an a (reflexivity);
- if an a is a b and a b is an a then a and b are the same (anti-symmetry).
- if an a is a b and a b is a c then an a is necessarily a c (transitivity).

The taxonomy described by figure 6.1 has a special property: it is a tree, i.e. no concept subsumed by one concept is subsumed by any other concept. This type of taxonomy will be studied in section 6.1.3.

6.1.1 Vector Lattice Embeddings of Taxonomies

Vector representations of meaning do not seem to sit nicely with ontological representations of meaning — the former make use of vector spaces and the latter make use of lattices. In fact, what we will show in this chapter is that the two types of representation can be combined within the structure of a vector lattice, a space that is simultaneously a vector space and a lattice. Taxonomies can be embedded in a vector lattice in such a way

that the lattice structure is preserved, and existing vector representations of meaning can be considered as implicitly carrying a lattice structure.

The relationship between concepts in a taxonomy is expressed by means of a partial order, and we wish to embed the partial ordering representation in a vector lattice; we call such an embedding a vector lattice completion. The partial ordering of the vector lattice representation must still therefore contain the partial ordering of the taxonomy, but in addition, we provide each meaning with a concrete position in some n-dimensional space. We define this formally as follows:

Definition 6.1 (Vector Lattice Completion). Let S be a partially ordered set. A vector lattice completion of S is a vector lattice V and a function ϕ from S to V that is a partial ordering homomorphism, i.e. $\phi(s_1) \leq \phi(s_2)$ if and only if $s_1 \leq s_2$, for all $s_1, s_2 \in S$.

Because the embedding will necessarily be a *lattice completion*, it will introduce new operations of meet and join on elements (see section A.3). Many taxonomies may already have some of the properties of a lattice, for example, most taxonomies are join semilattices. However the existing join operation is not usually directly useful since it does not correspond with our usual idea of logical disjunction. For example, in figure 6.1, the join of the concepts beech and oak is tree. If we say that something is a beech or an oak, it is definitely a tree, however the converse provides problems: something being a tree does not imply that that thing is necessarily a beech or an oak—since it could also be a chestnut. Thus the logical disjunction of the concepts beech and oak should sit somewhere between these two concepts and tree.

6.1.2 Probabilistic Completion

We are also concerned with the *probability* of concepts. This is an idea that has come about through the introduction of "distance measures" on taxonomies. Since words can be ascribed probabilities based on their frequencies of occurrence in corpora, the concepts they refer to can similarly be assigned probabilities. The *probability of a concept* is the probability of encountering an *instance* of that concept in the corpus, that is, the probability that a word selected at random from the corpus has a meaning that is subsumed by that particular concept.

We give a general definition based on this idea which does not require probabilities to be assigned based on corpus counts:

Definition 6.2 (Real Valued Taxonomy). A real valued taxonomy is a finite set S of concepts with a partial ordering \leq and a positive real function p over S. The measure of

a concept is then defined in terms of p as

$$\hat{p}(x) = \sum_{y \in \downarrow(x)} p(y).$$

The taxonomy is called *probabilistic* if $\sum_{x \in S} p(s) = 1$. In this case \hat{p} refers to the probability of a concept.

Thus in a probabilistic taxonomy, the function p corresponds to the probability that a term is observed whose meaning corresponds (in that context) to that concept. The function \hat{p} denotes the probability that a term is observed whose meaning in that context is subsumed by the concept.

Note that if S has a top element I then in the probabilistic case, clearly $\hat{p}(I) = 1$. In studies of distance measures on ontologies, the concepts in S often correspond to senses of words, in this case the function p represents the (normalised) probability that a given word will occur with the sense indicated by the concept. The top-most concept often exists, and may be something with the meaning "entity"—intended to include the meaning of all concepts below it.

The most simple completion we consider is into the vector lattice \mathbb{R}^S , the real vector space of dimensionality |S|, with basis elements $\{e_x : x \in S\}$.

Proposition 6.3 (Ideal Vector Completion). Let S be a probabilistic taxonomy with probability distribution function p that is non-zero everywhere on S. The function ϕ from S to \mathbb{R}^S defined by

$$\phi(x) = \sum_{y \in |x|} p(y)e_y$$

is a completion of the partial ordering of S under the vector lattice order of \mathbb{R}^S , satisfying $\|\phi(x)\|_1 = \hat{p}(x)$.

Proof. The function ϕ is clearly order-preserving: if $x \leq y$ in S then since $\downarrow(x) \subseteq \downarrow(y)$, necessarily $\phi(x) \leq \phi(y)$. Conversely, the only way that $\phi(x) \leq \phi(y)$ can be true is if $\downarrow(x) \subseteq \downarrow(y)$ since p is non-zero everywhere. If this is the case, then $x \leq y$ by the nature of the ideal completion. Thus ϕ is an order-embedding, and since \mathbb{R}^S is a complete lattice, it is also a completion. Finally, note that $\|\phi(x)\|_1 = \sum_{y \in |x|} p(y) = \hat{p}(x)$.

This close connection with the ideal completion is what leads us to call it the *ideal* vector completion. The completion allows us to represent concepts as elements within a vector lattice so that not only the partial ordering of the taxonomy is preserved, but the probability of concepts is also preserved as the size of the vector under the L^1 norm.

6.1.3 Distance Preserving Completion

Some attempts have been made to link ontological representations with statistical techniques. These centre around measures of *semantic distance* which attempt to put a value on semantic relatedness between concepts.

Jiang and Conrath (1998) defined a distance measure that has been shown to perform best out of five different measures in a spelling correction task (Budanitsky and Hirst, 2006). The measure is based on *information content* of concepts, which can be derived from their probabilities. We will show that this measure has the following property: concepts can be embedded in a vector lattice in such a way that the distance between concepts in the vector lattice is equal to the Jiang-Conrath distance measure.¹

We are able to show that the distances are preserved in certain types of taxonomy: the concepts must form a *tree*:

Definition 6.4 (Trees). A partially ordered set S is called a *tree* if every element x in S has at most one element $y \in S$ such that $x \prec y$ and there is an element I such that $z \leq I$ for all $z \in S$. The unique element preceding x is called the *parent* of x, it is denoted Par(x) if it exists.

Note that in a tree only the topmost element I has no parent.

The Jiang-Conrath measure makes use of a particular property of trees. It is easy to see that a tree forms a *semilattice*: for each pair of elements x and y there is an element $x \vee y$ that is the *least common subsumer* of x and y. For example, in figure 6.1, the least common subsumer of *oat* and *barley* is *cereal*; the least common subsumer of *oat* and *beech* is *plant*.

The measure also makes use of the information content of a concept; this is simply the negative logarithm of its probability. In our formulation, the information content IC(x) of a concept x is defined by

$$IC(x) = -\log \hat{p}(x).$$

The information content thus decreases as we move up the taxonomy; if there is a most general element I, it will have an information content of zero.

The Jiang-Conrath distance measure d(x, y) between two concepts x and y is then defined as

$$d(x,y) = IC(x) + IC(y) - 2IC(x \lor y).$$

There is a notable similarity between this expression, and a relation that holds in vector

¹Further investigation is required to determine whether the other distance measures possess this property.

lattices:

$$|u - v| = u + v - 2(u \wedge v),$$

for all u and v in the vector lattice. This formula provides the starting point for preserving distances in the vector lattice completion.

In its current form, in building a vector lattice we cannot simply replace the function \hat{p} with the information content, since \hat{p} must increase as we move up the taxonomy; instead we must invert the direction of the lattice. This allows us to embed concepts in the lattice while retaining the information content as the norm, and changes joins into meets, so that distances correspond to the Jiang-Conrath measure.

Proposition 6.5 (Distance Preserving Completion). Let S be a probabilistic taxonomy which forms a tree with partial ordering \leq . The function IC defines a positive real-valued function f_{IC} by

$$f_{IC}(x) = IC(x) - IC(Par(x)).$$

for $x \in S - \{I\}$, and $f_{IC}(I) = 0$. We define a new partial ordering \leq' on S by $x \leq' y$ iff $y \leq x$ (thus \leq' is the dual of \leq). Then f_{IC} together with the new partial ordering defines a real-valued taxonomy on S. Call the function that maps an element of S to its completion in the new taxonomy ψ . The vector lattice completion of the new taxonomy satisfies $\|\psi(x)\|_1 = IC(x)$ and $\|\psi(x) - \psi(y)\|_1 = d(x,y)$.

Proof. For the results about vector lattices used here see section A.4. By the tree nature of the taxonomy, f_{IC} is clearly a positive function satisfying $\|\psi(x)\|_1 = IC(x)$. To see the second part, we need to know that the vector lattice \mathbb{R}^S with the L_1 norm is an AL-space; this means that $\|s+t\| = \|s\| + \|t\|$ whenever $s \wedge t = 0$. We have here

$$(u - u \wedge v) \wedge (v - u \wedge v) = \frac{1}{2}(u + v - 2(u \wedge v) - |u - v|) = 0,$$

where we have used the above identity twice. Thus, using the same identity, we have

$$||u - v||_1 = ||u - v||_1 = ||u + v - 2(u \wedge v)||_1$$

$$= ||(u - u \wedge v) + (v - u \wedge v)||_1$$

$$= ||u - u \wedge v||_1 + ||v - u \wedge v||_1$$

$$= ||u||_1 + ||v||_1 - 2||u \wedge v||_1$$

For the last step, we used the fact that we are dealing with positive elements, with

 $u-u \wedge v \geq 0$ and thus, using the additive property of the L_1 norm, $||u|| = ||u-u \wedge v+u \wedge v|| = ||u-u \wedge v|| + ||u \wedge v||$.

Finally, note that the lattice completion is built from the dual of a tree, which is a join semilattice. Joins are preserved as meets in the completion since $\downarrow(x) \cap \downarrow(y) = \downarrow(x \wedge y)$, and thus we have $\psi(x) \wedge \psi(y) = \psi(x \vee y)$. This completes the proof.

Thus we have shown an important link between ontologies and vector lattice representations of meaning: it is possible to simultaneously preserve the partial ordering and distance measure of an ontology within a vector lattice representation. We believe this particular result opens up the potential for a wide range of techniques combining statistical methods of determining meaning with ontological representations. It has been suggested that distributional similarity measures can be used as a predictor of semantic similarity, originating in the distributional hypothesis of Harris (1968). There has not yet to our knowledge been a thorough analysis of the degree to which distributional similarity can be used to predict semantic distance, however our own preliminary investigations reveal that there is definitely some correlation between the two. If this correlation is strong enough then distributional similarity could in theory be used to place concepts in the vector lattice of meanings based on their similarity with surrounding concepts, opening up the field of determining meaning automatically, and making use of the fine-grained structure allowed by the vector lattice representations.

6.1.4 Efficient Completions

In this section we discuss the question of how many dimensions are necessary to maintain the lattice structure in the vector lattice completion. The representations discussed previously use a very large number of dimensions: one for each node in the ontology. To see that this is more than is generally needed, consider an ontology whose Hasse diagram is planar: that is it can be rearranged so that no lines cross (see figure 6.2). If we then position the nodes in the diagram such the lines between nodes are at an angle of less than 45° to the vertical (this can always be done by stretching the diagram out vertically), then if we rotate the diagram by 45° to the right, and set an origin, the position of each node in the two dimensional diagram can be considered as a representation of the concept in the vector lattice \mathbb{R}^2 . It is easy to see that the partial ordering is preserved — if $x \leq y$ in the partial ordering, then it will be the case in the two-dimensional vector lattice, although care has to be taken in the positioning of concepts to ensure that other unwanted relations can't be derived in the new space.

One problem with this simplistic vector lattice representation is that there is no obvious way to interpret the two dimensions. Another, more serious problem is that it is

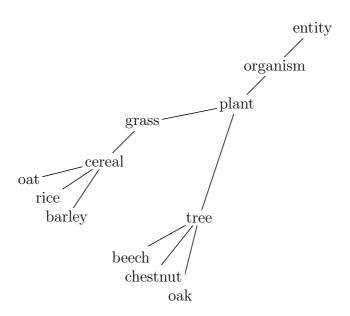


Figure 6.2: It is possible to embed a tree into a two dimensional vector lattice in such a way that the partial ordering is preserved. Two concepts s_1 and s_2 satisfy $s_1 \leq s_2$ if s_1 is to the left or level with and below or level with s_2 .

not unique: in general there are many ways we can draw the Hasse diagram, and each will correspond to a different representation. Concepts will necessarily be positioned arbitrarily according to which way the diagram is drawn, leaving us in doubt as to whether the vector aspect of the representation is meaningful. This arbitrary positioning means that distances between nodes are dependent on how we draw the Hasse diagram: in one representation a pair of nodes may be close together, while in another they may be far apart. For example, in the Hasse diagram of a tree, we can swap leaf nodes any way we wish to make a pair of nodes arbitrarily close or far apart.

We call representations that don't have this property symmetric: a representation is symmetric if the distances ||x - y|| between the representation of a pair of nodes is only dependent on the lattice properties of the nodes represented by x and y. Clearly symmetry goes along with uniqueness: if there is only one representation of a given lattice, the vector properties must be determined by the lattice.

Instead of this two dimensional representation then, we propose an efficient symmetric representation suitable for any partial ordering in which dimensions correspond to *chains* or *totally ordered subsets* of the partial order. This representation is unique up to isomorphism, given a certain requirement on the partial order which we expect to hold for most real world ontologies. This more efficient representation can generally be used in place of the vector ideal completion.

Definition 6.6 (Chains). Let S be a partially ordered set. A chain C of S is a totally

ordered subset of S, that is a subset of S which is a partially ordered set under the the partial ordering of S such that $x \leq y$ or $y \leq x$ for all $x, y \in C$.

A collection of chains is called *covering* if $\bigcup \mathcal{C} = S$. Clearly every partially ordered set has at least one covering collection of chains: that collection consisting of all chains containing just one node of S.

Proposition 6.7 (Chain completion). Let S be a real valued taxonomy and $C = \{C_1, C_2, \ldots C_n\}$ be a covering collection of chains for S. Let $Ch_{\mathcal{C}}(x) = \{i : x \in C_i\}$. Then define the function ξ_0 from S to \mathbb{R}^n by

$$\xi_0(x) = \sum_{i \in \operatorname{Ch}_{\mathcal{C}}(x)} \frac{p(x)}{|\operatorname{Ch}_{\mathcal{C}}(x)|} e_i,$$

where e_i are the basis elements of \mathbb{R}^n . Then the chain completion ξ is defined by:

$$\xi(x) = \sum_{y \le x} \xi_0(y).$$

The function ξ defines a vector lattice completion of S satisfying $\|\xi(x)\|_1 = \hat{p}(x)$.

Proof. By the definition of ξ it is clear that $u \leq v$ in S implies $\xi(u) \leq \xi(v)$ in \mathbb{R}^n . Conversely, if it is not true that $u \leq v$ then there will be some chain in \mathcal{C} containing v but not u, so it will never be true that $\xi(u) \leq \xi(v)$, showing that ξ defines an embedding of the partial ordering of S, and since \mathbb{R}^n is a vector lattice, it also defines a vector lattice completion. Finally, note that

$$\|\xi(x)\|_1 = \sum_{y \le x} \|\xi_0(x)\|_1 = \sum_{y \le x} p(x) = \hat{p}(x)$$

since all the vectors are positive, which completes the proof.

Providing we can find a covering with a low number of chains n, the previous proposition gives us an efficient vector lattice representation using n dimensions. The representation as it stands is not unique, since there are in general many ways we can cover a partially ordered set with chains. The task then, is to find an efficient, unique way of determining a covering collection of chains. We achieve this by considering $maximal\ chains$, chains containing all elements of S that they can whilst remaining a totally ordered set.:

Definition 6.8 (Maximal chains). A maximal chain C for S is a chain such that there is no element x of S-C such that if x were added to C then C would remain a chain. Let C be the covering collection of chains consisting of all maximal chains of S. S is said

to be uniquely minimally covered by C if for each $C \in C$ there is at least one element $x \in C$ such that x is not in any other chain of C; in this case, S is said to possess a unique minimal covering C.

Proposition 6.9. If S has a unique minimal covering C, then C is the unique covering for S with the least number of chains.

Proof. By definition \mathcal{C} is unique, since it consists of all maximal chains. To see that \mathcal{C} has the least possible number of chains, note that each chain contains as many elements of S as possible, while removing any of these chains and maintaining a covering collection is impossible, since each chain contains an element not in any other chain.

Thus if S has a unique minimal covering, we can represent it uniquely and efficiently using the number of dimensions corresponding to the number of chains in this covering. Any taxonomy that is a tree has a unique minimal covering: each maximal chain will have a leaf node that is not in any chain; in fact there will be a chain corresponding to each leaf node. Thus the chain completion gives a unique efficient representation for any taxonomy that is a tree, and we would expect taxonomies that are very tree-like to also have efficient representations.

6.1.5 Analysis of Application to Ontologies

While we know that the chain completion is a relatively efficient for trees, we don't know how useful it is likely to be in real-world applications. To find out, we analysed two real world ontologies. The first is the Semantic Network used in the Unified Medical Language System (National Library of Medicine, 1998), whose taxonomy consists of just 135 nodes representing broad categories of meanings related to medical concepts. In this case, the taxonomy has a simple tree structure, so each dimension corresponds to a leaf node. There are 90 leaf nodes, thus we can represent the 135 nodes using only 90 dimensions, a saving of a third.

It is also instructive here to consider a simple theoretical situation: a regular tree of depth n with each node having r branches. In this case, the total number of nodes is

$$\sum_{i=1\dots n} r^i = \frac{r^{n+1} - 1}{r-1} - r \simeq \frac{r^{n+1}}{r-1}$$

where the approximation is for large n and r > 1. The number of leaf nodes is r^n , thus in this approximation the ratio of leaf nodes to the total number of nodes will be $r^n(r-1)/r^{n+1} = (r-1)/r$. Thus the saving in the chain completion is greatest for low

r: in a binary tree, half the nodes will be concentrated in the leaf nodes. The semantic network we considered above has a saving corresponding to r = 3.

The second taxonomy we considered was that of WordNet (Fellbaum, 1989). This is a very different situation to that just considered, having a much greater number of nodes, and no tree structure — quite a large number of nodes have more than one parent. We looked at a subset of around 43,000 nodes using the hypernymy relation of nouns only; each node corresponds to a "synset" or concept corresponding to senses of words in WordNet. We found a covering collection of chains using around 35,000 chains: a saving in terms of dimensionality of around 20%. This does not give a unique representation however, and thus potentially suffers from some of the same problems as the two dimensional representations. The total number of maximal chains was around 60,000, meaning the unique chain-based representation would be less efficient than the straightforward vector lattice completion in which each dimension corresponds to a node.

It seems that chain-based representations are able to provide modest improvements in the efficiency of vector lattice representations, especially in the case of taxonomies with a tree structure. It is our hope, however, that techniques such as dimensionality reduction will eventually provide a means to find much more efficient representations, using good quality approximations which retain as much structure as possible of the original vector lattice.

6.1.6 Context-theoretic Taxonomies

The unifying mathematics of vector lattices allows us to view ontological representations and vector based representations based on the analysis of contexts from the same perspective. Just as we have endowed the partial ordering of taxonomies with a vector structure, we can view the vectors of context-based methods as having a lattice structure. We can view such a structure as a "context-theoretic taxonomy". It satisfies the mathematical requirements of a taxonomy, yet it is not considered as representing concepts that are necessarily related to the real world: it merely represents contexts that terms appear in. Thus the representations discussed in the first chapter — the vectors of latent semantic analysis and the feature vectors of distributional similarity — can be considered as describing context-theoretic taxonomies.

It is this unifying perspective that we hope will lead the way to new methods combining ontological (model-theoretic) and context-theoretic representations and techniques.

6.2 Representing Words

So far we have only really considered representing concepts, or *senses* of words; we have not been concerned yet with how to represent words themselves, which may be ambiguous with meanings covering many senses. For example, we view the structure of WordNet, which describes senses of words, as a partial ordering, or as elements of a vector lattice. If we want to combine the vector lattice representations of the senses of a word to form something representing the ambiguous meaning, what is the correct way to do this?

Context-theoretic techniques provide an answer: if we look at the most straightforward model of context, the representation of a word is given by the vector sum of the representations of its contexts. This can easily be seen by considering the model of context discussed in the first chapter: if we add sense tags to words in a corpus, then look at the vector representations of the individual senses, since the vector representation is formed linearly, summing these representations will give us the same vector as that arrived at by looking at occurrences of the word without sense tags. This also makes sense from a probabilistic perspective; the probability of the occurrence of a word in a corpus is the sum of the probability of the occurrences of its senses, and this property is carried over in the L^1 norm of the corresponding vector representations. Looking at the lattice structure, this construction behaves as we would expect: each sense of a word entails the word itself. Thus if word w has n senses $s_1, s_2, \ldots s_n \in S$, then the context vector of w would be

$$\hat{w} = \sum_{i=1}^{n} \hat{s}_i$$

where \hat{s}_i is the context vector of sense s_i .

When it comes to making use of vector representations of taxonomies, however, we run into a problem. We have constructed our vectors so that the L^1 norm corresponds to the probability of the concept, which depends on the taxonomic structure. According to the context-theoretic philosophy, the representation of a word should be constructed linearly from the representations of its senses, however the probability of the occurrence of a sense does not coincide with the probability of a concept. For example, the meaning of the word "entity" corresponds to the most general concept in some taxonomies, and thus the probability of the concept entity is 1. However the word itself occurs fairly rarely in corpora, and we would expect it to have a fairly low probability even with respect to words representing much more general concepts.

Looking at the situation from a context-theoretic perspective helps us to find an answer. We can view each node in the taxonomy as a context that words can occur in. In the ideal vector completion a concept s is represented as a sum over basis vectors

corresponding to the nodes representing concepts at least as general as s. When s is the sense of a word, we view the word as occurring in sense s in contexts corresponding to the concepts at least as general as s. We may know the probability of the sense, but we have no way to distribute this probability over the hypothetical contexts.

One way of getting around this problem is to renormalise the vectors representing the individual senses s_i and scale them according to the probability π_i that the word w occurs in sense s_i (so that $\sum_i \pi_i$ is equal to the probability of word w occurring):

$$\bar{w} = \sum_{i=1}^{n} \frac{\pi_i}{\|\bar{s}_i\|_1} \bar{s}_i$$

Thus we have a plausible way of representing words in terms of vectors. If we are to make use of these representations as part of a context theory, however, we have to be able to consider them as elements of an algebra. We have already seen the use of projections to represent lattice structures in the previous chapter, and again it is an algebra formed from projections that we will use to represent meanings of words within the setting of a context theory. In fact, as we will show, work in measures of distributional similarity supports the idea of representing meanings as projections.

6.2.1 Distributional Similarity and Projections

The work of Lee (1999) analyses distributional similarity measures with respect to the support of the underlying distribution. Let $f_t(c)$ denote the observed frequency of term t occurring in context c. The support S(t) of f_t is the set of contexts c for which $f_t(c)$ is non-zero;

$$S(t) = \{c \in C : f_t(c) > 0\}$$

where C denotes the set of possible contexts that terms may occur in, or the feature space. According to our previous analysis, we consider the function f_t as a vector in the space \mathbb{R}^C .

Lee considers measures of the degree of similarity between two terms u and v. She shows that the three best performing measures (which include the L^1 norm, $||f_u - f_v||_1$) all depend only on the behaviour of the functions f_u and f_v on the intersection of the supports of the two terms, $S(u,v) = S(u) \cap S(v)$. Those measures which placed emphasis on the behaviour of the functions outside of this set, such as the L^2 norm, generally performed poorly in comparison.

Weeds (2003) takes this analysis further, considering different functions D(t, c) measuring the degree of association between a term t and context c. The support with respect

to D is defined as $S_D(t) = \{c \in C : D(t,c) > 0\}$. She then considers the precision according to an "additive model" defined in terms of D:

$$\mathcal{P}^{\text{add}}(u,v) = \frac{\sum_{c \in S_D(u,v)} D(u,c)}{\sum_{c \in S_D(u)} D(u,c)};$$

recall can then be defined as the dual of precision, $\mathcal{R}^{\text{add}}(u,v) = \mathcal{P}^{\text{add}}(v,u)$. Weeds goes on to show how a general framework to describe distributional similarity measures can be described in terms of measures of precision and recall, and evaluates a range of measures within her framework. The best performing measure made use of the additive model of precision and recall together with a mutual information based function for D.

The details of Weeds' analysis are not so relevant for us; what is important to note is that in Weeds' additive model there is a move away from considering terms merely as vectors, and that this move is experimentally successful. What we will show is that we can view the additive model as representing terms as *projections*, special kinds of operators on a vector space.

The vector space we are considering is given by the set C of contexts that terms may occur in; we denote it \mathbb{R}^C since each element c of C has a corresponding basis element $e_c \in \mathbb{R}^C$ which determines a dimension in the vector space. Given a subset of contexts X, $X \subseteq C$, we can view the vector space \mathbb{R}^X as a subspace of \mathbb{R}^C . This subspace defines a projection P_X on \mathbb{R}^C .

To specify this in more detail, consider a vector f defined on \mathbb{R}^C in terms of its components α_c , where $f = \sum_{c \in C} \alpha_c e_c$. The effect of the projection P_X is then defined as follows:

$$P_X f = \sum_{c \in X} \alpha_c e_c.$$

Given two subsets X and Y of C, it is easy to see that $P_X P_Y = P_{X \cap Y}$, thus the projection encodes set-theoretic behaviour. Since the definitions of precision and recall depend on the intersection of supports, we can translate these definitions into ones based on projections:

$$\mathcal{P}^{\mathrm{add}}(u,v) = \|P_u P_v \Omega_D(u)\|_1,$$

where $P_t = P_{S_D(t)}$ and $\Omega_D(u)$ is a vector in \mathbb{R}^C given in terms of its components by

$$\Omega_D(u) = \frac{1}{\sum_{c \in C} D(u, c)} \sum_{c \in C} D(u, c) e_c.$$

This representation comes close to providing us with a context theory; words can be represented as operators on a vector lattice and thus are elements of an algebra; the

difference is that there is not a unique linear functional under consideration, the linear functional (which depends on $\Omega_D(u)$) is different depending on what element we are considering precision with respect to. The preceding analysis does however, point to the representation of meanings as projections on a vector lattice; we will show how such representations allow us to combine representations of concepts to form representations of the meanings of words.

6.2.2 Combining Concept Projections

First we show how concepts in a taxonomy can be represented in terms of projections together with a linear functional.

Definition 6.10 (Ideal Projection Completion). If S is a probabilistic taxonomy with probability distribution function $p \in \mathbb{R}^S$, then the *ideal projection* P_x associated with $x \in S$ is the projection $P_{\downarrow(x)}$ on the space \mathbb{R}^S . We define a linear functional ψ on $L^1(\mathbb{R}^S)$ by

$$\psi(A) = \|(Ap)^+\|_1 - \|(Ap)^-\|_1,$$

Proposition 6.11. The ideal projection completion defines a vector lattice completion for S, such that $\psi(P_x) = \hat{p}(x)$.

Proof. There is clearly a lattice isomorphism between the ideal completion representation $\downarrow(x)$ of $x \in S$ and the projection P_x ; for example

$$P_x P_y = P_{\downarrow(x) \cap \downarrow(y)}.$$

Then note that
$$\psi(P_x) = ||P_x p||_1 = \sum_{y \in \downarrow(x)} p(y) = \hat{p}(x)$$
.

The new representation encodes probabilities in the linear functional rather than directly in the representation of individual concepts, in contrast to the ideal vector completion introduced earlier. This gives us additional flexibility to combine concept representations in a way which preserves the partial ordering relation as we would expect from a context-theoretic perspective.

The ideal projection completion can in fact be used to define a context theory for an alphabet A if we have a way of associating elements of A with concepts in S; for example A may be a set of words and S a taxonomy of their meanings. If the words are unambiguous they will be associated with just one concept in S. Thus we can associate with each word a projection on $L^1(\mathbb{R}^S)$.

The new flexibility comes in being able to add these projections to create representations of words. If a word w has n senses $s_1, s_2, \ldots s_n \in S$, and the word w occurs in

the sense s_i with probability π_i , then we can represent w as a probabilistic sum of the projection representation of its senses:

$$\bar{w} = \sum_{i=1}^{n} \frac{\pi_i}{\psi(P_{s_i})} P_{s_i},$$

where \bar{w} is the representation of w in $L^1(\mathbb{R}^S)$. The factor $\pi_i/\psi(P_{s_i})$ ensures that $\psi(\bar{w})$ is equal to the probability of word w; it can be interpreted as the conditional probability that word w occurs in sense s_i given that some word has occurred in some sense t at least as general as s_i , that is $s_i \leq t$.

Because we represent words as operators, in addition to the usual lattice operations, which work in a similar way to the ideal vector completion, multiplication is also defined on the representations. We can think of the probabilistic sum of senses as representing our uncertainty about the meaning of a word. The product of two words then, would represent our uncertainty about the conjunction of their meanings. For example, if we approximate the meaning² of the word *line* by

$$\bar{w}_l = \frac{3}{10}P_{l_1} + \frac{1}{10}P_{l_2}$$

where l_1 represents the sense "a formation of people or things one beside another" and l_2 represents the sense "a mark that is long relative to its width", and the word mark by

$$\bar{w}_m = \frac{1}{5}P_{m_1} + \frac{1}{10}P_{m_2}$$

where m_1 represents the sense "grade or score" and m_2 represents the sense "a visible indication made on a surface", then the product is given by

$$\hat{w}_l \hat{w}_m = \frac{3}{50} P_{l_1} P_{m_1} + \frac{3}{100} P_{l_1} P_{m_2} + \frac{1}{50} P_{l_2} P_{m_1} + \frac{1}{100} P_{l_2} P_{m_2}.$$

If we further assume that the meanings of senses are disjoint, except for those referring to the sense "a mark that is long relative to its width" and the sense "a visible indication made on a surface"; that is we assume $P_x P_y = 0$ unless $x = l_2$ and $y = m_2$ or vice versa, in which case $P_{l_2} P_{m_2} = P_{l_2}$ since a line is a type of mark. Then $\hat{w}_l \hat{w}_m = \frac{1}{100} P_{l_2}$; the product has disambiguated the meaning of both words.

²Meanings are based on Wordnet definitions (Fellbaum, 1989), probabilities are invented.

Bibliography

- Y. A. Abramovich and Charalambos D. Aliprantis. *An Invitation to Operator Theory*. American Mathematical Society, 2002.
- V. M. Abrusci. Phase semantics and sequent calculus for pure noncommutative classical linear propositional logic. The Journal of Symbolic Logic, 56:1403–1451, 1991.
- Elena Akhmatova. Textual entailment resolution via atomic propositions. In Dagan et al. (2005b).
- Charalambos D. Aliprantis and Owen Burkinshaw. *Positive Operators*. Academic Press, 1985.
- Roy Bar-Haim, Ido Dagan, Bill Dolan, Lisa Ferro, Danilo Giampiccolo, Bernardo Magnini, and Idan Szpektor. The second pascal recognising textual entailment challenge. In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, 2006.
- Yehoshua Bar-Hillel. On syntactical categories. *Journal of Symbolic Logic*, 15(1):1–16, 1950.
- Yehoshua Bar-Hillel. Language and Information: Selected Essays on their Theory and Application. Addison-Wesley Publishing Co., Reading, MA., 1964.
- Samuel Bayer, John Burger, Lisa Ferro, John Henderson, and Alexander Yeh. MITRE's submissions to the EU pascal RTE challenge. In Dagan et al. (2005b).
- Garrett Birkhoff. *Lattice Theory*. Amer. Math. Soc. Colloquium Publications, New York, 1973.
- Patrick Blackburn and Johan Bos. Representation and Inference for Natural Language. CSLI, 2005.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

- Johan Bos. Towards wide-coverage semantic interpretation. In *Proceedings of Sixth International Workshop on Computational Semantics IWCS-6*, page 42?53, 2005.
- Johan Bos and Katja Markert. When logical inference helps determining textual entailment (and when it doesnt). In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, 2006.
- Jordan Boyd-Graber, David Blei, and Xiaojin Zhu. A topic model for word sense disambiguation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1024–1033, 2007. URL http://www.aclweb.org/anthology/D/D07/D07-1109.
- E. Briscoe, John Carroll, and R. Watson. The second release of the rasp system. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, Sydney, Australia*, 2006.
- Alexander Budanitsky and Graeme Hirst. Evaluating wordnet-based measures of semantic distance. Computational Linguistics, 32(1):13–47, 2006.
- Junfu Cai, Wee Sun Lee, and Yee Whye Teh. Improving word sense disambiguation using topic features. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1015–1023, 2007. URL http://www.aclweb.org/anthology/D/D07/D07-1108.
- Claudia Casadio and Joachim Lambek. A tale of four grammars. *Studia Logica*, 71(3): 315–329, 2002.
- Daoud Clarke. Meaning as context and subsequence analysis for textual entailment. In *Proceedings of the Second PASCAL Recognising Textual Entailment Challenge*, 2006.
- James R. Curran and Marc Moens. Improvements in automatic thesaurus extraction. In *ACL-SIGLEX Workshop on Unsupervised Lexical Acquisition*, Philadelphia, 2002.
- Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*, 2005a.
- Ido Dagan, Oren Glickman, and Bernardo Magnini, editors. *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*, 2005b.

- Ido Dagan, Lillian Lee, and Fernando C. N. Pereira. Similarity-based methods for word sense disambiguation. In *ACL*, pages 56–63, 1997.
- Ido Dagan, Fernando Pereira, and Lillian Lee. Similarity-based estimation of word cooccurrence probabilities. In 32nd Annual Meeting of the ACL, pages 272–278, 1994.
- Scott Deerwester, Susan Dumais, George Furnas, Thomas Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.
- Rodolfo Delmonte, Sara Tonelli, Marco Aldo Piccolino Boniforti, Antonella Bristot, and Emanuele Pianta. VENSES a linguistically-based system for semantic evaluation. In Dagan et al. (2005b).
- Christaine Fellbaum, editor. WordNet: An Electronic Lexical Database. The MIT Press, Cambridge, Massachusetts, 1989.
- John R. Firth. Modes of meaning. In *Papers in Linguistics* 1934–1951. Oxford University Press, London, 1957a.
- John R. Firth. A synopsis of linguistic theory 1930–1955. In F. Palmer, editor, Selected Papers of J. R. Firth. Longman, London, 1957b.
- Abraham Fowler, Bob Hauser, Daniel Hodges, Ian Niles, Adrian Novischi, and Jens Stephan. Applying COGEX to recognize textual entailment. In Dagan et al. (2005b).
- Maayan Geffet and Ido Dagan. The distributional inclusion hypotheses and lexical entailment. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, University of Michigan, 2005. URL http://www.aclweb.org/anthology/P05-1014.
- J. Y. Girard. Linear logic. Theoretical Computer Science, 50:1–102, 1987.
- Oren Glickman and Ido Dagan. A probabilistic setting and lexical cooccurrence model for textual entailment. In ACL-05 Workshop on Empirical Modeling of Semantic Equivalence and Entailment, 2005.
- Gregory Grefenstette. Explorations in automatic thesaurus discovery. Kluwer Academic Publishers, Dordrecht, NL, 1994.
- Petr Hájek. Basic fuzzy logic and BL-algebras. Soft Computing—A Fusion of Foundations, Methodologies and Applications, 2(3):124–128, September 1998.

- Zellig Harris. Mathematical Structures of Language. Wiley, New York, 1968.
- Zellig Harris. Distributional structure. In Jerrold J. Katz, editor, *The Philosophy of Linguistics*, pages 26–47. Oxford University Press, 1985.
- Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. Recognizing textual entailment with lcc's groundhog system. In *Proceedings of the Second PASCAL Challenges Workshop*, 2006.
- Donald Hindle. Noun classification from predicate-argument structures. In *ACL*, pages 268–275, 1990.
- P. Hinman. Fundamentals of Mathematical Logic. A. K. Peters, 2005.
- Thomas Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the 15th Conference on Uncertainty in AI*, pages 289–296. Morgan Kaufmann, 1999.
- Patrick Honeybone. J. R. Firth. In S. Chapman and C. Routledge, editors, *Key Thinkers in Linguistics and the Philosophy of Language*. Edinburgh University Press, 2005.
- J. M. Howie. An Introduction to Semigroup Theory. Academic Press, London, 1976. ISBN 0-12-356950-8.
- Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the International Conference on Research in Computational Linguistics*, 1998.
- Hans Kamp and Uwe Reyle. From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory, volume 42 of Studies in linguistics and philosophy. Kluwer, Dordrecht, 1993.
- Aaron Nathan Kaplan. A computational model of belief. PhD thesis, University of Rochester. Dept. of Computer Science, 2000.
- Adam Kilgarriff. Thesauruses for natural language processing. In *Proceedings of the Joint Conference on Natural Language Processing and Knowledge Engineering*, pages 5–13, Beijing, China, 2003.
- Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *Proceedings* of the 41st Annual Meeting of the Association for Computational Linguistics, 2003.
- Erwin Kreyszig. Introduction to Functional Analysis with Applications. Robert E. Krieger Publishing Company, Malabar, Florida, 1989.

- S. Kundu and J. Chen. Fuzzy logic or Lukasiewicz logic: a clarification. In Zbigniew W. Raś and Maria Zemankova, editors, *Proceedings of the 8th International Symposium on Methodologies for Intelligent Systems*, volume 869 of *LNAI*, pages 56–64, Berlin, October 1994. Springer. ISBN 3-540-58495-1.
- John Lafferty, Daniel Sleator, and Davy Temperley. Grammatical trigrams: A probabilistic model of LINK grammar. In *Proc. of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language. Cambridge, MA, 1992*, pages 89–97, Menlo Park, CA, 1992. AAAI Press.
- J. Lambek. The mathematics of sentence structure. American Mathematical Monthly, 65: 154–169, 1958.
- J. Lambek. From categorial grammar to bilinear logic. In Kosta Došen and Peter Schroeder-Heister, editors, Substructural Logics, pages 207–238. Oxford Univ. Press, 1993.
- Joachim Lambek. Type grammars as pregroups. Grammars, 4(1):21–39, 2001.
- Lillian Lee. Measures of distributional similarity. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-1999)*, pages 23–32, 1999.
- Dekang Lin. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, pages 768–774, Montreal, 1998.
- Dekang Lin and Patrick Pantel. DIRT-discovery of inference rules from text. In Foster Provost and Ramakrishnan Srikant, editors, *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-01)*, pages 323–328, New York, August 26–29 2001. ACM Press.
- Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins. Text classification using string kernels. *Journal of Machine Learning Research*, 2:419–444, 2002.
- Christopher Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA, 1999.

- Diana Mccarthy, Bill Keller, and John Carroll. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 2003.
- Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll. Finding predominant word senses in untagged text. In ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, page 279, Morristown, NJ, USA, 2004. Association for Computational Linguistics.
- W. D. Munn. Free inverse semigroup. Proceedings of the London Mathematical Society, 29:385–404, 1974.
- National Library of Medicine. UMLS Knowledge Sources. National Library of Medicine, U.S. Dept. of Health and Human Services, 8th edition, 1998.
- N. Nilsson. Probabilistic logic. ai, 28:71–87, 1986.
- Christos H. Papadimitriou, Prabhakar Raghavan, Hisao Tamaki, and Santosh Vempala. Latent semantic indexing: A probabilistic analysis. In Laura Haas and Ashutosh Tiwary, editors, *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, volume 27, pages 159–168, Seattle, Washington, 1998. ACM Press.
- Mati Pentus. Models for the lambek calculus. Annals of Pure and Applied Logic, 75: 179–213, 1995.
- Fernando Pereira, Naftali Tishby, and Lillian Lee. Distributional clustering of English words. In *Proceedings of ACL-93*, 1993.
- Rajat Raina, Aria Haghighi, Christopher Cox, Jenny Finkel, Jeff Michels, Kristina Toutanova, Bill MacCartney, Marie-Catherine de Marneffe, Christopher D. Manning, and Andrew Y. Ng. Robust textual inference using diverse knowledge sources. In Dagan et al. (2005b).
- Magnus Sahlgren and Jussi Karlgren. Vector-based semantic analysis using random indexing for cross-lingual query expansion. *Lecture Notes in Computer Science*, 2406: 169–176, 2002.
- Daniel D. Sleator and Davy Temperley. Parsing english with a link grammar. Technical Report CMU-CS-91-196, Department of Computer Science, Carnegie Mellon University, 1991.

- Daniel D. Sleator and Davy Temperley. Parsing english with a link grammar. In *The Third International Workshop on Parsing Technologies*, August 1993.
- Marta Tatu and Dan I. Moldovan. A logic-based semantic approach to recognizing textual entailment. In *ACL*. The Association for Computer Linguistics, 2006.
- Dan-Virgil Voiculescu. Free Probability Theory. American Mathematical Society, 1997.
- Julie Weeds. Measures and Applications of Lexical Distributional Similarity. PhD thesis, Department of Informatics, University of Sussex, 2003.
- Julie Weeds, David Weir, and Diana McCarthy. Characterising measures of lexical distributional similarity. In *Proceedings of the 20th International Conference of Computational Linguistics*, COLING-2004, Geneva, Switzerland., 2004.
- Dominic Widdows. Orthogonal negation in vector spaces for modelling word-meanings and document retrieval. In *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics*, 2003, Sapporo, Japan, pages 136–143, 2003.
- Dominic Widdows. *Geometry and Meaning*. Center for the Study of Language and Information, Stanford, 2004.
- Ludwig Wittgenstein. *Philosophical Investigations*. Macmillan, New York, 1953. G. Anscombe, translator.
- Mary McGee Wood. Categorial Grammars. Routledge, London, 1993.
- F. M. Zanzotto, A.AMoschitti, M. Pennacchiotti, and M. T. Pazienza. Learning textual entailment from examples. In *Proceedings of the Second PASCAL Challenges Workshop*, 2006.

Index

Abrusci, V. M., 98	context vector, 29
abstract Lebesgue space, 33, 43, 45, 131	context-theoretic
algebra, 70	framework, $7-10$, 42 , 44 , 78
formed from semigroup, 132	probability, 27, 34 , 43
lattice-ordered, 42, 44	taxonomies, 89
over a field, 27, 37–40, 43, 131	corpus model, 28
ambiguity, 60, 79	general, 70
lexical, 72–73	semantic, 71–72
anaphora resolution, 75	creation and annihilation operators, 102–
Banach space, 42, 125	105
Bar-Hillel categorial grammar, 95	data sparseness, 29
basis, 30, 39, 124	dimensionality reduction, 107, 119
orthonormal, 124	direct sum, 125
Bayesianism, 65–67, 77	distance measures
bilinear logic, 98–99	on taxonomies, 81
birooted word-trees, 110	distributional
Boolean algebra, 63, 74, 129	generality, 15
canonical form, 76	hypothesis, 6, 14–15
bracket semigroup, 109, 113	inclusion hypotheses, 16
ectomorial gramman OF 100	similarity, 16, 23–26
categorial grammar, 95–100	distributional similarity, 78 , $91-92$
and context theories, 99–100	distributivity, 40
and link grammar, 115	dot product, 124
chain, 86	enteilment 97
classification task, 55	entailment, 27
completeness, 42, 125	and lattice structure, 32
congruence, 121	degree of, 8, 37, 45 , 47, 76, 100
context algebra, 39	lexical, 16
context theory, 44 , 56, 58, 59, 67, 119	textual, 7, 47
strong, 44	PASCAL Challenge, 7, 47–56

field, 120	Lafferty et al., 106
filter, 129	Lambek calculus, 96–98
Firth, J. R., 6, 13–14	Lambek, Joachim, 96, 98
Fock space, 102, 103 , 108	language models, 26
free probability, 103, 116 , 119	latent Dirichlet allocation, 22–23, 29
free semigroup, 99, 120	latent semantic analysis, 16–20, 78, 102
and Lambek calculus, 97	probabilistic, 20–22
fuzzy logic, 74	lattice, 27, 127
	and logic, 63
Girard, J. Y., 98	distributive, 128
Glickman and Dagan	modular, 128
lexical entailment model, 51, 57–58	lattice homomorphism, 129
textual entailment framework, 51, 56	lattice homomorphism (operator), 133
group, 120	lattice ordered algebra, 99
Harris, Zellig, 6, 14–15	lexical overlap, 50, 59
Hasse diagram, 127	limit, 124
planar, 86	Lin, Dekang, 26
Hilbert space, 102–105, 125	link grammar, 95, 100–116
hypernymy, 16, 80	stochastic, 106
nypernymy, 10, 00	logical semantics, 6, 8, 60–77, 118
ideal (lattice), 129	and textual entailment, 53–56
completion, 130	l^p norm, 123
ideal completion, 62	L^p space, 125
ideal vector completion, 82	l^p space, 125 , 130
idempotents, 110	Lukasiewicz logic, 74
information content, 83	Datasatowicz logic, 11
inner product, 123	matrices, 102, 117
inverse semigroup	and link grammar, $106-107$
free, 113	meaning
inverse semigroups, 109	and lattice structure, 27, 30
free, 110	as context, 6, 13, 27, 29
I 12	as use, 6, 13
Jaccard's coefficient, 26	metric, 122
Jiang and Conrath, 83–84	Munn, W. D., 110
Kullback-Leibler divergence, 24	non-commutative probability, 41, 131
L^1 norm, 24, 33 L^{∞} , 29	ontologies, 78–80

operator (linear), 132 positive, 133 order embedding, 129 parsers, 60, 64, 74, 102, 116 and operators, 108 partial ordering, 27, 81, **126**, 129 parts of speech, disambiguating, 73 philosophy of meaning, 12–16 positive vector, 130 pregroup, **98**, 99 principle ideal, 129 probability of a string, 71 propositional calculus, 63–64, 74 quantum mechanics, 102 random projections, 107 residuated lattice, 97, 99, 100 retrieval, 102 scoring, 55 semantic disambiguation, 73 Semantic Network, 88 semigroup, 120 lattice ordered, 97 partially ordered, 97 semilattice, 127 Sleator and Temperly, 100 statistical disambiguation, 73 sub-vector lattice, 44 subsequence matching, 58 taxonomy, 78-94 probabilistic, 82 real valued, 81 tensor product, 126 tree, 80, 83, 88

Munn, 110, 113, 116 regular, 88 trigrams, 106 uncertainty, representing, 8, 60, 64–73 Unified Medical Language System, 88 vector lattice, 27, 30, 81, **130** vector lattice completion, 78, 79, 81 chain completion, 87 distance preserving, 79, 83–85 efficient, 85–89 ideal projection completion, 92 probabilistic, 79, 81–82 vector space, 121 finite dimensional, 122 partially ordered, 130 Weeds, Julie, 91 Wittgenstein, Ludwig, 6, 12–13

word sense disambiguation, 60, 64, 73, 75

WordNet, 79, 88, 89