A Mathematical Model for Context and Word-Meaning

Appeared in *Fourth International and Interdisciplinary Conference on Modeling and Using Context*, Stanford, California, June 2003, pages 369–382

Dominic Widdows
Center for the Study of Language and Information,
Stanford University
dwiddows@csli.stanford.edu *

February 10, 2008

Abstract

Context is vital for deciding which of the possible senses of a word is being used in a particular situation, a task known as disambiguation. Motivated by a survey of disambiguation techniques in natural language processing, this paper presents a mathematical model describing the relationship between words, meanings and contexts, giving examples of how context-groups can be used to distinguish different senses of ambiguous words. Many aspects of this model have interesting similarities with quantum theory.

1 Introduction

Context plays a key role in determining the meaning of words — in some contexts the word *suit* will refer to an item of clothing, in others a legal action, and so on. In the past decade, the challenge of incorporating contextual models into the way information is described has become very immediate and practical, in the wake of rapid technological advances. To compare and combine information which is readily available but varies across languages, domains of expertise and media, it is important to have some way of expressing what that information actually means in a common and flexible framework. Context can be very useful here — if someone is trying to buy a new computer they will be much more interested in the term *PC* if it occurs in a magazine called *'Computing Today'* than if it occurs in the *'Political Activism Quarterly'*. A word in one language can often have several possible translations in another language depending on

*This research was supported in part by the Research Collaboration between the NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University, and by EC/NSF grant IST-1999-11438 for the MUCHMORE project.

which meaning is appropriate (for example, English *drugs* can translate both to *drogen*='narcotics' and to *medikamente*='medicines' in German), and the correct translation can only be determined using context.

However, it is much easier to give examples of what context is used for and why it is important than it is to give a proper account what context is and how it is used to determine meaning. This paper attempts to bring some light on these issues, by describing techniques for resolving ambiguity in natural language processing within a particular mathematical framework. This investigation leads to a concrete definition of context, and to promising similarities with traditional branches of mathematics and physics which are much more precisely understood than our current knowledge of the way context is used to determine meaning.

This paper proceeds as follows. Section 2 introduces the field of Word-Sense Disambiguation (WSD), the branch of natural language processing which is concerned with finding the correct meaning for a word used in a particular situation. Section 3 presents a mathematical model which can be used to define and describe WSD and other linguistic phenomena within a unified framework based upon the three spaces $\mathcal{W}$ ('words'), $\mathcal{L}$ (a lexicon of 'meanings'), $\mathcal{C}$ ('contexts') and mappings between them, in particular 'sense-mappings' which map a word w in a context c to a particular lexical meaning l. Section 4 examines the space $\mathcal{C}$ in the light of these mappings, which suggests an approach to defining context itself. More exactly, we define groups of contexts, two linguistic situations being in the same 'context group' with respect to a particular word if that word gets mapped to the same meaning in both contexts. Examples are given of how homonyms (unrelated senses of ambiguous words) give rise to disjoint context groups, and systematically related senses give rise to overlapping context groups. Finally, Section 5 describes some notable if surprising similarities between our model for disambiguation and the process of making an observation in quantum theory. It is hoped that these analogies will be of interest to many researchers, scientists and even interested layfolk.

Context is of importance in linguistics, philosophy, sociology and many other disciplines, as well as to natural language processing. The approach taken in this paper is to cast the NLP problem in a mathematical setting, and through this to enable the use of ideas from this very traditional discipline. Many other more established areas of scholarship, way outside this author's expertise, could almost certainly contribute fruitful ideas to NLP if researchers become more aware of the problems and possibilities therein. While this paper attempts to present a precise model of meaning and context, the main goal is to stimulate inquiry and exchange of ideas between disciplines.

2 Word-sense disambiguation

There are several situations in information technology, and particularly in natural language processing, where it is necessary or at least very desirable to know what a particular word or expression is being used to mean. This begins in very simple ways in situations with which most of us will be familiar. If you're

booking an airline ticket online and say you want to travel to London, you will probably get a response such as:

> More than one airport matches 'london'. Please type a new airport name or code, or select an airport from the following list:
>
> | | |
> |---|---|
> | London, England | LGW - Gatwick |
> | London, England | LHR - Heathrow |
> | London, England | LON - all airports |
> | London, England | STN - Stansted |
> | London, OA, Canada | YXU - London |

The system which your browser is sending information to has a list of possible airports all over the world, each of which has a 3 letter code which means that airport and no other. A passenger requesting a flight to *London* could wish to travel to any of these airports. Effectively, each of these 3 letter code is a possible 'sense' of the word *London*, and to interact effectively the system must know which of these meanings are acceptable to the user. These meanings are not all created equal — *LHR*, *LGW* and *STN* are collectively referred to by *LON* [1] and this broader classification might be specific enough for some situations, so long as we know that *YXU* is completely different and must be distinguished. The user can see these options and choose the appropriate one for their travel needs — and if the system had more information about these travel needs it should be able to make the correct decision for itself. The word *London* is thus ambiguous (more particularly *polysemous*, meaning "many senses") and the process of deciding which meaning is appropriate is known as 'ambiguity resolution' or 'disambiguation'.

Most word-meanings in natural language have a much more complex structure than airport codes. There is a finite number of distinct airports and the system can rely on the user to say unequivocally which one is the correct choice. When a new airport is built, there is a recognised process for giving it its own unique code, the codes are universally recognised, and people do not start using a new and different code to refer to an airport which already exists. In natural language, people break all of these rules, and a careful analysis reveals all sorts of persistent difficulties in deciding what counts as a word-sense and which sense is being referred to in a particular instance [2].

Research on word-sense disambiguation (WSD) in natural language processing goes back several decades, and a historical survey of much of this work can be found in [3]. The central theme is that we should be able to work out which sense of a word is being used by examining the context in which a word is written (almost all work in WSD has been on text rather than speech). Initial work focussed on laboriously building 'expert' classifiers which would enumerate the different contexts in which a word might appear, with enough information to work out which sense was being used. Later on, as machine readable dictionaries (MRD's) became available, they were used automatically to provide information for disambiguation. An early and quite representative approach was that

[1] In a hierarchy of concepts, we might say that *LON* is a 'hypernym' [1, p 25] or 'parent node' of *LHR*, *LGW* and *STN*.

of Lesk [4], who used the words found in definitions from a MRD as clues that might indicate that one sense was being used rather than another. For example, one sense of the word *ash* is defined as

> 1) any of a genus (Fraxinus) of trees of the olive family with pinnate leaves, thin furrowed bark, and gray branchlets (Webster 1997)

In a sentence, one might very well encounter the words *tree* or *leaves* near the word *ash*, and use that as evidence that this sense of *ash* is being used. The problem here is that the information provided is very sparse, and often very different in register from the words one encounters in more normal text. For example, the definition of the other major sense of *ash* includes the words *residue* and *combustible* and makes no reference to the words *cigarette* or *dust*, which might be much better contextual clues. Lesk's solution to this mismatch was to use not only words occurring in the definition of *ash*, but words whose definitions share words with that of *ash*. For example, *dust* is related to the 'residue from combustion' sense of *ash* and both definitions include the word *particles*.

A method of relating "definitions" which more clearly reflect actual usage is to use the "one-sense per collocation" assumption [5]. This works upon the premise that idiomatic phrases can be used as 'seed-examples' of particular senses with very high reliability. For example, Yarowsky distinguishes between the 'living thing' and 'factory' sense of the word *plant* by assuming that almost every instance of the collocation *plant life* uses *plant* to mean a living thing, and almost every instance of *manufacturing plant* uses plant to mean a factory. These examples can then be used to find other words which indicate one sense or the other, and so gradually extend coverage [6].

Such a method involves taking a few labelled instances of one sense or another, examining their surrounding contexts, and extrapolating to achieve a similar labelling in other similar contexts. To get a more representative sample of instances of different senses, the initial labelling phase can be carried out (at some cost) by human annotators. This gives rise to the process of *supervised* word-sense disambiguation, where statistical tendencies observed in labelled examples are used to classify new unseen examples [3, §2.4.2]. The most standard model used is naive-Bayes, where "indicative words" are extracted from the training examples and used as evidence of the sense they occured with in training if they appear within a distance of n words from the target word.

By the late 1990's, WSD had become a recognised field within natural language processing, with its own internally defined standards and SENSEVAL evaluation contests [7]. These contests provide manually annotated training data and held-out evaluation data to compare the performance of different systems. This framework encourages supervised methods, which perform understandably better in such situations, though it is unlikely that enough hand-labelled data will ever be available to use such disambiguation techniques in real situations with many thousands of different words.

Even after much progress in the last ten years or so, one of the clearest results is that the 'success' of a disambiguation system depends critically on whether it is evaluated on easy or difficult ambiguities. Comparing decisions

made by different human judges shows that people often disagree with one another in labelling a particular usage with a particular sense [3, §3.3], and these discrepancies cast doubt on the significance of results obtained by comparing a word-sense disambiguation system against 'gold-standard' human judgments.

A major question is granularity: what distinctions in meanings are regarded as important enough to be considered different senses. Sometimes two senses are clearly different, such as the 'commercial-bank' and 'river-bank' senses of the word *bank*. This is an example of 'homonymy' or 'accidental polysemy', where two unrelated meanings share the same surface word-form almost accidentally. But many cases are more subtle — *bank* in the commercial sense is used to mean a financial institution and a building where that institution does business. Sometimes this distinction is not important — in the sentence

The bank closed at 5pm.

it probably means that the institution stopped doing business and the doors to the building itself were shut, so both senses are in use simultaneously. However, if one heard that

The bank collapsed at 5pm.

one would need to know whether it was the institution or the building which collapsed to determine whether to call in liquidators or rescue-workers. This sort of ambiguity is called 'regular polysemy' or 'systematic polysemy' — two senses are logically related, or one sense has different aspects which might be called upon in different situations. It is increasingly apparent that word-senses are not discrete, distinct units but adaptable and generative [8]. Rather than viewing ambiguity as a problem, one approach to creating lexicons is to use ambiguity as a guiding principle in defining systematically polysemous categories [9]. It is possible that the very choice of the term 'disambiguation' to describe the process of mapping word-usages to word-senses has led to a more divisive view of word-senses than is healthy.

Whether or not it is important to distinguish between senses depends very much on the task at hand. For translation, it is only important insofar as it determines the correct translation of a given word — if a single word in English with two possible senses is translated to the same word in French irrespective of which sense is being used, then the distinction does not matter for determining the correct translation. In information retrieval (IR), on the other hand, the distinction between two senses is important if and only if knowing the distinction could affect whether a particular document is relevant to a user. One solution to this problem in IR is to let the document collection itself determine which sense-distinctions are important. This entirely unsupervised approach to WSD was taken by Schütze [10, 11], finding different 'senses' of particular words by identifying distinct clusters of 'context-vectors'. For example, the word *suit* often occurs with the words *court, judge* and *attorney*, or with the words *jacket*, *tie* and *dress*, and words from each of these groups occur far more often with one another than with words from the other group. These two groups fall into very

distinct 'clusters', and once this has been noted the words in the different clusters can be used as clues that one sense or another is being used. This is one of the few methods for WSD that was shown to provide a reliable enhancement for information retrieval [12], possibly because it reflects the way words and senses are used in a particular document collection rather than how a dictionary says they should be used in general.

All of these methods for WSD share that feature that in some way, instances of words are being mapped to possible senses based upon information in a surrounding context. Unfortunately, the way 'context' is defined and used is often unstructured and inflexible. There are many kinds of context which might be relevant to determining meaning (for example, [3, §3.1] discusses *microcontext*, *topical context* and *domain context*), but much of this information is often neglected.

The practical challenge of combining information from different knowledge sources into a single system for disambiguation has been addressed with considerable success [13, 14], and in combining different knowledge sources these methods implicitly combine the different contextual information that the different knowledge sources rely upon. A theoretical challenge, which is of possible interest to a broad group of people, is to give an account for words and their meaning in context which can recognise and explain the wide variations we encounter in language.

3 A mathematical model for words, meaning and context

The techniques for word sense disambiguation described in the previous section have several factors in common. These similarities encourage the development of a more structured mathematical model which we now present, beginning by introducing three important sets which we need in order to phrase these questions mathematically. These are the spaces of words, of meanings, and of contexts.

$\mathcal{W}$	Words	Primitive units of expression Single words Parts of compound words (eg. *houseboat*) Independent multiword expressions
$\mathcal{L}$	Lexicon	The available meanings to which signs refer Traditional Dictionaries Ontologies, Taxonomies Meanings collected from training examples
$\mathcal{C}$	Contexts	Pieces of linguistic data in which signs are observed Sentences Immediate collocations (eg. blood *vessel*) Whole domains have conventions (eg. acronyms)

The main goal of this paper is to explore the structure of $\mathcal{C}$, the set of possible contexts, and we will do this in the following section. Before this, we devote a few remarks to the sets $\mathcal{W}$ and $\mathcal{L}$, and to some important mappings between these spaces [2].

In discussing a space of words or expressions, we are assuming that a sentence is somehow decomposable into basic units. These will not necessarily be handy strings of letters with spaces in between them — some words can be identified as compounds of more primitive units (such as with the word *fishbone* = *fish* + *bone*), and some multiword expressions or 'collocations' have a more or less independent meaning of their own (such as *United States of America*). Sometimes it is not clear whether something is really a multiword expression or two primitive units or both (such as *Catholic Church*. While it is possible (and depressingly easy) to use such objections to claim that 'words' don't really exist at all, even very simple computational and statistical techniques can analyse a corpus of English and identify words and collocations which most of us would agree are a reasonable choice of 'primitive units' [15, Ch 5].

The structure of the lexicon $\mathcal{L}$ is much more subtle and has been the subject of considerable research [16]. The lexicon describes the senses used in the domain in question, for example a traditional dictionary or thesaurus, an ontology or a taxonomy. If the domain is computer manufacturing, the lexicon $\mathcal{L}$ might be an ontology describing parts of computers, and information about where they go, what their purpose is, how they relate to other parts, etc. More generally, lexical resources such as WordNet [1] (which is domain-general) try to locate concepts by introducing broad, general categories such as *event* or *artifact* and giving relations between the entries such as IS_A, PART_OF, etc. A standard objection is that all these 'lexicons' are finite and discrete, but it is impossible to enumerate all the nuances of meaning a word can have in advance because new words are always being coined and old words being used with new meanings which are intuitively related to their old ones (examples include words which have acquired a technological meaning such as *web* and *mouse*). An alternative to the enumerative approach is provided by the theory of generative lexicons in which a core meaning can be systematically extended to new situations [8].

Another way to build a 'lexicon' is to take labelled training data. Instead of a definition being a statement about what a word means or what other words it is related to, concepts are defined by a list of contexts in which they are represented by a given word. This way of defining a lexicon has some drawbacks in common with those above. It is costly and static: once the labelled instances have been used, you can't then ask the lexicon for more information, or information about a different domain, hence the difficulty in applying WSD methods that require training data to any domain other than that for which the training data was designed.

[2] Note that authors sometimes use the term 'lexicon' to refer not only to the meanings and the way they are organised, but also to the words and the various possible mappings from word-forms to lexical entries, effectively amalgamating the spaces $\mathcal{L}$ and $\mathcal{W}$. The main reason here for using a notation where these spaces are kept separate it to be able to describe the process of assigning word-senses to word-usages as a mapping from $\mathcal{W}$ into $\mathcal{L}$.

Mappings between $\mathcal{L}$, $\mathcal{W}$ and $\mathcal{C}$

Many important linguistic phenomena can be defined as mappings between these spaces or products thereof.

Traditional WSD

The traditional WSD problem is as follows. Take a predefined collection of meanings $\mathcal{L}$ and a collection of pairs of words and contexts $(w, c) \in \mathcal{W} \times \mathcal{C}$. Produce a mapping $\phi : (w, c) \to \mathcal{L}$ and compare the results with a 'gold-standard' of human judgments. We will refer to such mappings ϕ as 'sense-mappings'. The possible images of this map for a given word w,

$$S(w) = \{\phi(w, c) \text{ for any } c \in \mathcal{C}\} \subset \mathcal{L},$$

can be referred to collectively as the senses of w.

The WSD problem can be posed using any one of the lexicon types described above. One significant difference between hand-coded dictionary or thesaurus definitions and definitions extracted from training data is that for the latter we do have a collection of examples of the form $\phi(w, c) = l$. (In fact, this is all we have.) The traditional WSD problem is to extrapolate ϕ to other pairs of symbols and contexts. In practice, WSD algorithms are only evaluated on how well they generalise to other contexts: to extrapolate from known instances $\phi(w, c_1)$ to unknown instances $\phi(w, c_2)$. This goes some way to explaining to the new reader how narrow much work in WSD has been, which must surely suggest partial answers to questions such as "why has WSD not proven very useful for information retrieval?".

Synonymy

The words $w_1, w_2 \in \mathcal{W}$ are said to be *synonymous* in the context of $c \in \mathcal{C}$ if $\phi(w_1, c) = \phi(w_2, c)$. Synonymy is the name given to the phenomenon that mapping from $\mathcal{W}$ into $\mathcal{L}$ will not in real life be injective. Synonymy in one particular context could accurately be called 'partial synonymy'. Two symbols are said to be *totally synonymous* if they have the same meaning in *all* contexts, so that $\phi(w_1, c) = \phi(w_2, c) \ \forall \ c \in \mathcal{C}$. It is known that total synonymy is very rare — there is nearly always some context in which two different words will have (at least slightly) different meanings.

WSD with learning or 'eavesdropping'

Introducing some learning operation is likely to help WSD. Along with the dictionary and the local context, the system is encouraged to look at a range of unlabelled data that has been collected along the way to see if it gives any further clues for accurate sense-mapping. Effectively, our system can eavesdrop on lots of conversations to see if it gets any further clues about what the symbols are used to mean.

Thus our sense-mapping is a function not just using one context $c \in \mathcal{C}$, but using any subset of contexts in $\mathcal{C}$. If we let $\mathcal{C}_w$ denote those contexts that are potentially relevant to the word w, our sense-mapping takes the form $\phi : (w, c, \mathcal{C}_w) \to \mathcal{L}$.

This method uses our initial definitions $\mathcal{L}$ as 'seeds of meaning' and allows those seeds to sprout, gathering more information, before we expect the seeds to accurately catch the correct meaning of all the symbols which refer to them. One way of doing this is to use a corpus to extract semantic similarity scores between pairs of words and to use these similarities for smoothing (assigning a non-zero probability to events that were not observed in the training corpus but are possible nonetheless) to improve disambiguation [17].

4 Context Groups

In this, the most important section of this paper, we use the mappings from the previous section to define the notion of context groups. This essentially topological notion can be used to address the question of how much context is appropriate in a particular situation for accurately determining meaning.

The structure of the set $\mathcal{C}$ is of particular interest and flexibility. A context, on the global level, might be any subset of the total 'universe of discourse', and on the local level is some piece of language containing the word w which we want to map into the lexicon. The elements of $\mathcal{C}$ might take many different (and nested) forms from a single word to a whole domain. How much context we need to distinguish meanings is an important question for determining a suitable 'granularity' for $\mathcal{C}$, and as we have stressed in this paper, the answer to this question will vary considerably from case to case.

Many approaches to disambiguation (such as naive-Bayes and Schütze's vector models [11]) have assumed a model where a 'context' c is simply a list of words $(w_1, \ldots, w_n)$ (often without taking the order of words into account). These disambiguation systems therefore provide a mapping from $\mathcal{W} \times \ldots \times \mathcal{W} = \mathcal{W}^n \to \mathcal{C}$. However, it is clear that the space $\mathcal{C}$ also contains broader 'meta' information which, though it can be described in words, is essentially not word-like. We would agree with the statement "in the context of medicine, *operation* usually means a surgical process (rather than a mathematical or a military operation". But this contextual information is certainly different from the usage of the single word *medicine* — we have an understanding of contexts as something that are often broader in scope than the individual words we use to name them.

One clear observation is that the relationship between granularity of contexts and granularity of senses is monotonic. That is to say that if two senses are to be clearly distinguished, then the contexts that include them must also be clearly distinguished and if the distinction between senses is relaxed, then the distinction between contexts can be relaxed. It follows that any measure of similarity or distance between senses will be mirrored in any corresponding measure of similarity or distance between contexts. This observation points to a way of defining the way context and meaning relate to one another without saying

exhaustively what each *individual* context can consist of. This is accomplished by defining 'context groups' [3].

The context group of a word w with meaning l consists of precisely those linguistic situations under which that particular word will have that particular meaning. Phrases like 'in the legal context, *suit* means the same thing as *lawsuit*' are then given the following interpretation. Suppose the lexicon gives two meanings for the word *suit*, the legal meaning and the clothing meaning. Then as far as we are concerned, the 'legal context' $\mathcal{C}_{\text{legal}}$ is precisely those situations in which *suit* has the same sense as *lawsuit*,

$$\mathcal{C}_{\text{legal}} = \{c \in \mathcal{C} | \phi(suit, c) = l\},$$

where l is the 'lawsuit' meaning of suit.

This definition of context as an inverse image of the sense-mapping ϕ is essentially topological. Words appear in a 'neighbourhood' of surrounding information, and it is these neighbourhoods which are used to resolve any ambiguity [4]. Other contextual information may well be available but often unnecessary — if we know for sure from the topic of one article, or from a single sentence, that *suit* is being used to mean lawsuit, then all other observations in (say) the context of that publication can safely be treated as factors under which the meaning of suit remains constant.

In this localised version of events, the context group can be reduced to a local vector space of contextual symmetries [20, Ch 8]. Placing a word in a particular context is then conceptually similar to placing a ball on a sloping surface. It is the local structure of the surface, in particular the plane tangential to the surface at that point, which determines the direction in which the ball will roll. These 'local surfaces' will be orthogonal (or at least transverse) for semantically unrelated homonyms, and will have some intersection or overlap for senses which are more closely related through systematic polysemy.

Figures 1 and 2 give 2-dimensional projections of this information, derived automatically by analysing local collocations such as "arms and legs" in the British National Corpus. (Details of the corpus processing used to obtain these graphs are given in [21].) Two disjoint senses have been found for *arms* ('part of the body' and 'weapons'), whereas several systematically related senses have been found for *wing*, including the wing of an aeroplane, the wing of a bird, the wing of a building and a wing of an organisation such as a political party. These examples show clearly how different senses have made themselves apparent by appearing with different context groups.

This theoretical analysis suggests a simple order in which contextual information for disambiguation should be sought. Extremely local information such

[3] This terminology is not intended to imply a group structure in the algebraic sense [18, Ch 10], since we have not yet defined a binary operation upon contexts, though this is an open and promising question. Any reasonable combination of contexts from the context group of a particular word-to-sense mapping would be expected to preserve that mapping, so closure would certainly be satisfied.

[4] This analysis suggests some similarity between the idea of contexts which are in the same context group and possible worlds which are accessible from one another in intensional logic [19, Vol II].

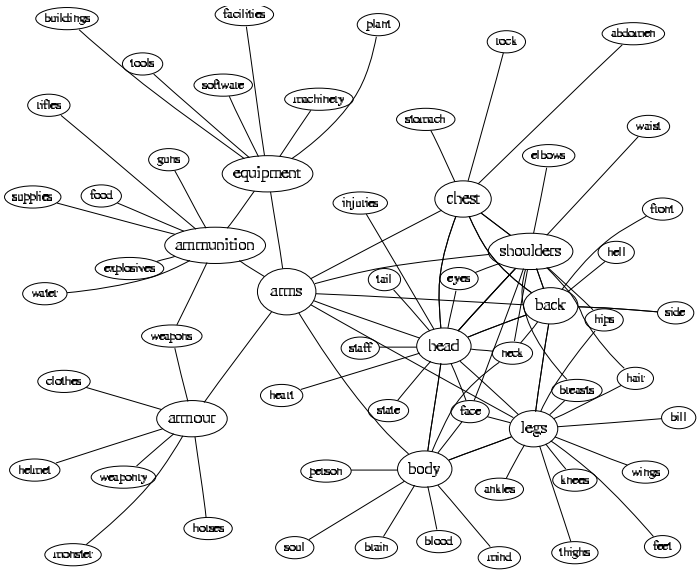

Figure 1: Words related to different senses of *arms*

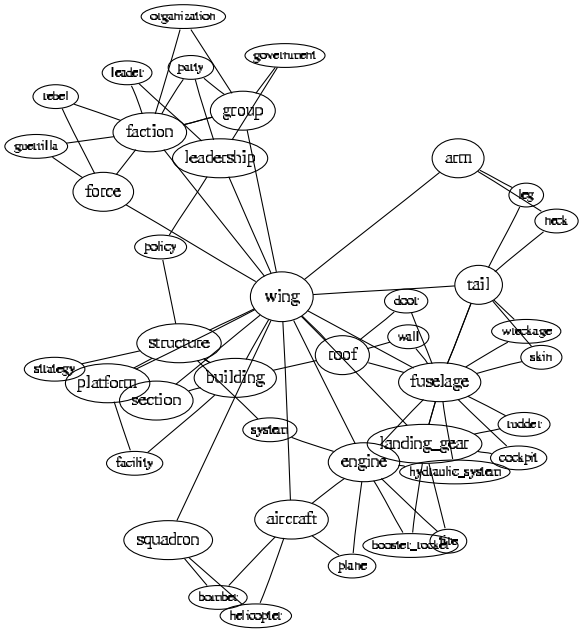

Figure 2: Words related to different senses of *wing*

as collocational usage should be prior, followed by local syntactic information, broader coccurrence information and finally broad discourse and domain information. The error made by statistical techniques [10, p 103] such as naive-Bayes for the sentence

> Sharpton said, "I have been on the attorney general's case, and I will be on his assistants like a suit jacket throughout the arraignment and the trial." (New York Times)

would be avoided by such a theory. Considering a context group as a topological neighbourhood around a particular word gives the simplest possible answer to the question "How much contextual information should be used to choose the correct sense?" — "Whatever is necessary".

5 The parallel with quantum theory

Consider an electron e orbiting the nucleus of a hydrogen atom. As is well-known, the electron may be in several different 'states'. Its spin may be 'up' or 'down' and it may be in a variety of different energy states [22], and the possible energy states are predicted by the localised action of a particular group of symmetries [20, Ch 11]. Without further observations, it is impossible to say for sure which of these energy or spin states the electron is in — all we know is the prior probabilities of the different states. To determine the 'actual' energy or spin state of the electron, a measurement of some sort must be made by a conscious observer.

Of course, there are many particles other than electrons, and many situations other than this simplest case of the single electron orbiting a hydrogen nucleus. With many electrons orbiting bigger nuclei, knowing the states of some electrons gives a lot of information about the rest. Many forces work in different ways on different particles — the protons and neutrons together in the nucleus are bound together by the strong nuclear force which is very powerful over a small locality, the electrons are chiefly affected by electromagnetic forces which are broader ranging, weak nuclear forces can cause radioactive decay, and all the particles influence one another by the force of gravity which has a weaker global effect.

This situation is curiously reminiscent of the relationship between words, senses and contexts described in this paper. A word $w \in \mathcal{W}$ may have several possible senses $l \in \mathcal{L}$, some of which are more common than others. Once the word is observed in some actual situation, it is possible to ask which sense is being used. The answer will depend on a number of forces — collocational, syntactic, broad cooccurrence and domain — with different strengths and ranges of influence. The best way to understand the possible states of a particle is to understand the group of symmetries which preserves each state, just as we have suggested that context groups give more insight on word-senses than individual contexts. Assigning meaning is an interactive, not an isolated process — knowing the senses of some words can greatly assist us in determining the senses of

others, and this process is often mutual. There are many common characteristics — the ambitious hope is that in years to come, scientific progress will provide theories which enable us to describe meaning with an accuracy similar to the accuracy with which quantum theory describes electrons.

A remaining difficulty is that the structure of word-senses is much more subtle and difficult to predict than the structure of energy-levels — at least for our current theories. However, this does not imply that meaning can only be written down exhaustively, guessed by statistical methods (both of which have already made great contributions) or left a mystery. The proposal in this paper is that careful scientific thought, together with empirical investigation, can provide models in which context and meaning are described clearly, flexibly, and accurately.

References

[1] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge MA, 1998.

[2] Adam Kilgarriff. Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, 26(1–2):365–387, 1993.

[3] Nancy Ide and Jean Véronis. Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics*, 24(1):1–40, March 1998.

[4] M. E. Lesk. Automated sense disambiguation using machine-readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the SIGDOC conference*. ACM, 1986.

[5] David Yarowsky. One sense per collocation. In *ARPA Human Language Technology Workshop*, pages 266–271, Princeton, NJ, 1993.

[6] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 189–196, 1995.

[7] A. Kilgarriff and J. Rosenzweig. Framework and results for english senseval. *Computers and the Humanities*, 34(1-2):15–48, April 2000.

[8] James Pustejovsky. *The Generative Lexicon*. MIT press, Cambridge, MA, 1995.

[9] P. Buitelaar. *CoreLex: Systematic Polysemy and Underspecification*. PhD thesis, Computer Science Department, Brandeis University, 1998.

[10] Hinrich Schütze. *Ambiguity resolution in language learning*. CSLI Publications, Stanford CA, 1997.

[11] Hinrich Schütze. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–124, 1998.

[12] H. Schutze and J. Pedersen. Information retrieval based on word senses. In *Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval*, pages 161–175, Las Vegas, 1995.

[13] Mark Stevenson and Yorick Wilks. The interaction of knowledge sources for word sense disambiguation. *Computational Linguistics*, 27(3):321–349, 2001.

[14] Mark Stevenson. *Word Sense Disambiguation: The Case for Combining Knowledge Sources*. CSLI Publications, Stanford, CA, 2003.

[15] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 1999.

[16] L Guthrie, J Pustejovsky, Y Wilks, and B Slator. The role of lexicons in natural language processing. *Communications of the ACM*, 39(1):63–72, 1996.

[17] Ido Dagan, Lilian Lee, and Fernando C. N. Pereira. Similarity-based models of word cooccurrence probabilities. *Machine Learning*, 34(1–3):43–69, 1999.

[18] Barbara H. Partee, Alice ter Meulen, and Robert E. Wall. *Mathematical Methods in Linguistics*. Kluwer, 1993.

[19] L.T.F. Gamut. *Logic, Language, and Meaning*. University of Chicago Press, 1991.

[20] William Fulton and Joe Harris. *Representation theory - a first course*. Number 129 in Graduate Texts in Mathematics. Springer-Verlag, 1991.

[21] Dominic Widdows and Beate Dorow. A graph model for unsupervised lexical acquisition. In *19th International Conference on Computational Linguistics*, pages 1093–1099, Taipei, Taiwan, August 2002.

[22] H. F. Jones. *Groups, Representations and Physics*. Adam Hilger/IOP Publishing, 1990.