In Merlo & Stevenson, Eds., The Lexical Basis of Sentence Processing, 1-38, 2002. [Draft version]

Words, numbers and all that: the lexicon in sentence understanding

Suzanne Stevenson and Paola Merlo

1 Cross-disciplinary Issues in Lexical Theories

It is hardly a controversial statement that the acquisition and processing of language require knowledge of its words. Yet, the type and use of information encoded in a lexical entry, the relation of words to each other in the lexicon, and the relationship of the lexicon to the grammar, are complex and unsettled issues, on which researchers hold very different views. While there is a recent consensus that mechanisms operating in the lexicon are not substantially different from those operating in the syntax, there are differences on whether the syntax is in the lexicon, or the lexicon is in the syntax. On the one view, the lexicon is a static repository of very rich representations, which regulate the composition of words to an extent that goes beyond the phrase, while on the other view the lexicon is dynamically generated as a result of composition and competition mechanisms, largely syntactic in nature. Roughly speaking, computational linguistics and psycholinguistics in general follow the first view, and the two fields are converging on some similar lexicalised, probabilistic models of grammars. Theoretical linguistics has recently proposed models of the latter type.

In computational linguistics, work on the lexicon has stemmed from two different areas of research: parsing and grammar formalisms, and construction of electronic databases (lexicography). In the area of parsing, the interest in probabilistic models and lexicalised grammars did not develop simultaneously. Parsers based on probabilistic context-free grammars were motivated by the difficulties in building robust, large-scale systems using the explicit representation of linguistic knowledge. Large corpus annotation efforts and the creation of tree-banks (text corpora annotated with syntactic structures) enabled researchers to develop and automatically train probabilistic models of syntactic disambiguation (Marcus, Santorini, and Marcinkiewicz 1993). In an attempt to take advantage of the insights gained in the area of statistical speech processing, computational linguists initially adopted very simplified statistical models of grammar and parsing, abandoning the more sophisticated lexicalised feature-based formalisms (Magerman and Marcus 1991; Magerman and Weir 1992; Resnik 1992; Schabes 1992).

However, it soon became apparent that the success of probabilistic context-free grammars was limited by the strong (and incorrect) assumption of probabilistic independence of rule applications (Black, Jelinek, Lafferty, Magerman, Mercer, and Roukos 1992; Charniak 1996; Johnson 1998). The search for more context-sensitive models of disambiguation led to the development of probabilistic models that rely heavily on lexical heads. Current models of probability assume a lexicalised grammar, in which a syntactic rule is conditioned on its lexical head, as well as on the heads of its dependent constituents. In this way, a probability model is defined that takes into account lexical dependencies that go beyond a single context-free rule (Brew 1995; Collins 1996; Abney 1997; Charniak 1997; Collins 1997; Ratnaparkhi 1997; Johnson, Geman, Canon, Chi, and Riezler 1999; Srinivas and Joshi 1999). These models derive their formal specifications from (non-statistical) lexicalised grammar formalisms, where the needs for empirical coverage have similarly led to precise definitions of such grammars (e.g., Bresnan 1982; Pollard and Sag 1987; Mel'cuk 1988; Joshi and Schabes 1997).

The lexicalisation of computational models of grammar and parsing, and the emphasis on robust systems, has brought to the forefront one of the major practical and scientific problems in large-scale linguistic applications, namely the difficulty in acquiring and encoding lexical information. Manually building the rich lexical representations that are a central component of linguistic knowledge is time-consuming, error prone, and difficult, as shown by the effort required to produce electronic databases such as Wordnet (Miller, Beckwith, Fellbaum, Gross, and Miller 1990; Fellbaum 1998) and verb classifications such as Levin's (Levin 1993). The complexity of the task of hand-building lexical entries has sparked interest in extending the learning approaches developed for parsing to the inductive learning of fine-grained lexical classifications. By exploiting the implicit syntactic and lexical information encoded in annotated corpora, lexical knowledge can be induced from the statistical analysis of distributional data (Brent 1993; Briscoe and Carroll 1997; Dorr 1997; McCarthy 2000; Lapata and Brew 1999; Schulte im Walde 2000; Stevenson and Merlo 2000; Merlo and Stevenson 2001; Siegel and McKeown 2001).

In sentence processing, the path of development of lexical, probabilistic theories has been different, but reaching similar conclusions. In this context, a key question is the degree to which lexical information underlies distinctions in on-line processing difficulty, especially in the process of ambiguity resolution. Early work in sentence processing mirrored the emphasis on the syntactic component in grammatical theory, by focusing on the large-grained structural properties of interpretations as they are developed incrementally. For example, the most widely known statements of preference, Minimal Attachment and Late Closure (Frazier 1978), rely solely on general properties of the size and locality of incremental additions to a partial syntactic structure. Beginning with the work of Ford, Bresnan, and Kaplan (1982), there has been a gradual shift to emphasizing more and more the influence of individual lexical information within such structure-based accounts. For example, Pritchett (1992) examines the role of the lexical head of a phrase in determining its basic properties, and the impact of the argument structure of a verb on preferred structural representations. However, while some lexical properties are assumed to influence processing decisions in this type of structure-based account, they are viewed as secondary to structural information.

The recent lexicalist constraint-based approach in sentence processing takes the lexical basis of language comprehension much further, suggesting that perhaps all of the relevant processing distinctions can be traced to distinctions in lexical information (e.g., MacDonald, Pearlmutter, and Seidenberg 1994; Spivey-Knowlton and Sedivy 1995; Spivey-Knowlton, Trueswell, and Tanenhaus 1993; Trueswell 1996). On this view, interpretation is an incremental process of satisfying constraints associated with lexical entries. Dynamical models inspired by the connectionist literature are used to describe the integration of multiple numerically-weighted constraints (e.g., Spivey and Tanenhaus 1998; Tabor and Tanenhaus 1999). Typically, the frequency of lexical features and their co-occurrence is believed to be the basis of the weights, and therefore frequency plays a primary role in determining sentence processing behaviour. These models, analogously to the corpus-based work in computational linguistics, tie together processing and learning, as much of the lexical frequency information needed for parsing is learned from exposure.

In both disciplines, increased lexicalisation raises interesting issues concerning the role of frequency (or probability) in parsing, the conception of parsing itself, and the issue of incrementality. In computational linguistics, the automatic acquisition of lexical knowledge through statistical analysis entails an emphasis on frequency information, as corpus counts are used to estimate probabilities. This brings up the important issue of what are the relevant entities or features of entities to count in order to achieve accurate probabilistic parsers. In lexicalist sentence processing as well, this issue of what to count has been a focus of much attention—i.e., for exactly what types of lexical information do humans keep track of frequencies (morphological, syntactic, thematic, semantic, etc.), and what is the grain of frequency information that influences human parsing (Mitchell, Cuetos, Corley, and Brysbaert 1995; Gibson, Schütze, and Salomon 1996)?

One also sees an influence of lexicalisation on how parsing is conceived. Both in computational linguistics and in lexically based models of sentence processing, there is consensus that the processing primitives are rather large units specifying a lexical head and all its argument relations. In this way, the primitive grammatical objects correspond to a lexicalised specification of the non-recursive set of substructures that come into play in building a parse. This notion can be realized in one of two ways. One option is an explicit representation of grammar-in-the-lexicon, such as in tree-adjoining grammars or lexicalist constraint-based models (MacDonald, Pearlmutter, and Seidenberg 1994; Srinivas and Joshi 1999), in which the lexical entries are themselves tree structures. Lexicalised probabilistic models of parsing adopt an alternative approach which implicitly associates grammatical rules with lexical entries, by conditioning each rule on the lexical heads of it and its children. In conjunction with the emphasis on frequencies and probabilities, parsing is viewed within both these formulations as a competition of structures, whose respective probabilities depend both on the substructures composing them and on the words that they contain. In connectionist models, this competition is represented directly in the component processors, through their levels of activation. In a probabilistic parser, the competition is encoded as the ranking given by the probability of each parse.

Lexicalisation raises further questions related to incremental parsing. First, lexicalisation tests the limits of incrementality, as it assumes that each word can be directly integrated into the previously determined structure. Also, the related assumption that the domain of influence—the domain of locality—of each individual word extends to units larger than the immediate phrase raises interesting technical issues of how exactly the integration of each word in the parse tree occurs. In this view, each incoming word gives rise to a complex set of interactions with the existing structure, where several mutual constraints—determined by both the word and by the structure—have to be satisfied. The proper description of these mechanisms has consequences for the choice of both the grammatical representations encoded in the lexicon and the parsing architectures.

While computational linguistics and sentence processing have focused on the locus of lexical statistics and the process of lexically-guided parsing, theoretical linguistics has focused on the organization of the lexicon. The increased complexity of lexical entries in computational linguistics and sentence processing appears in contrast with the final goal in theoretical linguistics, which consists in reducing the lexicon to its primitive components and describing its regularities. Much linguistic knowledge is situated in how lexical entries are organized with respect to each other. On this view, lexical entries tend to be simplified, while the organisation of words in the lexicon and the integration of words with sentence construction becomes more complex.

For example, Levin's (1993) work on verb classes aims at reducing the information in a lexical entry to its primitive meaning components (see also Levin 1985; Pinker 1989). Under the hypothesis that semantic properties of verbs largely determine their syntactic behaviour, the linguistic knowledge about a verb consists in its specific set of meaning components along with general relations between each meaning component and its possible syntactic expressions. The architecture of Wordnet (Miller et al. 1990), an electronic lexicon organised on psycholinguistic principles, provides another example of the complexity of organisation in the mental lexicon. Components of meaning do not always appear in the actual definition of the word, rather they are sometimes part of the organisational structure of the lexicon, as relational notions between entities. For example, the notion of causation does not appear as part of the meaning component of certain verbs, such as *melt*, but as the relation connecting two senses of the verb *melt*, the intransitive (not causative) and the transitive (causative).

Thus, as lexical entries are simplified, the organizational knowledge itself becomes more complex, in many cases appearing syntactic in its formal nature. Furthermore, as the lexical entries become impoverished, the notion of "projecting" syntactic structure from the lexicon becomes less tenable. In fact, taking this form of reasoning to its logical conclusion, current conceptions in Optimality Theory view the lexicon no longer as the input to the sentence level but as the result of syntactic competition (Prince and Smolensky 1993). Constructionist approaches to the lexicon also propose to simplify the lexical entries and devise more complex relations to link them. For instance, some have argued that the lexical entries themselves are syntactic—that is, lexical structure is not only predictive of syntactic structure, but is itself subject to syntactic processes within the lexicon, such as the Head Movement Constraint (Hale and Keyser 1993). Others have proposed that syntactic structures themselves are the bearers of meaning (Fillmore 1988; Goldberg 1995).

Thus, there seems to be consensus across disciplines that there is more structure in the lexicon than previously assumed, and that similar mechanisms operate in the lexicon and in the syntax. One area of disagreement is in the conception of lexical entries themselves. In theoretical linguistics, these are generally assumed to be minimal, while in computational linguistics and sentence processing, they are by contrast entire subtrees. A second issue on which the fields differ is the nature of the competitive processes at work in the lexicon. In sentence processing and in computational linguistics, lexical processes are fundamentally a numeric competition, while in Optimality Theory the lexicon can be viewed as the result of a non-probabilistic syntactic competition process. In evaluating the arguments on the differing sides of these two issues, it is important to broaden one's view beyond a particular discipline, as the evidence for an encompassing theory of language will come from many sources.

For example, the lack of consensus on what constitutes a lexical entry is often resolved in favour of the most succinct description, on the basis of Occam's razor. However, this is appropriate only if the sole criterion is simplicity of representation. Given that the language faculty has a computational component, simplification in the description of knowledge might lead to more complex computations. A more appropriate metric would be one in the spirit of the Minimum Description Length principle, which characterizes simplicity as a function of both the complexity of the data and the complexity of the computation (Rissanen 1989). This points to the inherent multidisciplinarity of the lexical enterprise, as this principle clearly cannot be applied, and consequently the preferable approach cannot be decided on, in the absence of a precise model of both the representations and the computations involved.

Another example arises concerning the nature of lexical competition. The existence of competition effects in sentence processing has been assumed to argue in favour of a probabilistic, connectionist architecture. But frequency effects can be incorporated in a natural way in other types of models, such as probabilistic parsers that are not connectionist in nature, and competition can be naturally expressed in a symbolic model (as in Optimality Theory). As some of the contributions in this volume argue convincingly (e.g., Spivey et al.), it is premature to infer architectural organisations from currently available empirical evidence. Rather, a broader range of both human and computational experiments must be pursued to provide mutually constraining evidence about plausible language processing architectures.

The resolution of these open questions in lexical theorising thus requires accumulation of evidence from all three disciplines of theoretical linguistics, computational linguistics, and sentence processing. Very precise theories must be developed, along with more elaborated computational models and new experimental methods, that, as evidence accrues, will be mutually constraining. This volume represents an attempt to bring together leading research on the lexicon across these three fields, in particular focusing on work that influences computational views of the human sentence processor. In the remainder of this introduction, we outline brief summaries of the contributions to the volume, and discuss current issues in theories of lexical information and processing that cut across the groups of papers collected here.

2 Tour of the Volume

This volume derives from the special conference session entitled "The Lexical Basis of Sentence Processing: Formal and Computational Issues," held in conjunction with the 11th Annual CUNY Conference on Human Sentence Processing, March 19–21, 1998. The special session focused on current theories of the lexicon from the perspective of its use in sentence understanding. Participants included a multidisciplinary slate from theoretical linguistics, computational linguistics, and psycholinguistics, representing various theoretical frameworks within each of these disciplines.

By analogy with the conference special session, the volume is organized into three parts: Part I presents theoretical proposals on the lexicon and discusses their relation to sentence processing; Part II explores the relationship between syntactic and lexical processing; and Part III investigates more specific issues about the content of lexical entries. Here we briefly discuss the contributions within each part; we return in Sections 3 and 4 to issues that cut across these broad topic areas.

2.1 Part I: Fundamental Issues

Part I of the volume contains papers that elaborate on foundational issues concerning the nature of the lexicon and its connection to processing. Bresnan and Fodor address two fundamental aspects of a lexical item: its lexico-syntactic paradigm (i.e., its relation to other items in the lexicon), and its combinatory properties (i.e., its relation to phrases in which it can occur). Both chapters bring up issues surrounding the key topics discussed earlier, of the kind of structure in the lexicon and the type of operations used within it. Bresnan espouses a view of language as a system of contrasts, where the meaning and use of a word are determined by competition with the other members of its paradigm (which includes both words and phrases), rather than by the intrinsic features of the word. Fodor on the other hand is concerned with the syntagmatic relations of a word—the hierarchical and compositional properties of words and phrases that mutually determine their semantic content. Johnson and Weinberg are direct commentaries, respectively, on how each of these views relates to parsing and sentence processing—in particular, to current probabilistic proposals. In the final chapter of Part I, Steedman tackles similar issues from a computational perspective, exploring the complex problem of how to integrate a competitive-ranking view of language processing with the structure-building necessary to derive adequate semantic representations.

Bresnan, "The Lexicon in Optimality Theory" In her chapter, Bresnan elaborates on the proposal that the lexicon is not the source but rather the result of syntactic variation. She focusses on markedness facts, providing an account of the lexical forms that surface in instances of variation, specifically explaining the gap in the paradigm of negation in English. In Bresnan's view, words and phrases are elementary units that can compete with each other to be the optimal expression of an underlying form; specific lexical items result from syntactic competition induced by a language-specific ranking of universal constraints.

The proposed account views language as a system of inter-related and competing levels. In particular, Bresnan elaborates a generalised view of paradigmatic competition, within a framework that combines the violable constraint optimization of Optimality Theory with the rich feature specifications of Lexical-Functional Grammar. The classical markedness view of contrasting words is generalised to a competition of elements that are not necessarily lexical items, but also phrases or even larger fragments. For example, periphrastic and lexical forms of negation can compete with each other within a uniform competitive process. All forms are subjected to the very same constraints, whose ranking gives rise to the surfacing of the observed words.

Johnson, "Optimality-theoretic Lexical Functional Grammar" Johnson's paper focusses on the relevance of Bresnan's proposal for processing. The explanation of linguistic universals and markedness in terms of optimisation of well-formedness makes contact with current views of grammaticality and parsing in computational linguistics, brought about by the recent interest in probabilistic language models. Both Optimality Theory and maximum likelihood parsing involve selecting a parse of the input string over an ordinal scale. In this view, grammaticality becomes a relative or comparative notion. Some importants details differ across the two fields—for instance, Optimality Theory does not use continuous scoring functions as in statistical parsing approaches. However, there are also several similarities, especially with recent probabilistic exponential models.

Johnson also notes that Bresnan's general integration of Optimality Theoretic optimization with Lexical-Functional Grammar feature mechanisms may increase the formal complexity of the framework. The view of lexico-syntactic processes as competitive satisfaction of constraints departs from previous Lexical-Functional Grammar theory in a way that might affect its decidability. Specifically, it may not be decidable whether a given string belongs to this kind of grammar formalism, since competitors that are unboundedly different in structure need be compared. This is a topic of on-going research.

Fodor, "The Lexicon and the Laundromat" Fodor's chapter also concerns what constitutes the lexicon. But while Bresnan's paper focuses on the explanation of markedness effects in a paradigm, Fodor's paper concentrates on relations that are purely syntagmatic. The relationship of the words to each other in the lexicon are not of concern here, but rather the relationship of each word to its host—the sentence or phrase it can occur in.

The theoretical decision of what is in the lexicon, and (as importantly) what is not, is guided, according to Fodor, by two necessary principles for lexical well-formedness and interpretability—compositionality and reverse compositionality. The principle of compositionality states that the linguistic structural description of a host is entirely determined by the linguistic structural description of its constituents (plus principles of construction). Reverse compositionality states that the grammar of the constituents is exhausted by what they contribute to their host.

Under this view, lexical entries contain the minimal information that supports compositionality without violating reverse compositionality—that is, lexical entries cannot contain more than what they contribute to the interpretation of the phrases they occur in. Fodor claims as a consequence of this view that frequency information cannot be associated with lexical entries, because it would violate reverse compositionality. Specifically, according to Fodor, since the relative frequency of a host does not depend on the relative frequency of its components, frequency cannot be a lexical property. (We discuss alternative views on this particular assumption in Section 3.2 below.)

Weinberg, "Semantics in the Spin Cycle: Competence and Performance Criteria for the Creation of Lexical Entries" Weinberg's commentary focuses on the apparent mismatch between the constraints imposed by Fodor's reverse compositionality principle, and the recent success of probabilistic (or frequency-based) lexicalised approaches to parsing in computational linguistics and sentence processing. Weinberg reconciles these differing views by noting that reverse compositionality is most appropriately seen as part of the competence theory. Theories that assume a competence/performance distinction do not enforce a one-to-one mapping for representations at those two levels. Thus, lexical representations for processing that extend the competence theory are not precluded provided they are learnable.

To illustrate the necessary competence/performance distinction, Weinberg shows that speakers can keep track of frequencies of properties that are either not distinguished in the competence theory, or are distinguished in a discrete, as opposed to graded, manner. Weinberg also notes that current lexicalist constraint-based approaches to sentence processing are not incompatible with a lexicon organised around compositionality and reverse compositionality. In the lexicalist constraint-based view, lexical entries are dynamically constructed, putting features together that are strongly associated. Highly correlated features (almost always) correspond to lexical entries. This is compatible with a view in which the definitional features are both compositional and reverse compositional, with principles of construction that include performance notions such as frequency or plausibility.

Steedman, "Connectionist and Symbolist Sentence Processing" In his chapter, Steedman notes that competition-based views of processing typically leave unspecified the mechanisms for structure building (see Bresnan's chapter, for instance). Steedman addresses the issue of whether such theories can be integrated with compositional approaches to language and parsing, whose output is a complete structural description, which he assumes to be necessary for semantic interpretation.

Steedman notes that the connectionist models known as simple recurrent networks have been claimed to be models of syntactic parsing; in fact, since their memory fades with time and space, they are effectively equivalent to finite-state devices, such as n-gram part-of-speech taggers. Instead of structured representations, their output is the prediction of the next grammatical element in the input. Such devices are very effective, and can be used to resolve a large amount of ambiguity, in lexicalised, sense disambiguated grammars (see the chapter by Kim et al., described below). However, neither n-gram part-of-speech taggers or simple recurrent networks produce structural interpretations that can support semantic processing; for example, they cannot disambiguate structural ambiguities such as PP attachment.

Steedman proposes that associative memory devices are more promising as a basis for structured lexical representations that support semantic interpretation. If these kinds of devices were used to learn lexical entries, in particular verbs, in a highly lexicalised grammar (such as combinatory categorial grammar, or lexicalised tree-adjoining grammar), then acquisition of these lexical items would carry a large amount of structural information.

2.2 Part II: Division of Labour Between Syntax and the Lexicon

The chapters in Part II address the general question of what is the basic architecture of the human sentence processor. The contributions use a range of computational methodologies, such as modelling and corpus analysis, and experimental methodologies, both behavioural and neuro-imaging, often offering complementary insights.

All of the papers focus on the particular issue of how lexical and syntactic information are integrated, or kept apart, in sentence processing. Kim et al., Crocker and Corley, and Stowe address the problem globally, by presenting computational or functional models. Questions that arise in this kind of enterprise are whether the processor manipulates different types of information at different levels, and whether these information types, defined functionally and computationally, belong to different modules of the processor or not. Spivey et al. and Lombardo and Sturt focus on a more restricted question related to the interaction of lexical and syntactic information. In both cases, they focus on a specific property of an existing model, exploring the conceptual and empirical consequences. Spivey et al. ask what kind of evidence is needed to argue for different stages of processing, while Lombardo and Sturt provide the corpus data to quantify the feasibility of a fully incremental lexicalised processor.

On one hand, the proposed models are increasingly sophisticated and make fine-grained predictions. On the other hand, the recent introduction of new conceptual elements (such as studying frequency of usage) or new methodologies (such as neuro-imaging) have not yet given rise to a convergence on a range of models that is more restricted than previously.

Kim, Srinivas, and Trueswell, "A Computational Model of the Grammatical Aspects of Word Recognition as SuperTagging" Kim et al. investigate whether current lexicalist constraint-based approaches can be precisely defined at the grammatical level and implemented on a large scale. They illustrate the implementation of a model of lexico-syntactic processing (Srinivas and Joshi 1999) based on the formalism of lexicalised tree-adjoining grammar, in which the grammar is entirely stored in the lexicon, as a forest of trees (Schabes 1991). The lexicon thus contains explicit calculation of (at least some) syntactic operations, leading to very rich lexical descriptions. By contrast, the grammatical operations for creating structure are minimal. The lexical trees are known as "supertags", by analogy with the simpler category-based part-of-speech tags typically associated with a lexical entry, and assigning the appropriate trees to lexical items during the parsing process is known as "supertagging". In this approach, the balance of work is shifted from purely syntactic combinatory operations to choosing the best lexical tree for a word.

In a representation of this kind, many potential syntactic ambiguities are grounded in a lexical ambiguity—i.e., the choice of lexical tree or supertag. Kim et al. therefore argue for a model in which the lexical disambiguator accomplishes most of the work of syntactic disambiguation. This view of the lexicon/grammar is then encoded in a distributed representation of the supertags, within a connectionist architecture that directly reflects the lexicalist constraint-based approach, with frequencies influencing the likelihood of choice of lexical tree. They proceed then to show that this general purpose implementation presents some of the behavioural effects that have been documented in the psycholinguistic literature, such as the frequency-by-regularity interaction and the bias-by-contextual-cues interaction.

Lombardo and Sturt, "Incrementality and Lexicalism: A Tree-Bank Study" Purely lexicalist theories propose that large amounts of the information necessary to parse a sentence must be prestored together with a given lexical item and activated during parsing. In these kinds of models, as with the proposal by Kim et al., the determination of how structure building exactly works is left under-specified. Lombardo and Sturt explore a crucial computational issue that arises in such an approach: how much non-lexically driven structure building is required in a parsing model that is fully lexicalised, but that also respects one of the basic assumptions of current sentence processing, namely that interpretation is incremental.

They investigate this issue through an analysis of a structurally annotated corpus—that is, a tree-bank. The use of a tree-bank enables them to provide an operational definition of incrementality in terms of structure: incrementality is the requirement of building a fully connected tree at all times. Their results show that 80% of the words can be incrementally integrated into a parse without the use of a headless projection (i.e., non-lexically-based structure), and that very few words require more than one headless projection. Moreover, in all the cases requiring a headless projection, there are systematic patterns related to the current word and the left context that can help in building structure. This result is important because it shows that a large amount, but not all, of incremental syntactic structure building can be accomplished through lexical projection.

Crocker and Corley, "Modular Architectures and Statistical Mechanisms: The Case from Lexical Category Disambiguation" In contrast to the lexicalist investigations above, Crocker and Corley argue in favour of a more traditional modular and pipe-lined architecture, along the lines of several large scale applications in computational linguistics (Brants 1999; Ratnaparkhi 1999). Like Kim et al., they propose a model in which lexical frequencies play a critical role in disambiguation, but in contrast to the supertagging approach, they claim that the distinction between lexical processing and syntactic processing is clearly demarcated. Conceptually, they base their assumption on the observation that having frequency information for the units at each level of grammar increases the amount of information available in each module. (We can remark that this intuition is supported by formal results about the greater power of statistical formal mechanisms, compared to their non-probabilistic counterparts (Cortes and Mohri 2000).) Consequently, they argue, the use of statistics supports a better encapsulation of the modules.

The chapter presents empirical evidence in support of their view. On one hand, they argue that the lexical level of processing makes use of a statistical model, based on data from an experiment showing that lexical statistics are used in disambiguation. On the other hand, they show that the lexical processor does not have access to syntactic information in its statistical calculation, through an experiment revealing that lexical statistical constraints are stronger than syntactic constraints in lexical disambiguation. Together, these results favour a statistical, but modular, language processing architecture.

Stowe, Withaar, Wijers, Broere, and Paans, "Encoding and Storage in Working Memory during Sentence Comprehension" Stowe et al. turn to the methodology of neuro-imaging to address the issue of the division between syntax and the lexicon, by attempting to localize different functions involved in sentence comprehension. Specifically, they claim that the following three language processing functions are associated with distinct areas of the brain: the encoding of lexical information, the storage of lexical and phrasal information in memory, and structural processing. Since theories of sentence processing typically appeal to notions of memory load and processing complexity as the basis for observed behaviour, Stowe et al. argue that their results have important ramifications, as they show that working memory and structural processing can be dissociated. For example, they note that a straightforward interpretation of theories that equate memory load with processing (such as Just and Carpenter 1992; Gibson 1998) cannot be strictly maintained.

In addition, of high relevance to the ideas in this part of the volume, we think that the evidence in the chapter by Stowe et al. challenges the degree to which (and the manner in which) syntax and the lexicon can be unified in models of human sentence processing. While they show that lexical and phrasal memory are not distinct—drawing on the same resources, in the same area of the brain—they also reveal distinctions between lexical encoding and memory on one hand, and structural processing on the other. These results indicate that proposals in which lexical and syntactic processing are the same (e.g.,MacDonald, Pearlmutter, and Seidenberg 1994, Kim et al., this volume) may be oversimplified when taken at face value. The Kim et al. chapter downplays the degree of processing involved in attachment ambiguities, but Stowe et al.'s neuro-imaging results suggest that the structuring of phrases, rather than just the encoding and storage of lexical information, is a central aspect of sentence processing.

Spivey, Fitneva, Tabor, and Ajmani, "The Time-Course of Information Integration in Sentence Processing" The papers by Crocker and Corley and Stowe et al. bring new kinds of evidence in favour of a modular organisation of processing, on functional or computational grounds. Spivey et al., on the other hand, question the very notion of stages of processing. Moreover, they present human and computational experimental results supporting the view that previous evidence taken to indicate stages of processing can be equally well explained within an interactive, non-modular constraint-based system.

Specifically, Spivey et al. compare the results and conclusions of McElree and Griffith (1995, 1998) to the conclusions that can be drawn on the basis of a dynamical model of parsing. McElree and Griffith had argued that subcategorization (as syntax) and thematic role information (as lexical semantics) come into play at different times in sentence processing. Spivey et al. re-interpret the results of the experiments to show that a different explanation can be found within Spivey's normalized recurrence model (Spivey-Knowlton 1996). In their computational system, all the factors influencing ambiguity resolution are lexicalized constraints, which, regardless of information content, are simultaneously activated at the moment of lexical access. However, the influence of various factors is felt at different points in time, due to differential weightings of the types of information. Thus, constraints that are simultaneously available may still have a "staged" influence on behaviour.

Spivey et al. conclude that it is important to distinguish differences in timing versus differences in strength of information sources, in both computational and human experiments, in order to differentiate current theories of human sentence processing.

2.3 Part III: Details of Lexical Entries

Part III of the volume contains papers that elaborate, through human experimental studies as well as corpus analysis, details concerning the information that is stored within a lexical entry. Whereas papers in Part II focus on a larger grain of analysis of a lexical entry—how "syntactic" is it?—here the emphasis is on finer-grained details of individual pieces of the stored information.

All of the papers focus particularly on argument structure properties of verbs—their representation and role in processing—reflecting the importance of verbs in guiding language understanding, both in computational linguistics and in theories of sentence processing. A theme that runs through the papers is the need to elaborate more clearly what constitutes the representation of argument structure for a verb, and how this representation relates to frequency biases that influence human behaviour. Some of the questions being raised are the following: Is the information conceptual in nature (i.e., real-world knowledge) or a set of more circumscribed formal properties (i.e., a more "linguistic" view)? How fine-grained is the representation of differences in argument structure across verbs? What is the relation between verb sense, verb argument structure, and frequency biases for verb/argument relations?

Three of the papers (Mauner et al., Filip et al., and Altmann) focus on determining the precise nature of the information that influences sentence processing, using psycholinguistic experiments to build evidence for the early use of more finely grained thematic and/or semantic information. The corpus-based work of Argaman and Pearlmutter, and Roland and Jurafsky, addresses a somewhat more general issue of the appropriate level of representation of frequency biases—focusing less on exactly how frequencies influence human processing and more on which aspects of lexical information they are stored with.

Mauner, Koenig, Melinger, and Bienvenue, "The Lexical Source of Unexpressed Participants and their Role in Sentence and Discourse Understanding" The main idea investigated by Mauner et al. is that, as part of the argument structure of a verb, implicit (i.e., unexpressed) arguments are made available to and used by the human sentence processor immediately at the verb. They show that, by contrast, this is not the case for entities that are merely implied by general conceptual knowledge. They conclude that the participants in an action that are *linguistically* licensed can influence processing earlier than those that are only inferrable.

This proposal is supported by experiments on two structures that have the same logical necessity of an Agent, but have different linguistic representations—a short passive in which the unexpressed Agent is nonetheless part of the argument structure of the verb, and an intransitive in which an unexpressed Agent is implied by the semantics of the situation, but is not part of the representation of the verb. Evidence from eye-tracking experiments reveals differences in processing between conditions with linguistically and conceptually derived implicit agents as early as at the verb, and in first pass reading times.

The authors further argue that their results indicate that thematic roles in processing must be finer-grained than the traditional labels such as Agent or Patient commonly adopted in linguistic theory. For example, they compare volitional and non-volitional agents of causation and show that "volitionality" influences the availability of an implicit agent in processing (precisely, it affects the suitability of an implicit agent as an antecedent for a volitional anaphor).

Filip, Tanenhaus, Carlson, Allopenna, and Blatt, "Reduced Relatives Judged Hard Require Constraint-Based Analyses" Filip et al. develop the role of fine-grained thematic relations in sentence processing even further, arguing that the linguistic differences between verbs in different lexical semantic classes can be captured with semantic features based on Dowty's (1991) featural decomposition of Proto-Agent and Proto-Patient roles. Using a questionnaire study, they verify Stevenson and Merlo's (1997) observation that class-based distinctions between verbs arise in processing, but argue for modelling these results with verb frequency values and thematic fit biases. Filip et al. further note that a fine-grained analysis of thematic roles (in terms of Dowty's linguistic properties) is key to a full understanding of the relation between verb class and behaviour.

A key novel observation in their work is that the difficulty of processing a reduced relative construction—such as *The horse raced past the barn* in *The horse raced past the barn fell*—is influenced by the main verb (i.e., *fell* in the example). They account for this effect in terms of the compatibility between the two sets of Proto role features assigned to the subject of the sentence by the two verbs—the verb in the reduced relative (*raced*) and the main verb (*fell*). They suggest that on-line processing of this type of construction is facilitated when the two verbs assign compatible features to the initial noun phrase, and made more difficult when the two verbs assign incompatible features.

The Filip et al. proposal draws on converging evidence from linguistics, psycholinguistics, and computational modelling, by embedding Dowty's Proto features within a constraint-based linguistic theory (HPSG), and modelling the ambiguity resolution data using Spivey's normalized recurrence algorithm (see Spivey et al., above).

Altmann, "Predicting Thematic Role Assignments in Context" Altmann also develops a proposal involving finer-grained semantic information to define thematic relations, and an emphasis on probabilistic use of this information. In contrast to Filip et al., however, his proposal for thematic roles is that they are verb specific selectional restrictions based on real world knowledge, thus moving even further from the standard linguistic theoretic notion of thematic roles. His primary claim is that discourse context establishes probabilistic relationships that serve as predictions concerning the entities which will play a role in the predicate of a subsequent verb. That is, the sentence processor *at the verb* (and therefore, in English, before the object position) tries to assign the verb's object thematic role to an entity already in the discourse. In support of this, Altmann finds that people experience an anomaly at a verb such as "injured" when it follows a context in which nothing is "injurable".

Altmann offers two hypotheses concerning how the sentence processor encodes "probabilistic contingencies" that represent relations particular to a lexical item. One suggestion is that the processor projects empty structure at the verb corresponding to its object thematic role, and then attempts to link this, following the attendant thematic (selectional) restrictions, with a prior entity in the discourse. Under this view, anomaly detection occurs when the processor attempts to anaphorically link the empty object position to a prior entity and finds no suitable antecedent. An alternative hypothesis is that the prior nouns in the discourse restrict the range of expected verbs to those which can fill a role with one of the entities. Anomaly detection in this scenario occurs when earlier entities have activated possible verbs of which they could be arguments and the current verb is not one of them. Altmann discusses the implications of the experimental data and these possible processing explanations within a connectionist framework for sentence processing.

Argaman and Pearlmutter, "Lexical Semantics as a Basis for Argument Structure Frequency Biases" In their chapter, Argaman and Pearlmutter shift the focus from what is the precise nature of argument structure relations in a verb's lexical entry, to what determines the argument structure frequency biases. Following Pinker (1989) and Levin (1993), they assume that primitive meaning components licence particular argument-taking properties of predicates. Since, under this view, argument structures map to "partial semantic representations", they note that argument structure selection is essentially sense disambiguation. Then, by analogy with frequency effects in word sense disambiguation, argument structure disambiguation involves frequency biases that are associated with the lexical semantic meaning components which license the argument structures. Just as meaning frequencies for ambiguous words are determined by properties of the world (the frequency of the things they refer to), so too are argument structure frequencies under Argaman and Pearlmutter's view. They claim then that argument structure biases can be independently determined through a theory of real world semantics.

The specific hypothesis investigated in the chapter is that words closely related in meaning will have similar argument structure frequencies. To determine this, Argaman and Pearlmutter compare a set of verbs and their derived nouns, and find a highly correlated bias between the two for a particular complement type (the sentential complement), in both corpus and completion studies. Conversely, a comparison of the sentential complement bias across a group of Levin verb classes reveals a significant difference. The question remains whether finer-grained differences (within-class semantic distinctions) also influence frequency biases. They found partial support for this hypothesis in the form of marginally significant correlations between verbs and their derived nouns within a class.

Argaman and Pearlmutter conclude that, while preliminary, their results provide clear evidence for a connection between frequency biases of argument structures and their underlying semantic basis.

Roland and Jurafsky, "Verb Sense and Verb Subcategorization Probabilities" Roland and Jurafsky set out to explore the issue of frequency bias for lexically determined structural information, such as subcategorization, from a different and complementary point of view to that of Argaman and Pearlmutter. Specifically, they explore the extra-linguistic factors that underlie variations in subcategorization frequencies, in an attempt to distinguish differences attributable to the sense of a verb, from differences due to the modality or style of usage.

Using a comparative method of corpus analysis, they examine the differential frequencies of subcategorization frames of selected verbs across modalities, in written and oral corpora. They further compare elicited to spontaneous production for written corpora, and also balanced corpora to simple text collections (hence unbalanced). Interestingly, they notice that while the difference between experimental data (elicited language) and corpora (spontaneous language) is large, the difference between corpora in the same modality, whether balanced or not, is not significant, once controlled for sense.

By factoring out extra-linguistics properties, Roland and Jurafsky confirm the hypothesis that subcategorization probabilities vary according to word senses, once the context effects have been taken into account. This conclusion converges with that of Argaman and Pearlmutter. Together, these two papers provide evidence that the sense of a word is the grain at which frequency is expressed, and that therefore the underlying events that give rise to frequency differentials must be found in lexical semantic primitives.

3 Discussion of Cross-cutting Issues

Although we have divided the contributions into sections of the book that emphasize the primary focus of the individual chapters, many of the papers address issues across these boundaries, and there are numerous points of contact between papers in different sections. Here we highlight three major topics that run through the papers, which capture the primary issues in current lexicalized theories of language and language processing: the organizational basis of the lexicon, the role of statistical information in lexical entries and processing, and the impact of lexicalization on incremental processing algorithms.

3.1 Lexical Organisation

A number of issues cut across the chapters concerning the fundamental nature of the lexicon: how it is organized and how that organization influences processing. Here we first discuss the nature and role of lexical classes, and their basis in semantics. Then we turn to the timing of structural versus semantic information in on-line interpretation, and how they interact. Finally, we address the basic character of the lexicon itself—as a static set of pre-stored tree structures versus a dynamically generated set of structures created from simple universal primitives.

3.1.1 The Role of Lexical Classes

The lexicon is not simply a list of irregularities. Each word in the lexicon is the bearer of unique, idiosyncratic information, but also of information similar to that contained in the lexical entries of many other words. If the regularities and underlying organisation of the lexicon are not taken into account, lexicalist approaches are prone to be redundant. To avoid such redundancies, lexical theories have relied on the notion of a lexical class as a means for capturing regularities. In the current volume this issue is reflected in several chapters that address the problems of determining precisely how (or even whether) classes are defined, and what role they serve in processing.

Filip et al. focus on the issue of whether verb classes are organized on the basis of structural or semantic properties. Not only do they claim that the structural distinctions between classes proposed by Stevenson and Merlo (1997) are unnecessary, they further argue that a categorical view of argument structure representations is insufficient to fully account for data on ambiguity resolution. Thus, they elaborate a view of verb class distinctions as semantic, not syntactic, and as graded, not categorical. Spivey et al. reinforce this latter point, in stressing that behaviour that appears categorical may result from the interaction of continuously weighted constraints. From this perspective, verb classes are (possibly overlapping) fuzzy sets.

Like Filip et al., Argaman and Pearlmutter also view verb classes as semantically defined, but assume a more discrete view of classes. Specifically, they adopt the approach of Pinker and Levin, in which verbs in a class share primitive meaning components that license particular argument-taking properties. Since, under Argaman and Pearlmutter's proposal, these meaning primitives serve as a site for frequency counts, the class of a verb plays a direct and observable role in processing phenomena. Argaman and Pearlmutter also put forth some preliminary evidence suggesting that frequency differentials arise from within-class meaning distinctions as well. Thus, under their view, semantic regularities lead to coarse-grained commonalities in behaviour across verbs within a class, while finer-grained distinctions in meaning lead to correspondingly finer-grained differences in behaviour among those verbs.

3.1.2 Timing of Different Information Types

In addition to the issue of which type of lexical information (syntactic or semantic) underlies particular processing effects, there is also the issue of the relative timing of the information in on-line interpretation. In contrast to previous sentence processing models, lexicalist approaches raise the possibility of having both syntactic and semantic information arise from a common source, with both playing an immediate role in processing. This view is different from more traditional structure-based models, in which syntactic information is maintained separately from semantics, and comes into play much faster. The lexicalist view is also different from interactive, staged models in which syntactic and semantic information can influence each other, but the separation of different types of information into distinct levels is clear.

With respect to this issue, Filip et al. and Spivey et al. take an approach similar to a number of lexicalist constraint-based theories, in which both types of information are available from the outset. The general approach accounts for effects in ambiguity resolution with separate but interacting frequencies corresponding to either syntactic or semantic information.

Two other chapters attempt a finer grained elucidation of the relationship between linguistic and real world information, and appear to come to contradictory conclusions. Mauner et al. claim to find an earlier influence of linguistic information (knowledge about argument structure), compared to the later influence of general semantic plausibility. Altmann, on the other hand, finds that real world constraints come into play immediately. However, the two approaches are probing somewhat different aspects of the use of fine-grained lexical information. Mauner et al. test for the immediate use of linguistically-specified arguments, as opposed to plausible (non-argument) participants. Altmann, by contrast, tests whether plausibility plays an immediate role when processing a linguistically-specified argument, and finds that it does so. In combining these two insights, it seems that a more complex relationship must be elaborated, in which real world knowledge may have an early influence if it is closely associated with a linguistic argument (as is the case with selectional restrictions).

Argaman and Pearlmutter, from a very different perspective, provide additional evidence that linguistic and real world information are closely linked through the relation between semantics and argument structure. Thus, it may be the case that syntactic and semantic information not only arise from a common source (a lexical item or its class), but are tightly connected to each other in a way that directly influences processing.

What all of these chapters demonstrate more generally is that lexicalist theories of processing may involve very fine-grained interactions of different levels of information, requiring more sophisticated experimental methodologies and analyses to elucidate them. Spivey et al. point to one particular methodological difficulty, that of using claims about timing to determine the primacy of some particular organizational principle, such as structural information. Specifically, they propose that observed differences between syntactic and semantic factors in processing depend on their initial and accrued strength, and not on a distinction in availability. Thus, evidence for differences in staging of information during processing must be carefully examined, and new methods developed for distinguishing between availability of information and its strength.

3.1.3 Static vs. Dynamic Organization of the Lexicon

Under a lexicalist constraint-based account, much observable behaviour is viewed as the result of competition among constraints of varying strengths during on-line processing. Recently, lexical theories in theoretical linguistics have extended the role of competition to determining not only on-line behaviour, but the content of the lexicon itself. What we think of as the lexicon of words or morphemes for a particular language may not be a pre-existing database of linguistic facts (whether organized by classes or not), but rather the result of a competitive process over universal primitives.

This view is espoused in Bresnan's chapter, in which the lexicon is proposed to be the result of syntactic competition. In her Optimality Theory account, words can compete with entire phrases and structures to be the optimal expression of an underlying input, so that words and phrases lie within the same level of representation and are subject to the very same (syntactic) operations. At least at the level of the competence theory, it is not only syntactic structure that is generated when a sentence is formed, but the lexical entries themselves as well. The pre-existing lexicon consists of a universal set of primitives, from which the competitive process for an input selects the optimal combination for a particular language (Prince and Smolensky 1993).

At first blush, this contrasts starkly with the view in sentence processing, where the notion of competition as central to interpretation has led to more elaborated structures in the lexicon (MacDonald, Pearlmutter, and Seidenberg 1994 and others). The paper by Kim et al. exemplifies this view, in which lexical competition occurs between syntactic trees within the lexicon which are activated in response to an input. In other words, in sentence processing, the notion of competition has led to extensive pre-compiled structure in the lexicon, while in theoretical linguistics, it has conversely led to less pre-existing structure—to the point where even words or morphemes result from syntactic competition, rather than being the input to it.

The differences between these views may not be as great as first appears, however. In his chapter, Johnson points out the modifications that would have to be applied to Bresnan's competence theory to render it computable (and therefore able to form the basis of a model of sentence processing). The GEN function in Optimality Theory raises similar issues to those introduced by other theories that presuppose an infinite generating function: a computational algorithm depends on at least some of the filtering constraints being interleaved with the generation of possible alternatives (compare Fong 1991; Tesar 1996). Johnson proposes to apply the phonological string as the first filter, to guarantee that the computation will be possible (see also Johnson 1989). Perhaps the difference then between the lexicalist sentence processing view and the view proposed by Bresnan is not so much one of kind as of degree—it remains for future work to determine how much information can, and must, be precompiled into lexical entries.

3.2 Frequencies and statistics

An emphasis on lexical influences and on frequency effects have gone hand-in-hand in sentence processing research. Frequency is thought to be a prototypical example of lexical information due to standard word-based frequency effects. Furthermore, the apparent ease of associating frequencies with pre-stored information in the lexicon, rather than with syntactic constructions, argues for a lexicalist view of frequency effects. An emphasis on frequency has also led to a corresponding emphasis on semantics, as we saw in previous discussion, since frequency provides an observable encoding of the exposure to the external world and its influence on the sentence processing mechanism.

Before turning to the issues involving the precise specification of lexical frequencies, their impact on processing, and their origin, we first discuss the possibility that frequencies are not an integral part of the lexicon.

3.2.1 Are there frequencies in the lexicon?

One possible view is that frequencies (of any granularity) simply do not occur in the lexicon. In his chapter, Fodor claims that such is the case, based on the assumption that lexical entries project the entirety of their content to their host. According to Fodor, then, frequency cannot be in the lexicon, as this would require that the relative frequency of the host be determined by the relative frequency of its parts, an assumption Fodor disavows.

However, probability models in computational linguistics hold precisely this assumption: any complex event is assumed to be decomposable into smaller, independent events, and the probability of the complex event is the product of the probability of the independent subevents. The independence assumptions are acknowledged to be too strong, and violated in practice, but the response is to develop more sophisticated and accurate probability models, rather than to abandon the approach. Contingent frequencies may be seen as rules of composition over the raw frequencies stored with an individual lexical item; what is required is to determine the appropriate combination algorithm.

Another view on this issue, reflected in sentence processing work, is that contingent frequencies are strengths of association *between* lexical items—i.e., contingent frequencies are not part of a lexical entry, but rather are part of the organization of the lexicon. From both of these perspectives, it is clear that Fodor's assumption that reverse compositionality cannot apply to frequencies is not a given, but rather highlights a known challenge of determining an appropriate representation and processing algorithm for lexical frequencies.

Weinberg replies to Fodor's arguments in a different vein, calling on the competence/performance distinction. Weinberg notes that frequencies do not need to be part of the representation of a lexical entry to be relevant to parsing, rather they can be part of the performance system. As such, the principles constraining lexical entries proposed by Fodor do not rule out current lexicalist theories of processing. However, Weinberg proposes a view that is not what many lexicalist proponents would endorse, we think. The distinction between competence and performance is not very clear in those approaches that are crucially based on a continuous as opposed to discrete representation. Hence, lexicalist approaches generally appear to say that frequency is part of the lexical entry. If a representation is distributed, the strength of the association of certain features is crucial to the representation itself.

If we accept that frequencies are associated to lexical items at some level of representation, then several issues must be addressed when studying the influence of frequency on processing. The primary one seems to be what the frequencies are associated with—i.e., what level and type of information carries frequency information (sense of a word, lemmas, phrases, constructions, among the many possible candidates). It is also important to study how to determine which frequency information comes into play at different points in processing, and where frequency differentials come from.

3.2.2 What do we count?

Roland and Jurafsky address the first question above of determining the lexical unit with which frequencies are associated, and they argue convincingly that the indexing unit is the individual word sense. Argaman and Pearlmutter, from a very different perspective, reinforce this view with their evidence that frequencies are associated with argument structures, which are themselves linked to semantic primitives. This result raises a practical and a theoretical problem. Practically, many experiments and data collections on lexical frequency are not based on the sense of the word, but on the lexical string, thus potentially confounding frequencies of very different word meanings and uses. Theoretically, the senses of a word are often not clearly defined; they may not even be enumerable, but rather the result of the interaction with the sentence context (Pustejovsky 1995).

If we combine these two observations to their logical conclusions, we might envisage a picture of lexical frequency as being the result of a process of syntactic analysis. If a word sense is determined by a compositional operation, and lexical frequencies are associated with senses, then lexical frequencies are the result of a compositional process. This view is similar to the one espoused by Bresnan, discussed above, in which the lexicon itself is the result of a compositional process. It also impacts on the discussion of Fodor's claim above, reinforcing the view that it is the combination of frequencies that must be addressed in a model of interpretation.

3.2.3 Frequencies in processing

If lexical frequencies are sensitive to syntactic structure and, more generally, the influence of context, then one might envision that several types of lexical frequencies are associated with the same lexical item. Moreover, the various types of frequencies may refer to increasingly larger domains or classes of information within a lexical hierarchy. Questions then arise concerning how these different levels of lexical frequencies are used in the time course of processing. When several levels of statistics come into play, it must be determined whether they do so all at the same time, or according to some predetermined ordering procedure. The first method is envisaged by proponents of distributed representations and architectures in psycholinguistics, in which interacting syntactic and semantic constraints, weighted by frequency, simultaneously determine the overall activation of an interpretation. In computational linguistics, this simultaneous, non-linear combination of frequencies is only rarely used (Henderson 2000), and simpler techniques based on linear interpolation are more current. Backing-off techniques are also used, which impose explicit ordering on the use of different grains of frequencies. These techniques have been developed to handle the problems of sparse data by supplementing the core fine-grained probabilities with coarser grained ones that play a secondary role (Katz 1987; Collins and Brooks 1995; Jelinek 1997).

Several positions concerning the timing of frequency information are illustrated in the current volume. Kim et al. propose a model where several levels of frequency are activated at the same time, and the appropriate level of specificity is found automatically. Crocker and

Corley argue against this view and propose a more traditional architecture where frequency information is exclusively related to lexical items and encapsulated in a lexical preprocessor. Argaman and Pearlmutter argue that frequency is a property of partial semantic components that make up the meaning of a word and that hold across boundaries of syntactic categories. However, while Argaman and Pearlmutter assert that processing behaviour is correlated with argument structure frequencies, they leave open the precise role and timing of such information. Specifically they note that it isn't yet known whether it is stored frequencies that influence sentence processing, or the underlying semantic representations themselves that are directly at work on-line.

3.2.4 The origin of frequencies

By grounding frequency in real world semantics, the proposal by Argaman and Pearlmutter also addresses the problem of the origin of frequencies. They delimit the space of possibilities as including non-causal and causal explanations. In one case, frequency is an accident, a random variation that has reached larger proportions over time (as in Tabor (1995)). Argaman and Pearlmutter adopt a different view, in which frequency is the effect of an underlying cause that accounts for similarities and differences in individual frequencies. According to Argaman and Pearlmutter, this underlying cause is the salience of objects in the world that the words refer to.

A connection between real-world salience and frequency is not a new idea for the explanation of word sense frequencies (e.g., the use of the word *bank* as an institution is more frequent than *bank* as the edge of a river because the former are more commonly talked about today). What is novel in the Argaman and Pearlmutter proposal is that frequencies of more structural notions such as argument structure also directly reflect differences in the world. This is an interesting proposition, that connects some features of the world, or our knowledge of the world, directly to our linguistic behaviour. Given the connection between argument structure and syntax, the position is rather radical.

A more indirect relation is usually assumed in order to explain a certain arbitrariness in the lexicalisation and grammaticalisation of real world knowledge. In particular, a relation between (low) complexity on the one hand, and (high) frequency and (wide) typological distribution on the other, is a widely attested phenomenon, captured by the notion of linguistic *markedness* (Moravcsik and Wirth 1983). An instantiation, and in part an explanation, for the relationship betwen complexity, frequency, and cross-linguistic variation is illustrated in Bresnan's paper. In her account, more complex structures violate more grammatical constraints, and therefore surface less frequently. Thus Bresnan's account includes an intra-linguistic component to frequency differentials, which are not exclusively a function of salience of referents in the external world.

3.3 Incrementality

The increasing emphasis on lexical information brings a corresponding emphasis on a-word-at-a-time processing. In sentence processing, this is due to the observed rapidity of interpretation (which in standard views requires a connected parse); in computational linguistics, this is due to the need for efficiency and support for semantic interpretation (Charniak 1997;

Roark and Johnson 1999; Brants and Crocker 2000; Sturt, Lombardo, Costa, and Frasconi 2001). The chapters here investigate the degree to which incremental word-based processing is possible and conducive to interpretation, and the effects on incremental processing of having enriched lexical entries.

3.3.1 The Limits of Incrementality

Lombardo and Sturt show that while a large amount of lexically projected syntactic structure-building can be performed incrementally, some attachments cannot be resolved in a fully lexically projected and incremental approach. In this context, Steedman's implicit reminder of the difference between lexical semantics and sentential semantics help situate the claims of lexicalist approaches with respect to parsing proper. Steedman's remarks apply both to simple recurrent network architectures and to recent lexicalist approaches to parsing (Kim et al, this volume). Any approach equivalent to part-of-speech tagging will leave some structural attachments undone, and is therefore not supportive of full semantic interpretation, for which a fully connected structure is required. Note too the empirical support from Stowe et al. that suggests the existence of a functional area of the brain for structure-building as opposed to lexical processing.

These points indicate that structure building beyond lexicalist projection must be accomplished if one is to build a complete interpretation. However, recent research, both in psycholinguistics and computational linguistics, has called into question the completeness assumption in parsing. Work on reanalysis of garden-path sentences has shown that comprehenders often end up with an interpretation that is based on both the initially incorrect and ultimately correct structures (Christianson, Hollingworth, Halliwell, and Ferreira 2001). These results are interpreted as suggesting that alternative syntactic analyses may not be completely constructed; rather, the reanalysis mechanism may be satisfied with a "good enough" parse. A different way of relaxing the requirements for a full parse is exemplified in the paper by Kim et al. in this volume, and similar approaches have been suggested in the computational parsing literature (Abney 1996). Here a large amount of lexical projection is performed, which gives rise to a partial parse, where fragments of the entire structure are constructed. Some difficult structure building decisions are left undone (PP attachment is a typical case), under the assumption that they will be resolved at later stages by knowledge based on pragmatics, or lexical associations. Both these lines of proposals raise interesting questions regarding exactly what constitutes "an interpretation," and the precise interaction required between lexically-driven and discourse-driven processes.

3.3.2 The Influence of Rich Lexical Information

The chapters by Mauner et al. and Altmann expand the view of incremental interpretation, by proposing rich lexical information that licenses early postulation of hypothesized entities. These empty elements play a central role in on-line interpretation. Mauner et al. provide evidence that linguistically-licensed entities (e.g., arguments to a verb) influence interpretation even when those entities are not expressed in the sentence. Altmann suggests further that real-world properties of such arguments can affect subsequent integration of words in the input. Both of these proposals go far beyond the early establishment of an empty element in the syntactic representation of an input, as proposed in previous models of sentence processing (e.g., Crocker 1995; Gibson and Hickok 1993; Stevenson 1993; Stevenson 1995). In the Mauner et al. view, fine-grained thematic properties of empty discourse elements play an early role in interpretation. According to Altmann, this role extends to the incremental computation of expectations concerning the real-world properties of entities and events in the input. In both cases, the representation of empty elements is more sophisticated than previously assumed, and thus their role in incremental processing is potentially more complex and influential.

4 Methodological Concerns

Many of the chapters use computational methods to investigate psycholinguistic questions, either by modelling experimental results or by investigating corpus data. Both types of approaches raise important methodological issues in the study of human language processing, and increase the connections between work in psycholinguistics and computational linguistics. Other chapters take a more traditional human experimental approach, but expand the repertoire of experimental methodologies in order to address the novel questions raised by a lexicalist view. We discuss the import of each of these methodological insights here.

4.1 Computational Modelling

Building a computational model is a complex business, as computer programs can be interpreted at two levels of abstraction, usually referred to as the representational level and the algorithmic level (Marr 1982; see also the discussion in Brent 1996). These different perspectives are concerned with what is computed and how it is computed, respectively. For a particular set of observational data from human experiments, there could be many underlying representational schemes, each of which has many possible ways of being computed. The consequence is that experimental data generally underspecify the set of possible computational models. The chapter by Spivey et al. addresses this issue in detail, with a concrete demonstration that certain experimental evidence can be explained equally well by two different kinds of models founded on very different representational schemes and algorithms.

The under-specification of models by the data leads to a range of approaches to computational modeling illustrated in the chapters here, which lie along a gradient of specificity, or degree of abstractness. Starting from a full implementation, one can propose a very detailed model, such as the "almost" parsing model of Kim et al. By assuming a specific computational architecture, this approach provides theoretical justification for hypotheses for which the empirical evidence is insufficient. One can also propose models more limited in scope, such as that of Crocker and Corley. They develop a detailed model of disambiguation, situated in a well-understood, although not implemented, parsing framework. An even greater degree of abstraction can also profitably be adopted, as seen in the chapters by Filip et al. and Spivey et al. Each of these papers uses a model with a completely underspecified parsing framework, in order to highlight very specific aspects of the competition process, and the influential constraints, in disambiguation.

4.2 Corpus-Based Investigations

The availability of large collections of annotated (part-of-speech-tagged or parsed) text has recently introduced a new opportunity to explore even more abstract computational models. The paper by Lombardo and Sturt exemplifies this methodology, by developing a very abstract model of "possible parsers," founded on representations derived from corpora. Corpora also support the investigation of models of lexical information without any explicit relation to parsing, as seen in the papers by Roland and Jurafsky, and Argaman and Pearlmutter.

More generally, these papers underscore the fact that annotated corpora are implicit repositories of grammars (Merlo and Stevenson 1998). In this regard, corpora go beyond idealized grammatical knowledge, and serve as an approximation to a speaker's linguistic experience, containing important frequency information. It has become common practice in psycholinguistic approaches, illustrated in numerous chapters here (e.g., Argaman and Pearlmutter, Filip et al., Kim et al., Spivey et al.), to use frequency data collected from corpora as representative of a speaker's knowledge. More helpful still will be the sophisticated lexicalized grammars currently being developed by computational linguists through automatic extraction from parsed corpora (e.g., Xia, Palmer, and Joshi 2000). The result of such efforts would be grammars that incorporate statistics over usage, providing the integrated grammatical and statistical knowledge needed to evaluate sentence processing proposals that rely on lexicalized frequency effects.

It is important to note, though, that as an approximation to an actual linguistic experience, data from a corpus must be confirmed through analysis and comparison to actual behavioural and linguistic studies (Gibson and Pearlmutter 1994; Merlo 1994; Roland and Jurafsky 1998; Lapata, Keller, and Schulte im Walde 2001). The chapter here by Roland and Jurafsky shows that such studies can both evaluate properties of experimental stimuli, and lead to preliminary hypotheses which can then be tested experimentally. From a practical point of view, the paper by Roland and Jurafsky provides very useful evidence on the materials and methods needed to estimate subcategorization and argument structure frequencies, a crucial problem in developing sophisticated probability models in both psycholinguistics and computational linguistics.

4.3 Experimental Advances

An emphasis on lexical information has similarly led to innovations in experimental methodologies as well, illustrated in several of the chapters here. For example, motivated by fine-grained predictions from a lexicalist constraint-based perspective, Spivey et al. introduce a very interesting new methodology, called speeded sentence completion. This technique allows them to access representations at different points in on-line processing, by eliciting completions of a sentence fragment after differing time delays. This type of data is needed to support or disprove the claims that multiple interpretations are simultaneously, but differentially, activated, as factors of differing weights compete over time.

Both the Mauner et al. and Altmann chapters extend experimental approaches to determine the role of lexically-specified information about arguments that are not (or not yet) explicitly present in the input. Both add to the repertoire of experimental methods for eliciting information about the early use of argument structure in sentence processing. Mauner et al. detail methods for detecting elements of argument structure (the semantic arguments of a verb) independently of subcategorization (the syntactic expression of the arguments). Their techniques are also independent of the plausibility of the argument, allowing for arguments to be detected at the verb itself. Altmann also seeks to elucidate the immediate role of argument expectations in processing, using a method of anomaly detection at the verb to elicit responses when no previously introduced entities are compatible with the selectional restrictions on its object (what he terms the object's "thematic role"). These selectional restrictions involve general semantic "fit" with the verb, and thus go beyond the purely linguistic information suggested by Mauner et al.

The chapter by Stowe et al. demonstrates the need for additional neuro-imaging data and techniques to help constrain possible models of sentence processing. Their results indicate a complex relationship between functional areas of the brain, and the division of sentence processing labor into encoding, storage and processing. Approaches that equate lexical and syntactic (phrase-level) processing, a common assumption in lexicalist theories, initially appear incompatible with the evidence from Stowe et al. As we have noted at several key points of discussion above, a lexicalist approach raises many new interesting questions concerning the precise representation of different information types, and the nature of their interaction. Advances in neuro-imaging studies will ultimately be required to elicit the fine-grained data needed to distinguish among the logical possibilities those that are compatible with the functional architecture of the brain.

5 Conclusions

The contributions to this volume illustrate the wide range of issues that arise as a consequence of the increased role that the lexicon plays in current theories of syntax, of parsing and of sentence processing. The findings here reveal the complexity of both representations and algorithms required in theories of language that capture the richness of lexical information and its interaction with syntax and structure-building. In dealing with this complexity, researchers face a tension between the descriptive need to represent the full range of lexical variability in language, and the conceptual need to explain the regularities and organisation of the lexicon. We think that the accumulated evidence that lexical effects are strong ("it is all in the words") can be reconciled with the theoretical needs for generalisation and succinctness by further exploring the notion of *classes* of words. The investigation of the notion of class promises to be informative to some of the common concerns that have appeared across many of the papers in the volume.

A class structure for lexical items implicitly assumes that the lexicon is organised, since it imposes regularity on lexical variability. Studying the potential principles that underlie lexical organisation is important practically, as a means of reducing redundancy and rendering lexicalised approaches manageable, and is also important conceptually, as it highlights regularities and generalisations (Daelemans, Smedt, and Gazdar 1992; Briscoe, Copestake, and de Paiva 1994). Recent proposals in computational linguistics for automatic verb classification have investigated classifications based on both syntactic and semantic information, such as subcategorisation (Xia, Palmer, and Joshi 2000), argument structure (Merlo and Stevenson 2001), Levin's classes (Lapata and Brew 1999), and finer-grained classes than Levin's

(Dang, Kipper, Palmer, and Rosenzweig 1998). A possibility unifying these approaches is that these different types of classes correspond to different levels in a hierarchical lexicon, which simultaneously captures generalizations at different levels of abstraction (Palmer 2000; Merlo and Stevenson 2001).

The notion of lexical class further provides the conceptual locus to integrate symbolic linguistic notions and probabilistic concepts. For example, the recent work in computational linguistics on the verbal lexicon has shown that there are pervasive regularities in statistical distributions of verbs belonging to the same semantic class. These statistical regularities are attested at several levels of granularity of lexical organisation (Lapata and Brew 1999; Merlo and Stevenson 2001). These findings lend further support to the idea that the lexicon is hierarchically organised along several levels at the same time, and extends this view by suggesting an organisation that is sensitive to frequency. This type of rich lexical organization, in terms of a frequency-informed hierarchy, supports the sophisticated probabilistic modeling techniques (using back-off and smoothing) that have been so useful in computational linguistics.

Many of the profitable definitions of classes in computational linguistics have been based on structural notions, such as subcategorisation, alternations in the expression of arguments, and argument structure. If words are systematically grouped in classes organised around structural notions, then the integration of each word in the sentence representation during parsing requires the integration of a little piece of structure. Thus, the notion of class is directly relevant to issues of incrementality in parsing, concerning what kind of information is immediately available, and the projection or prediction of such information. Furthermore, a structure-based notion of classes has the consequence that the relation between structure and frequency can be productively studied, with each investigated as the predictor of sentence processing complexity.

Finally, the systematic relationship between class and frequency can support the integration of research on processing and acquisition. Recent work on the automatic acquisition of properties of verbs has shown that even in cases where the surface syntactic representation (subcategoristion) does not distinguish between classes, the statistical differentials related to verb class properties are strong enough to generalise and enable the semantic classification of previously unseen verbs (Merlo and Stevenson 2001). If words are systematically grouped into classes organised around structural notions, then learning words is already in part learning structure. This means that the structural notions related to classes can be learnt by exposure to their frequency differentials, and then both the structures and frequencies can be used in processing (Merlo and Stevenson 1999).

Thus, the idea of a hierarchical lexicon organized according to different classes of information yields a unified framework for further exploration of the important issues raised in this volume concerning: lexical organization; the interaction between lexical, syntactic and semantic information and processing; and the role of lexical statistics in guiding the acquisition and interpretation of language. While this is just one of the possible developments in the study of the interface between the lexicon and sentence understanding, the notion of class provides a fruitful ground from which both variability and regularity in language and processing can be successfully investigated.

References

Abney, S. (1996). Partial parsing via finite-state cascades. In J. Carroll (Ed.), *Proceedings of the Workshop on Robust Parsing at the 8th Summer School on Logic, Language and Information*, Number 435 in CSRP, pp. 8 – 15. University of Sussex, Brighton.

Abney, S. (1997). Stochastic attribute-value grammars. *Computational Linguistics 23*(4), 597–618.

Black, E., F. Jelinek, J. Lafferty, D. Magerman, R. Mercer, and S. Roukos (1992). Towards history-based grammars. In *Proceedings of the 30th Meeting of the Association for Computational Linguistics*, Newark, DE, pp. 31–38.

Brants, T. (1999). Cascaded markov models. In *Proceedings of 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL-99*.

Brants, T. and M. Crocker (2000). Probabilistic parsing and psychological plausibility. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, Saarbrücken, pp. 111–117.

Brent, M. (1993). From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics 19(2)*, 243–262.

Brent, M. (1996). Advances in the computational studies of language acquisition. In M. Brent (Ed.), *Computational Approaches to Language Acquisition*, pp. 1–38. Cambridge, MA: MIT Press.

Bresnan, J. (1982). *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Brew, C. (1995). Stochastic HPSG. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics (EACL'95)*, Dublin, Ireland, pp. 83–89.

Briscoe, T. and J. Carroll (1997). Automatic extraction of subcategorization from corpora. In *Proceedings of the Fifth Applied Natural Language Processing Conference*, Washington, D.C., pp. 356–363.

Briscoe, T., A. Copestake, and V. de Paiva (Eds.) (1994). *Inheritance, Defaults and the Lexicon*. Cambridge University Press.

Charniak, E. (1996). Tree-bank grammars. In *Proceedings of Thirteenth National Conference on Artificial Intelligence (AAAI'96)*, Portland, Oregon, pp. 1031–1036.

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the 14th National Conference on AI*, Providence, Rhode Island, pp. 598–603.

Christianson, K., A. Hollingworth, J. Halliwell, and F. Ferreira (2001). Thematic roles assigned along the garden path linger. *Cognitive Pychology 42*, 368–407.

Collins, M. and J. Brooks (1995). Prepositional Phrase Attachment through a Backed-Off Model. In *Proceedings of the Third Workshop on Very Large Corpora*, Cambridge, MA, pp. 27–38.

Collins, M. J. (1996). A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA, pp. 184–191.

Collins, M. J. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain, pp. 16–23.

Cortes, C. and M. Mohri (2000). Context-free recognition with weighted automata. *Grammars 3*(2/3), 133–150.

Crocker, M. (1995). *Computational Psycholinguistics: A Interdisciplinary Perspective.* Dordrecht: Kluwer Academic Publishers.

Daelemans, W., K. D. Smedt, and G. Gazdar (Eds.) (1992). *Computational Linguistics, Special Issue on Inheritance*, Volume 18. MIT Press.

Dang, H. T., K. Kipper, M. Palmer, and J. Rosenzweig (1998). Investigating regular sense extensions based on intersective Levin classes. In *Proceedings of the 36th Annual Meeting of the ACL and the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, Montreal, Canada, pp. 293–299.

Dorr, B. (1997). Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation 12*(4), 1–55.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language 67*(3), 547–619.

Fellbaum, C. (1998). *Wordnet: an Electronic Lexical Database.* MIT Press.

Fillmore, C. (1988). The mechanisms of "construction grammar". In *Berkeley Linguistics Society*, Volume 14, pp. 35–55.

Fong, S. (1991). *Computational Properties of Principle-based Grammatical Theories.* Ph. D. thesis, MIT, Cambridge, MA.

Ford, M., J. Bresnan, and R. Kaplan (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, pp. 727–796. Cambridge: MIT Press.

Frazier, L. (1978). *On Comprehending Sentences: Syntactic Parsing Strategies.* Ph. D. thesis, University of Connecticut. Available through the Indiana University Linguistics Club, Bloomington, IN.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition 68*, 1–76.

Gibson, E. and G. Hickok (1993). Sentence processing with empty categories. *Language and Cognitive Processes 8*, 147–161.

Gibson, E. and N. J. Pearlmutter (1994). A corpus-based analysis of psycholinguistic constraints on prepositional phrase attachment. In C. Clifton, L. Frazier, and K. Rayner (Eds.), *Perspectives on Sentence Processing*, pp. 181–198. Laurence Erlbaum.

Gibson, E., C. Schütze, and A. Salomon (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research 25*(1), 59–92.

Goldberg, A. (1995). *A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Hale, K. and J. Keyser (1993). On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser (Eds.), *The View from Building 20*, pp. 53–110. MIT Press.

Henderson, J. B. (2000). A neural network parser that handles sparse data. In *Proceedings of the 6th International Workshop on Parsing Technologies (IWPT2000)*, Trento, Italy, pp. 123–134.

Jelinek, F. (1997). *Statistical Methods for Speech Recognition*. MIT Press.

Johnson, M. (1989). The use of knowledge of language. *Journal of Psycholinguistic Research 18*(1), 105–129.

Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics 24*(4), 613–632.

Johnson, M., S. Geman, S. Canon, Z. Chi, and S. Riezler (1999). Estimators for stochastic unification-based grammars. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park,MD, pp. 535–541.

Joshi, A. and Y. Schabes (1997). Tree-adjoining grammars. In G. Rozenberg and A. Salomaa (Eds.), *Handbook of Formal Languages*, Volume 3, pp. 69–124. Berlin,New York: Springer Verlag.

Just, M. A. and P. A. Carpenter (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review 99*(1), 122–149.

Katz, S. (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustic, Speech and Signal Processing 35*(3), 400–401.

Lapata, M. and C. Brew (1999). Using subcategorization to resolve verb class ambiguity. In *Proceedings of Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, College Park, MD, pp. 266–274.

Lapata, M., F. Keller, and S. Schulte im Walde (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*.

Levin, B. (1985). Introduction. In B. Levin (Ed.), *Lexical Semantics in Review*, Number 1 in Lexicon Project Working Papers, pp. 1–62. Cambridge, MA: Centre for Cognitive Science, MIT.

Levin, B. (1993). *English Verb Classes and Alternations.* Chicago, IL: University of Chicago Press.

MacDonald, M., N. Pearlmutter, and M. Seidenberg (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review 101*(4), 676–703.

Magerman, D. and M. Marcus (1991). Pearl: A probabilistic parser. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, Berlin, Germany, pp. 15–20.

Magerman, D. and C. Weir (1992). Efficiency, robustness, and accuracy in Picky chart parsing. In *Proceedings of the 29th Meeting of the Association for Computational Linguistics*, Newark, DE, pp. 40–47.

Marcus, M., B. Santorini, and M. Marcinkiewicz (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics 19*, 313–330.

Marr, D. (1982). *Vision.* S.Francisco, CA: W.H.Freeman.

McCarthy, D. (2000). Using semantic preferences to identify verbal participation in role switching alternations. In *Proceedings of ANLP-NAACL 2000*, Seattle, WA, pp. 256–263.

Mel'cuk, I. (1988). *Dependency Syntay: Theory and Practice.* Albany, NY: State University of New York Press.

Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research 23*(6), 435–457.

Merlo, P. and S. Stevenson (1998). What grammars tell us about corpora: the case of reduced relative clauses. In *Proceedings of the Sixth Workshop on Very Large Corpora*, Montreal, CA, pp. 134–142.

Merlo, P. and S. Stevenson (1999). Language acquisition and ambiguity resolution: the role of frequency distributions. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society (CogSci'99)*, Vancouver, Canada, pp. 399–404.

Merlo, P. and S. Stevenson (2001). Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics 27*(3).

Miller, G., R. Beckwith, C. Fellbaum, D. Gross, and K. Miller (1990). Five papers on Wordnet. Technical report, Cognitive Science Lab, Princeton University.

Mitchell, D., F. Cuetos, M. Corley, and M. Brysbaert (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research 24*(6), 469–488.

Moravcsik, E. and J. Wirth (1983). Markedness – an Overview. In F. Eckman, E. Moravcsik, and J. Wirth (Eds.), *Markedness*, pp. 1–13. New York,NY: Plenum Press.

Palmer, M. (2000). Consistent criteria for sense distinctions. *Special Issue of Computers and the Humanities, SENSEVAL98: Evaluating Word Sense Disambiguation Systems 34*(1–2), 217–222.

Pinker, S. (1989). *Learnability and Cognition: the Acquisition of Argument Structure*. MIT Press.

Pollard, C. and I. Sag (1987). *An Information-based Syntax and Semantics*, Volume 13. Stanford University: CSLI lecture Notes.

Prince, A. and P. Smolensky (1993). Optimality Theory: Constraint interaction in generative grammar. Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University.

Pritchett, B. (1992). *Grammatical Competence and Parsing Performance*. Chicago, IL: University of Chicago Press.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Ratnaparkhi, A. (1997). A linear observed time statistical parser based on maximum entropy models. In *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*, Providence, RI, pp. 1–10.

Ratnaparkhi, A. (1999). Learning to parse natural language with maximum entropy models. *Machine Learning 34*, 151–175.

Resnik, P. (1992). Probabilistic tree-adjoining grammars as a framework for statistical natural language processing. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, Nantes, France, pp. 418–424.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*. New Jersey: World Scientific.

Roark, B. and M. Johnson (1999). Efficient probabilistic top-down and left-corner parsing. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, MD, pp. 421–428.

Roland, D. and D. Jurafsky (1998). How verb subcategorization frequencies are affected by corpus choice. In *Proceedings of the 36th Annual Meeting of the ACL and the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, Montreal, Canada, pp. 1122–1128.

Schabes, Y. (1991). *Mathematical and Computational Aspects of Lexicalized Grammars.* Ph. D. thesis, University of Pennsylvania.

Schabes, Y. (1992). Stochastic lexicalized tree-adjoining grammars. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, Nantes, France, pp. 424–432.

Schulte im Walde, S. (2000). Clustering verbs semantically according to their alternation behaviour. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, Saarbruecken, Germany, pp. 747–753.

Siegel, E. and K. McKeown (2001). Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics 26*(4), 595–628.

Spivey, M. and M. Tanenhaus (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition 24*, 1521–1543.

Spivey-Knowlton, M. (1996). *Integration of visual and linguistic information: Human data and model simulations.* Ph. D. thesis, University of Rochester.

Spivey-Knowlton, M. and J. C. Sedivy (1995). Resolving attachment ambiguity with multiple constraints. *Cognition 55*(3), 227–267.

Spivey-Knowlton, M., J. Trueswell, and M. Tanenhaus (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology 47*(2), 276–309.

Srinivas, B. and A. K. Joshi (1999). Supertagging: An approach to almost parsing. *Computational Linguistics 25*(2), 237–265.

Stevenson, S. (1993). Establishing long-distance dependencies in a hybrid network model of human parsing. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pp. 982–987.

Stevenson, S. (1995). Arguments and adjuncts: A computational explanation of asymmetries in attachment preferences. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pp. 748–753.

Stevenson, S. and P. Merlo (1997). Lexical structure and processing complexity. *Language and Cognitive Processes 12*(1-2), 349–399.

Stevenson, S. and P. Merlo (2000). Automatic lexical acquisition based on statistical distributions. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, pp. 815–821. Saarbruecken, Germany.

Sturt, P., V. Lombardo, F. Costa, and P. Frasconi (2001). A wide-coverage model of first-pass structural preferences in human parsing. In *Paper presented at the 14th CUNY Conference on Human Sentence Processing*, University of Pennsylvania, pp. 23.

Tabor, W. (1995). Lexical change as nonlinear interpolation. In J. D. Moore and J. F. Lehman (Eds.), *Proceedings of the 17th Annual Cognitive Science Conference*, Hillsdale,NJ. Lawrence Erlbaum Associates.

Tabor, W. and M. K. Tanenhaus (1999). Dynamical models of sentence processing. *Cognitive Science 23*(4), 491–515.

Tesar, B. (1996). Computing optimal descriptions for Optimality Theory grammars with context-free position structures. In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, CA, pp. 101–107.

Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language 35*, 566–585.

Xia, F., M. Palmer, and A. Joshi (2000). A uniform method of grammar extraction and its applications. In *Proceedings of the Fifth Conference on Empirical Methods in NLP (EMNLP-2000)*, Hong Kong.