

Book Reviews

Karen Sparck Jones and Julia Galliers, *Evaluating Natural Language Processing Systems: An Analysis and Review*. Berlin: Springer-Verlag, 1996. ISBN 3 540 61309 9, Price DM54.00 (paperback). 228 pages.

Evaluation, like democracy, is one of those things which it is impossible to be against. The problem is, what does it mean?

In the context of software systems, and more specifically NLP systems, evaluation means radically different things depending on whether you adopt the perspective of, say, a business attempting to choose between machine translation systems, or a researcher attempting to extend the coverage of a parser, or a funder trying to decide whether R&D money has been well spent. Even identifying the system to be evaluated can pose serious challenges: NLP application systems are almost always hybrid systems containing significant amounts of non-NLP software with which the NLP software may interact in complex ways (for example, the knowledge base of an expert system) and these interdependencies make the assignment of merit and fault to individual components an extremely sticky business. Further, we must decide whether we are interested in how well an NLP (sub)system performs with respect to its own internal design goals or objectives, or how well it performs with respect to the larger system or setup within which it is embedded. Then, are we interested in how well the system performs relative to some idealised notion of correct performance (but, what, for example, is a correct dialogue?), or relative to human performance at the task (but, some tasks humans cannot realistically perform – such as extracting information from gigabytes of text on the WWW), or relative to other systems which purport to do the same thing (but, comparing two poor systems may tell us little)? Will our evaluation allow us to make generalisations about the techniques employed or support predictions about application in new domains? As if these worries were not enough, evaluation is also costly and time consuming – so getting it right is of significant practical concern. Nevertheless, despite all of these difficulties, we must strive to obtain some analytical purchase on evaluating complex software artefacts such as NLP systems: otherwise we are condemned to arbitrary choices between competing systems in the marketplace and stagnation or reinvention in research.

Sparck Jones and Galliers' new book is, as its subtitle suggests, an analysis of the concept of evaluation in the context of NLP systems followed by a review and assessment of what work has been done to date; it ends with the authors' own modest proposals for how evaluations ought to be designed, illustrated by means of an extended, hypothetical example.

The first third of the book sets forth a conceptual framework for addressing evaluation issues in relation to NLP systems. This requires first attempting to delimit just what will be meant by 'natural language' (in contrast with 'sub', 'pseudo' and 'quasi' languages), by NL 'processing' (nothing as trivial as keyword indexing), by 'systems' (contrast 'subsystems', 'components', 'l-system' (language (sub)system), 'n-system' (non-language (sub)system)) and by 'setups' (system plus its operational context, including users). A set of features by which systems and setups are to be described is introduced.

Turning to evaluation, the authors propose that broad notions of evaluation 'criteria' need to be made specific through 'measures' and 'test methods'. These can be applied both to systems and to setups. Criteria applied to systems or setups are either 'extrinsic' or 'intrinsic'

depending on whether or not they address the system or setup's 'function' within its embedding environment, or merely with respect to its own 'objectives'. To carry out a thorough evaluation we must identify the 'performance factors' that will affect a system or setup's behaviour and investigate systematically the effects of altering them. Performance factors divide into 'environment factors' and 'system factors' and are specified respectively via 'environment variables' and their corresponding 'values' and 'system parameters' and their corresponding 'settings'.

To illustrate these concepts an extended example is used of a NL database query system for a manufacturing enterprise. Various setups and systems can be identified here, depending, for example, on whether one takes the whole database system including the NL query subsystem to be the system, and whether one takes all of the varying company operations (warehouse, workflow, finance) to be part of the setup. Consider the limited setup of NL queries issued from the workflow. The purpose of this setup is to satisfy information needs, provide a timely response, etc. for users on the workflow. An environment variable for this setup is database correctness – something determined external to the setup, but whose value may affect the setup significantly. The system within this setup is the DBMS with NL query interface only (i.e. no database update), comprising an l-subsystem (the NL interface) and an n-subsystem (data retrieval). Environment variables for the system will be various characteristics of the users, such as proneness to spelling mistakes, frequency of use; system parameters could be such things as range of NL queries supported and output format. Extrinsic criteria for the system would include, e.g. success in meeting data needs, while intrinsic criteria would include appropriateness of response (that is, given a query, is the answer as the system is intended to produce it? – at this level reference to the user's information need must be excluded and the database must be assumed correct). A suitable measure for this intrinsic criterion would be percentage of queries answered appropriately; a suitable methodology would involve specifying a representative sampling of queries from the user population.

No doubt this clearing of this conceptual undergrowth is necessary, but it makes for hard sloggng. The authors have provided a glossary at the front of the book for the 200-odd terms they introduce for use in a semi-formal way. This eases the burden somewhat, but be prepared for a severe strain on working memory and some doubts over whether at points the ramification of distinctions is carried too far (e.g. users have 'roles', 'categories', 'aims', 'activities' and 'behaviours'; evaluations have 'kinds' and 'types', 'orientations', 'forms', 'styles' and 'modes').

The next half of the book discusses the actual state of the art in NLP evaluation. The material here is divided into two sections. The first deals with task-specific evaluations, examining in detail the three main areas of organised evaluation so far – machine translation (MT), message understanding (MU) or information extraction (IE), and NL front ends to databases – and then, in less detail, spoken language understanding, miscellaneous task evaluations (summarisation and categorisation) and text retrieval. The second section addresses general developments in the field. These include recent workshops (including the US DARPA conferences – MUC, SLS, ATIS, Machine Translation, TIPSTER and TREC), recent tutorials on evaluation, the work of the EAGLES standards group in Europe, proposals for methods such as 'Wizard of Oz' simulations, various existing test corpora/suites/collections, the prospects for generic NLP system evaluation (generic NLP systems are task-dependent but application (domain) independent, e.g. an NL front end for databases in any subject domain), and the prospects for multi-task, or 'mega-evaluations'.

This part of the book will be extremely valuable for those who are interested in getting a broad and current perspective on actual evaluation activity across the whole NLP spectrum. It contains many valuable references to material otherwise difficult to trace – for much of this work lives in technical reports of various institutions and is difficult to find without such a compendium.

The final and briefest chapter of the book returns to the author's analytical framework of the first chapter and presents their recommendations for how evaluations ought to be carried

out. They summarise these as (1) evaluations should be designed by an ‘unpacking’ process which successively decomposes the subject of the evaluation and requires at each stage of the decomposition the answering of a set of questions (which they supply), and (2) evaluations should be viewed as a set of related component evaluations based on a performance factor grid. These general recommendations are illustrated by reference to a hypothetical application which is intended to tutor architecture students on house design and incorporates an NL interface. They consider five different evaluation cases, involving different setups and systems, subsystems or components. These demonstrate very well the overall approach they recommend: establishing the evaluation remit (motivation, goals, etc.), analysing the system and its setup, defining (sub)setup/(sub)system boundaries, identifying setup purposes and system objectives, determining performance factors, specifying extrinsic and intrinsic evaluation criteria, translating these criteria into measures, collecting evaluation data and devising an evaluation procedure. Given such a design one then simply executes the evaluation procedure, systematically varying variables and parameters, and records the results in a performance factor grid.

The authors conclude with a slogan which they suggest is an adequate conclusion for the whole book: “in evaluation it is always essential to look at environment factors”. The gloss being that while NLP evaluation is fine as such it is of limited value – what matters is the setup.

While sympathetic with this view, I believe it does raise several points for discussion. First, there may be NLP ‘applications’ (such as IE) which because of their novelty do not, as yet, have a realistic setup. At a number of places in the book the authors cite IE, and particularly the DARPA MUC evaluations, as somewhat artificial because of this lack of setup. No doubt this is true; but, while we must be aware of the dangers of unconstrained technology-driven research, too much insistence on evaluation within a context of ‘real-world’ environment may stifle the genuinely novel. Translation, by contrast, is something which is and has been done regularly by humans without machine assistance; hence it is essential to evaluate MT systems within the human and institutional setups in which they are to be deployed, and not merely on their own terms. But IE is an activity which does not have as clear a correlate as MT in existing practices. Of course, this may be because it is something which will never prove useful in practice, in which case the DARPA MUC exercises will, in hindsight, be viewed as a colossal waste of time. But on the other hand, it may only be by bootstrapping a technology to a certain level that it becomes a plausible candidate for inclusion in setups, and indeed may then even transform them. The Wright brother’s Flyer must not have seemed much of a threat to the transatlantic steamship business, and any evaluation based on a setup purpose of intercontinental mass transport would probably have discarded airplanes as technology of no interest.

The second point to make here is that while the performance factor grid the authors promote is the only way to gather comprehensive, reliable information about a system, there are practicalities which may prevent it being used beyond a certain limit. Suppose for a given setup that we have m environment variables and n system parameters, each of which can take, say, k values or settings. Then the number of setup variations to be tested is $k^{(m+n)}$ and each of these may need to be tested on a range of test data which needs to be defined and collected. Further, the unpackaging approach suggests that for any NL system there will likely be a number of such setups to investigate. Given the brevity of human life and the limitations of funding, the consequence is that such complete evaluations, however defensible in principle, will never be carried out on any but the simplest systems. Instead, researchers will continue to rely on intuitions about what aspects of setups/systems are likely to be playing crucial roles, and use these intuitions to focus evaluation efforts. This may lead to incorrect analyses, but the combinatorics may leave no alternative.

Interestingly, this same issue arises in designing multi-factorial experiments in all areas in which a systems perspective is appropriate. In cognitive psychology, for example, the object of study may be the individual within an environment – we might, perhaps, be interested in the effect of various teaching strategies on learning. In medical biology, we might be interested in

the effects of drug treatments on the functioning of cells within an organ. In microeconomics we may be interested in understanding the effects of the broader economy on individual firms. For each of these areas we must isolate and systematically investigate performance factors. The problem is that (1) we cannot necessarily identify in advance all factors in the system or setup that significantly affect performance; and (2) the combinatorics of the factor space may preclude exhaustive experimentation.

None of this is to take serious issue with the utility of the book or of the approach it describes. It will prove invaluable as a reference, both for pointers into the current literature on evaluation, and for its articulation of the concept space around that overworked, but underanalysed, term 'evaluation'. As the authors say, there are no magic bullets for NLP evaluation; but my advice is, if you set out to do it, start here.

ROBERT GAIZAUSKAS
University of Sheffield, UK

email: R.Gaizauskas@dcs.shef.ac.uk

Lila Gleitman and Barbara Landau, *The Acquisition of the Lexicon*. MIT Press, 1994. ISBN: 9 780262 571099, Price: £33.95.

How do children acquire lexical knowledge? They are not usually presented with negative evidence (which makes learning easier), the speech signal is often noisy, and as is well known, the general induction problem of language is intractable. Even with these hurdles, most children acquire large numbers of words with relative ease. Quite apart from research in human cognition, insight into how children so effortlessly acquire their mental lexicon would undoubtedly help us to build large-scale lexica for natural language processing systems. This volume, consisting of a series of papers, examines how children acquire their mental lexicon. The book is organized into six sections, as follows.

The first section ('Nature of the mental lexicon') is concerned with generalities of lexical learning. Edwin William's paper argues that because of lexical phrases, 'linguistic and categorization principles are not enough' for successful acquisition. Levin and Hovav present a mainstream linguistic paper on causative alternation verbs. The argument for inclusion of this paper in the volume is that it highlights the complexity of lexical items.

We next have a section ('Discovering word units'), with a paper by Anne Cutler discussing how prosody helps segment the speech signal into discrete entities, and a paper by Kelly and Martin commenting on the implications of animal learning for language learning.

The third section 'Categorizing the world' contains two papers, one by Susan Caray and another by Frank Keil. Both papers are concerned with learning lexical semantics. The first paper suggests that infants have an evolving sortal hierarchy (used to learn concepts such as mass and count nouns), whilst the second paper argues that there is no difference between the mentalities of children and adults: both adults and children learn using probabilistic induction and theory-driven knowledge.

Following this, we have a section 'Categories, words, and language', with four papers, also concerned with lexical semantics. However, this time, the accent is on usage of other principles to constrain learning. Ellen Markman's paper discusses three constraints: the whole-object assumption (terms refer to objects as a whole rather than their parts); the taxonomic assumption (words are extended to objects of like kind); and the mutual exclusivity assumption (each object has only one label). Sandra Waxman suggests that infants expect words to refer to objects and object categories (these references are termed 'linkages'). Afterwards, the infant refines these linkages into particular forms such as nouns versus adjectives. Barbara Landau's paper proposes that the human spatio-cognitive system imposes constraints on how objects and places can be represented during first language learning. Paul Bloom's paper is concerned with how language can be learnt in the face of constraint violation (such as names for parts, rather than wholes). He suggests that the learner uses general syntax-semantic mappings (akin to semantic types in Montague semantics) to guide learning.

The section 'The case of verbs' considers verbal acquisition (as opposed to nominal acquisition). Fisher, Hall, Rakowitz and Gleitman comment that verbal acquisition differs from nominal acquisition. In particular, they argue that verbs are acquired in a syntax-orientated manner. By contrast, Steven Pinker argues against this stance, suggesting a more semantics-orientated approach. Jane Grimshaw attempts to reconcile these last two papers.

Finally, we have a section called 'Procedures for verb learning', with a paper by Michael Brent describing an implementation of a system that learns subcategorization frames, using a mixed statistical and knowledge-based approach. The section ends with a paper by Mark Steedman, commenting on Brent's approach, and summarizing lexical acquisition in general.

I would have preferred if there were more papers describing actual implementations of systems. Working systems clarify otherwise hazy theories, and pinpoint where the real problems are. For example, many of the theories presented in this volume do not make clear the relative contributions that induction and constraints make to learning. Furthermore, few, if any of the theories give a definition of what a lexical entry is supposed to be. It does seem that many of these workers are implicitly working in a Government-and-Binding framework. Hence, lexical entries are presumably relatively simple, and so easier to learn than say rules of syntax. However, this need not be the case. If instead the mental lexicon was encoded in a highly lexicalized theory, such as Head-driven Phrase Structure Grammar, lexical acquisition would include syntax learning, and so would need to be more involved than some of the approaches considered in this volume. Either way, a definition would not go amiss.

Editorially, the first two papers could easily be dropped, without loss to the volume as a whole. Most of the remaining papers offer useful insights into the lexical acquisition problem. The picture that seems to be emerging is that learning needs to use induction, augmented by constraints to reduce the hypothesis space of possible lexical entries (nothing new there). However, judging by the amount of contrary positions, it is clear that the field, as represented by these papers, has not reached a consensus. For example, witness the arguments between the role of syntax and semantics in acquisition. Consequently, people building practical NLP systems will have to wait a while before using much of this work. These quibbles aside, I found the majority of the papers interesting and accessible.

MILES OSBORNE

Kings College, University of Aberdeen

e-mail: mosborne@csd.abdu.ac.uk

Frank Palmer, editor. *Grammar and Meaning: Essays in Honour of Sir John Lyons*. Cambridge: Cambridge University Press, 1995. ISBN: 0 521 46221 5, £37.50 (hardback). 265 + xiii pages.

This book is a *Festschrift* for John Lyons and begins with the editor's Foreword. It provides a summary of each contribution, which I shall not attempt to emulate mainly because of the lack of space. For those readers who are not familiar with or interested much in theoretical linguistic work, it will help them decide which chapter(s) to read. Another place to start is the last chapter, which is rather unusually written by Lyons himself. It has the same title as the book itself, and some issues from each of the other chapters are taken up and related to Lyons's past work and current thinking.

Roughly, the contributions in the book can be divided into two groups. A majority (Kilgariff and Gazdar, Miller, Trudgill, Comrie and Anderson) provide arguments based on discussion of specific types or class of data. The others (Lehrer and Lehrer, Matthews, Kempson and Levinson) are rather programmatic and/or polemic. Dixon's chapter has characteristics of both.

Kilgariff and Gazdar's contribution concerns one of the much discussed topics in computational morphology and lexicography: this is perhaps the most directly relevant chapter

to language engineers. They discuss 'regular polysemy' (Apresjan, 1974), and argue that it is actually 'subregular' and, as such, it can be described by the same formalism as inflectional morphology. They use a lexical knowledge representation language that utilizes multiple default inheritance, DATR, an outline of whose essential machinery is provided in their chapter; for more details, readers are referred to Evans and Gazdar (1990). Their use of multiple default inheritance to deal with polysemy can be contrasted with work by, say, Copestake (1995), in which lexical rules account for this type of sense extension. Kilgariff and Gazdar conclude that 'once we have a fully developed theory of subregularity, it is unlikely that a distinction between "irregular polysemy" and "homonymy" would serve any purpose in a synchronic description of the lexicon' (p.2). This, despite a contrary impression one gets from their discussion, is not substantially different from what Lyons (1977, pp. 554–555) states in his discussion of the relationship between homonymy and polysemy.

Lehrer and Lehrer begin their chapter by noting the ambiguity of meaning. They argue for 'a model of sense, reference and interpretation of a word as an aggregation of input vectors' (p.38). It is an attempt at rather an ambitious goal of providing a comprehensive semantic theory that accounts for vagueness and indeterminacy, variation among individuals, and so on. Although I do not feel confident enough to make any comments on technicalities of their formalism, I am not sure whether providing an 'average' value of reference of an expression can indeed explain indeterminacy.

Matthews demonstrates that the extant distinction between semantics and pragmatics (and also between syntax and semantics) was a theoretical necessity for Bloomfield and structuralists. He then questions whether the distinction should be retained; his own answer apparently is in the negative. Until he shows us how a linguistics without the rejected boundaries might look, however, it would be rather difficult to concur with his argument without reservation.

In contrast, Kempson takes for granted a fairly traditional, yet by no means universally accepted, view of formal semantics. She provides a constructional model of interpretation which 'blurs syntax-semantics dichotomies' (p.85). Indeed, her 'content' seems syntactic, just like Logical Form (LF) in the Principle-and-Parameter Approach (cf. Chomsky, 1981) was syntactic. Just on the basis of this chapter, I cannot see why the properties that her system assigns to strings (= sentences) are properties of 'logic which ... are motivated by quite general properties of reasoning' (p.85).

Levinson's chapter is the first widely available version of his theory of Generalized Conversational Implicatures (GCIs). This should probably be seen as a trailer to his overdue book (Levinson, forthcoming), and the evaluation of his theory needs to wait until it comes out. In the abridged version in this chapter, the reason why what he calls GCIs should be bagged together is not made clear, as there are different types of them and one type of GCI ('Q2 inferences') behave rather differently from others ('Q1 and M inferences'; see p.103). I, for one, should like to see exactly how an inference system of GCIs is to be modelled, as this is one of the major issues in computational approaches to implicature generation and/or interpretation as evidenced by the recent AAAI symposium on the topic (see <http://www.isp.pitt.edu/implicature/>) [working as of 14/1/1997]. It is hoped that in a forthcoming, complete theory, Levinson will provide a full answer to criticism by, for example, Hirschberg (1990), that implicatures should not be treated as defaults, and that he will explain how non-monotonic logic of his choice would be restricted enough or explicate the *ceteris paribus* part of his theory.

Miller argues, at least for spontaneous speech, that there is no justification for having sentence as a theoretical construct, and that clause should be considered a basic unit instead. He further argues that some part of the language system is medium-dependent, and differences between the systems could be dealt with by 'discourse rules'. As Lyons (pp.235–236) notes in his response to Miller, even if clause were more basic than sentence, there would be little gain in abandoning the latter altogether as it has been useful in syntactic descriptions of many, if not all, languages.

Trudgill shows how a sociological factor may contribute to grammaticalization, which is explicated with examples from East Anglian dialects. Although it is a fine example of well-

argued sociolinguistic work, its relevance to language engineering is unclear to me, except perhaps for speech processing work capable of handling East Anglian dialects. (I must confess I do not know of any such projects, however.)

One of the issues that Comrie discusses in his chapter is the infamous semantic-pragmatic distinction. This he illustrates with German Perfekt and Praeteritum, corresponding approximately to the Present Perfect and the Simple Past in English, respectively. Comrie contends that the basic difference between these two is that the former has ‘time reference before a non-past reference point’ (p.160), whereas the latter refers to the past. In addition, the Praeteritum instructs ‘the addressee not to relate the situation as relevant to a non-past reference point’ (*ibid.*), and the Perfekt, in contrast, does no such thing. Everything else follows from this: for instance, it explains the reason why the Perfekt is usually employed to describe a situation that is somehow related to the present. Comrie appeals to the notion of relevance and, like Levinson, appeals to the notion of default interpretation. Moreover, like Levinson, Comrie mentions the possibility of contextual cancellation of default interpretation without formalizing it. Although his argument sounds intuitively plausible, it is difficult to evaluate his claim without some idea about a formalized cancellation mechanism.

Two traditional grammatical concepts that are central in Anderson’s chapter are concord and rection (or government). He argues that both of them underlie the possessive construction, using data from Old English, Makonde and Hausa. As I am ignorant of dependency grammar and other work by Anderson in particular, all I can say is that it would be rather difficult to model his analysis computationally because of a seemingly contradictory relation between concord and rection.

Dixon, another contributor with ‘his own individual research programme’ (to borrow Lyons’s characterization of Anderson’s work), discusses complement clauses. His examples are taken from English, Fijian and Dyirbal, three typologically diverse languages that use different complementation strategies. Dixon claims that selection of grammatical constructions to express a given meaning is semantically based; this I do not find controversial. Dixon’s idea of clause linking (i.e. how clauses may be realized as arguments such as subject and object) seems to me to be related to the series of work by Beth Levin and her coresearchers (e.g. Levin, 1993; Levin and Rappaport Hovav, 1995). This line of work with empirical evidence is promising in explicating the syntax-semantics interface.

The gem of this book is the final chapter by Lyons. As a former student of his, I am perhaps biased, but it certainly provides some insight to understand his work, and, to borrow Levinson’s words, it contains less ‘medieval casuistry’. As his chapter makes clear, Lyons, I believe, is a formally-oriented semanticist. The fact that his work typically does not contain logical symbols or formulae I do not think is consequential. Looking back at his own past work (Lyons, 1963, 1968), Lyons notes that he would now recast it with ontology and possible worlds, which, judging from his comment (and Lyons, 1968), could be linked with categorial grammar. An explicitly formulated version of this can serve as a basis of a rigorously formulated semantic theory. Perhaps he is just suggesting this as a possibility, but I, for one, should like to see how he materializes it.

References

- Apresjan, J. D. (1974) Regular polysemy. *Linguistics* 14 (2): 5–32.
- Chomsky, N. (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Copetake, A. (1995) Representing lexical polysemy. *Proc. AAAI Spring Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity and Generativity* Palo Alto, CA. AAAI Press. [Available from <http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html>].
- Evans, R. and Gazdar, G. (editors) (1990) *DATR Papers. Cognitive Science Research Paper. No. CSRP 139*. Falmer, East Sussex: University of Sussex.
- Hirschberg, J. (1990) *A Theory of Scalar Implicature*. New York: Garland.

- Levin, B. (1993) *English Verb Classes and Alternations: A Preliminary Investigations*. Chicago: University of Chicago Press.
- Levin, B. and Rappaport Hovav, M. (1995) *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.
- Levinson, S. C. (forthcoming). *Generalized Conversational Implicature*. Cambridge: Cambridge University Press.
- Lyons, J. (1963) *Structural Semantics*. Oxford: Blackwell.
- Lyons, J. (1968) *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Lyons, J. (1977) *Semantics, Vol. 2*. Cambridge: Cambridge University Press.

MAYUMI MASUKO

International Christian University, Tokyo

e-mail: mayumi@icu.ac.jp

Bas Aarts and Charles F. Meyer, editors, *The Verb in Contemporary English: Theory and Description*. Cambridge: Cambridge University Press, 1995.
ISBN 0 521 46039 5, \$37.50/\$59.95 (hardback, 313 pages).

The book starts with the editors' introduction which illustrates historical development of theory and important concepts. Aarts and Meyer state that 'an adequate understanding of the verb in contemporary English is best achieved if theoretical treatments of the verb are accompanied by studies that describe its usage' (p 2). When I read this sentence, I thought they meant a theoretical work incorporating some descriptive work. That indeed seems to me how linguistic theory should be: either theory has been proved by 'hard' evidence (e.g. large-sized corpora) or actually occurring data has lead to theory. Sadly, however, that probably happens in the ideal world only: the book consists of two separate parts, one theoretical and the other descriptive, each consisting of eight chapters. Partly because of the limitations of space and partly because of my interest, I shall focus on the first part.

The theoretical part consists of eight chapters. In their introduction, the editors claim that this part 'contains chapters that provide differing theoretical perspectives on the syntax and semantics of the English verb' (p 2). I find, however, that this part is biased towards the so-called Government-and-Binding theory (GB: cf. Chomsky, 1981) with some exceptions.

Chapter 2 by Meyer is more of a brief history of theories on verb syntax than a fully-fledged theoretical piece. He maintains that it is essential to consider two different kinds of relations: clause-level and general functions. The former corresponds to what are often called grammatical functions or roles, and the latter is more structural and includes complementation, modification and what he calls disassimilation, which seems a misnomer judging from the definition ('a relation in which there are varying degree of integration between constituents', p 34). His discussion focuses on one example each: complement (a grammatical function) and complementation (a general function). His rather informal account seems to me to resemble a more formally-oriented treatment in Lexical Functional Grammar (LFG: cf. Bresnan, 1982), of which he makes no mention. Rather surprisingly, as he himself admits, Meyer '[does not deal] with the issue of what particular relations needs be posited at each level in order to yield a satisfactory description of English and other languages' (pp 37-38). I find this baffling for a chapter supposedly concerning relations.

Hudson discusses a related issue in Chapter 3: complementizers or 'Comps'. His account is couched in Word Grammar, a variety of dependency grammar (see Hudson, 1990). He concludes that complementizers (*that, if, whether* and *for*) do not form a single word-class or category as they behave syntactically differently and the abstract syntactic position called 'Comp' does not exist. He maintains that traditional concepts such as 'head' and 'complements' are clearly identifiable from data and that subordination is a useful notion. Hudson's chapter is concise and convincing enough for those who do not subscribe to or know much about his framework (myself included).

In Chapter 4, direct object is defined as 'any Noun Phrase that is a verb complement and is not preceded by a preposition' (p 54). Schlesinger states that direct object could not be described in terms of semantic (or thematic) roles without an account of how a given role will be linguistically expressed. This is because the same role could be expressed in different ways, as a subject, an indirect object or direct object. Schlesinger postulates a Defining Participant (DP) which is a participant in the event described by the main verb in a given sentence. It does not have to be lexically expressed and its existence may only be implied by verb semantics. It is argued that direct object 'usually' expresses the DP. A participant in the event that is not a DP but 'affected' can also appear as a direct object. What is meant by 'affected' is not explicated and it is rather difficult to evaluate Schlesinger's account.

B. Aarts examines secondary predicates which he defines as 'phrases which are predicated of a constituent which stands in a thematic relation to the main verb of the sentence in which the secondary predicate occurs, typically a direct object' (p 75). There are two dimensions, resultative or depictive and subject- or object-related, distinguishing four different types. Aarts's argument is cast in the GB framework, more precisely its subtype called Small Clause Theory (cf. Hoekstra, 1988): an object-related depictive secondary predicate such as *her coffee black* in *Mary drank her coffee black* can be regarded as an example of a Small Clause. As for the question of whether object-related resultative predicates (e.g. *the cups clean* in *John rinsed the cups clean*), Aarts claims that they are not lexically selected. This is because if secondary predicates were to be lexically selected (either by subcategorization or selectional restriction), they would have to be arguments and predicates at the same time. That would pose an insoluble problem in GB as arguments have to get a theta-role from the verb but predicates cannot do so. Aarts concludes that they are adjuncts and not predicates. It seems to me that his account is given only to avoid theory-internal problems. This suspicion is strengthened by the fact that his account rests on the validity of Small Clause Theory which not all proponents of GB accept.

Huddleston's chapter is perhaps the least polemic in the theoretical ones and does not presuppose any specific framework. He argues, as the chapter title indicates, that the perfect in English is a secondary past tense. The perfect basically means that T_r (time referred to) precedes T_o (time of orientation). Although Huddleston does not provide clear definitions, this seems to mean that the event described by the past participle happened before 'now' in the case of the present perfect, and, with the past perfect, before some specific time in the past. The preterite (i.e. the simple past) has the same meaning, but it does not do so in two cases: (i) modal remoteness (e.g. If they *were* busy, they'd let us know) and (ii) backshifted preterite (e.g. I told John that I *was* tired). In these cases, the perfect can explicate the anteriority of T_r (cf. If they *had been* busy, they'd have let us know; I told John that I *had been* tired). A second difference is that the perfect is continuative in the sense that T_r begins before and may extend up to T_o . Thirdly, the preterite is usually deictic whereas the perfect is prototypically non-deictic. Moreover, according to Huddleston, the perfect is less grammaticalized than the preterite. Huddleston rejects alternative analyses such as Comrie's (1976, 1985) which regard the preterite as tense and the perfect as aspect on the ground that both the preterite and the perfect have as their basic meaning the anteriority of T_r relative to T_o and that the present and the past perfects behave differently. How exactly Huddleston's own claim that the perfect is a compound tense differs from positing aspect is not entirely clear from the discussion in this chapter.

Rosta discusses English mediopassives, which are perhaps better known as the middle construction. He argues that any verb can become a mediopassive if certain syntactic and semantic conditions are met. The subject of the mediopassive is not necessarily agent. This is comprehensible but Rosta appears to equivocate 'the subject' and 'the referent of the subject', which Rosta contends is identical with 'agent', when he mentions that some mediopassives have no subject (p 126). This is rather confusing. He then argues that the subject of a mediopassive is the 'archagonist' of the event described by the verb. Rosta does not give any formal definition of the concept. Although what exactly the 'archagonist' means thus remains unclear, it is said to convey a necessary condition for the event to take place. As he notes,

mediopassives elude the notion of grammaticality as some speakers accept (presumably) all mediopassives and other speakers are less receptive. When appeals to pragmatic or contextual factors are made, one would expect to find a detailed description of data and/or theoretical analysis of controlling factors. When the data in question involve varying judgements of native speakers of the language, as is the case with the mediopassive, this would indeed seem requisite. Rosta, however, does not quite do this, which is a pity as the data he deals with are interesting in their own right and have not been discussed in detail by other researchers.

The last chapter in the theoretical part examines the distinction between root/epistemic modality. Roughly speaking, epistemic modality concerns the speaker's judgements about the truth of propositions, thus conveying possibility or necessity. Root modality, on the other hand, involves necessity and obligation as well. Coates argues that the root/epistemic contrast is rather weak regarding possibility, and that this reflects the tendency of root modals to develop epistemic interpretations. This is clearly seen with *may* and *can*, though there is some difference between American and British English concerning the latter auxiliary. Coates's conclusion seems reasonable. I, however, would expect analyses of this kind to provide actually occurring examples, preferably from a reasonably-sized corpus, especially when claims concerning diachronic changes are made.

The eight chapters in the latter half of the book examine certain verbs or constructions. Not surprisingly, each of them provides information about the nature and peculiarity of corpora used. Among them, Aarts and Aarts's general comment is pertinent: using grammatical functions and categories in description is useful for making generalizations, but it has to be supplemented by lexical and/or collocational information if one wishes to provide insight into actual usage (p 180). They also note the necessity of adding semantically-based analyses to predominantly syntactic descriptions.

As I stated at the beginning, the most disappointing characteristic of the book is that theory and description is not satisfactorily connected in any of the chapters. This, I suspect, is due, at least partly, to the fact that the chapters are short. It certainly is very difficult, if not impossible, to provide a description leading to a convincing theoretical argument within approximately 20 pages. The Preface implies that this book may perhaps have been envisaged as a Festschrift for Sidney Greenbaum. As such, the editors might have felt the need to include as many researchers who have had contact with Greenbaum as possible. It is, however, rather frustrating to read papers that are literally restricted by their lengths. Another disappointment is that although many researchers appeal to semantic and/or pragmatic factors to deal with cases where syntactic accounts fail, what these factors are or how they behave is sadly not given. It would have been more instructive if the book consisted of longer chapters containing a sufficient description of data and a theory. Such work is more likely to be produced by computational linguists with access to large corpora and the need to provide an elegant and cost-effective account, rather than theoretical or descriptive linguists with typically limited resources, computational or otherwise.

References

- Bresnan, J. (editor) (1982) *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Comrie, B. (1976) *Aspect*. Cambridge: Cambridge University Press.
- Comrie, B. (1985) *Tense*. Cambridge: Cambridge University Press.
- Hoekstra, T. (1988) Small clause results. *Lingua* 74 101–139.
- Hudson, R. (1990) *English Word Grammar*. Oxford: Blackwell.
- Levin, B. and Rappaport Hovav, M. (1995) *Unaccusativity. At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.

MAYUMI MASUKO
Wasedo University, Japan

email: mayumi@mn.waseda.ac.jp

Michael L. Geis, *Speech Acts and Conversational Interaction*. Cambridge: Cambridge University Press, 1995. ISBN 0 521 46499 4 (hardback). £37.50. xiv + 248 pages.

This book presents Dynamic Speech Act Theory (DSAT) as an attempt to account for competence in naturally occurring dialogue. The book is relevant both to conversation theorists and to developers of conversational devices such as advanced spoken dialogue systems. The book is well-written and interesting reading.

DSAT departs from speech act theory, conversation analysis, and artificial intelligence approaches to natural language processing. The ‘dynamic’ part of DSAT comes from the claim that for a theory of conversational competence to be useful, it must be incorporable in a computational model of utterance generation and understanding. The presentation of and argumentation behind DSAT is based on analyses of transcriptions of real-life conversations, a very sound approach.

The book starts out with a review and critique of traditional speech act theory *a la* Searle. The traditional theory is claimed to fail, for two main reasons:

1. There is no effective mapping of complex multiturn interactions into traditional speech acts
2. Even if there was such a mapping, associating speech acts with individual utterances does not explain the interactions in which they occur.

Geis states that the traditional speech act theory claim that any sentence has a primary force is false. Communicative acts in dialogue are not results of single utterances but of the continued effort of participants. Moreover, in accordance with conversation analysis speech acts are social, not linguistic. The alternative offered by DSAT is that utterances have transactional significance related to the ostensible goal and/or interactional significance reflecting ‘face work’, the interpersonal side of interaction.

The central means of administering the utterance effects during conversation is the DSAT interaction structure. This structure contains (final) transactional and interactional effects, initial-state conditions, satisfaction conditions (ability and willingness), and a set of domain predicates for each action referred to in the effects and conditions. DSAT interaction structures are highly domain dependent. Geis uses a ‘slot and filler’ representation for interaction structures. They seem to originate from AI-based natural language processing and clearly have roots back to Marvin Minsky’s frames. Geis does not explain how such interaction structures (frames) may be designed, nor how to exploit the similarity of many such frames nor their domain predicates. However, for the construction of restricted task-domain system such as current task-oriented spoken language dialogue systems the approach seems practical and promising. Two related approaches from dialogue modelling are the interaction-as-filling-missing-axioms (Smith and Hipp, 1994) and the use of dialogue patterns (Novick, 1996).

The interaction structure corresponds to Searle’s speech act structure as follows:

- (a) The transactional effect corresponds to Searle’s essential condition. An important point of DSAT is the addition of an interactional effect.
- (b) The initial-state conditions correspond closely to Searle’s sincerity condition, but emphasises more strictly the psychological state.
- (c) The satisfaction conditions correspond to Searle’s preparatory conditions, but adds willingness to ability, and restricts the scope to conditions that must be satisfied before the transactional effect can be achieved, hence the name.
- (d) The domain predicates replace Searle’s propositional content condition because while single utterances might have a propositional content, multiturn interactions cannot.

According to Geis, terms like request, offer etc. are useful in informal discussions of what participants are doing in multiturn interaction, but do not play a role in achieving and

recognising goals. DSAT instead maps utterances in multiturn interactions into interaction structure conditions and domain predicates, instantiating slots of the DSAT interaction structure. Given that Searle's five general categories of speech acts (assertives, directives, commissives, expressives, and declaratives) seem to have psychological significance (Searle, 1983), and that sets of concrete instances of speech acts are widely employed in dialogue managers in natural language dialogue systems, the relation between these 'traditional' speech acts and the DSAT interaction structure mappings deserves more consideration than given by Geis. In DSAT the acts seem to pop up again as multiturn communicative actions, called encounters, such as service requests and invitations. Each interaction structure corresponds to a certain encounter in a certain domain (note that a given interaction may address several interaction structures). I believe that this approach and Geis' claim that speech acts are not properties of individual utterances may be right, but I would like to see more real-data tests of the respective qualities of DSAT and individual utterance speech acts.

A nice consequence of the framelike interaction structures is that computations of Gricean implicatures are reduced or even avoided. Conditions and subconditions are directly instantiated from one another according to the structure hierarchy, essentially providing 'precompiled inferences'.

Also, Geis argues and illustrates that simply storing representations of transactional and interactional significance of utterances in DSAT interaction structures accounts for adjacency pairs and insertion sequences. Adjacency pairs are in focus in conversation analysis but is criticised by, for example, Searle (1992).

Moreover, the complexity of indirect speech acts is replaced with the simple notion of direct communication (utterances that instantiate primary conditions of interaction structures) and indirect communication (utterances that instantiate preconditions of primary conditions).

Finally, I should mention that for the purpose of illustrating the computability of DSAT it is embedded in the Discourse Representation Theory (DRT) of Kamp and Reyle. Arguably this embedding enables DRT to account for conversational interaction. This part of the presentation is illustrative and sound, even though I still fail to see the first practical use of DRT.

Taking all this together, Geis convincingly presents Dynamic Speech Act Theory as part of the basis for further theoretical and practical work on conversation, relating the structure of conversation with utterance understanding and utterance generation, via mappings to an underlying interaction structure.

References

- Novick, D. G. and Sutton, S. (1996) Building on experience: Managing spoken interaction through library subdialogues. In: LuperFoy, S., Nijholt, A. and Veldhuijzen van Zanten, G. (editors), *TWLT 11. Dialogue Management in Natural Language Systems. Proceedings of the 11th Twente Workshop on Language Technology*, Universiteit Twente, Enschede, Holland, pp 51–60.
- Searle, J. R. (1983) *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle, J. R. (1992) Conversation. In: Parret, H. and Verschueren, J. (editors) (*On*) *Searle on conversation*. Amsterdam: John Benjamins, pp 7–29.
- Smith, R. W and Hipp, D. R. (1994) *Spoken Natural Language Dialog Systems: A practical approach*. New York: Oxford University Press.

HANS DYBKJÆR

Prolog Development Centre A/S, Denmark

email: dybkjaer@pdc.dk

Yorick A. Wilks, Brian M. Slator and Louise M. Guthrie, *Electric Words: Dictionaries, Computers and Meanings*. Cambridge, MA: MIT Press, 1996. ISBN 0 262 23182 4, £24.95 (hardback), ix + 289 pages.

Overview

Electric Words is the closest I have seen to a textbook in computational lexicography, charting as it does the philosophical, theoretical and computational aspects of the construction of large scale lexicons for use in Natural Language Processing (NLP). The 14 chapters which make up this volume are briefly overviewed below.

Brief contents

In the Introduction the authors reiterate the need for wide coverage lexicons in NLP and propose the use of Machine Readable Dictionaries (MRDs) intended for human use as a rich source for the construction of computational lexicons. Since dictionaries are mostly concerned with meaning, 'A Short History of Meaning' gives us a brief tour of 11 approaches to meaning, ranging from protoreferential theories (e.g. the meaning of 'Paris' is the actual city in France) to procedural views to meaning as use (i.e. the meaning of a word is defined by its relationship to other words). 'Symbolic Accounts of Definitional Meaning' elaborates on three approaches: truth conditions, meaning as deductive or inferential relations, and meaning as symbolic structures. 'Primitives in Meaning Definition' focuses on arguments for and against the possibility of atomic meaning units. 'Wordbooks as Human Artifacts: Dictionaries and Thesauri' argues for the notion of word senses as a useful concept in dictionary construction, and further defends the use of MRDs in lexical investigations. 'Early Computational Approaches' summarizes pioneering work on thesauri (Sparck Jones, 1986), taxonomies (Amsler, 1980) and MRDs (Michiels, 1982). 'Text Analysis and Its Relationship to Dictionaries: Dictionaries as Text' describes features of LDOCE (Procter, 1978) and COBUILD (Sinclair, 1987), as well as a variety of experiments to extract, use and elaborate on the lexical information contained in these and other MRDs. 'The Construction of Modern Lexicons' reviews a number of existing lexical resources used in NLP, ranging from Machine Translation (MT) systems to hand-crafted lexical databases. 'Automatic Techniques for Analyzing Machine-Readable Dictionaries' describes parsing and pattern matching of definitions and other techniques for eliciting information from MRDs. 'Genus Hierarchies and Networks' concentrates on the acquisition of disambiguated hierarchies of (mainly) nouns from MRDs. 'Numerical Methods for Language Processing' presents an overview of algorithms for dictionary construction and for word sense disambiguation. 'Automatic Processing of Bilingual Dictionaries' summarizes a variety of projects concerned with the construction of bi- and multilingual lexical resources, using parallel corpora and/or monolingual and bilingual MRDs. 'Consumer Projects Using Machine-Readable Dictionaries' describes two MT projects and two natural language parsing projects which have exploited MRDs in their construction or operation. 'The Present' concludes the book by outlining projects which have adopted diverse strategies to dictionary or computational lexicon construction, including the ACQUILEX project and the Japanese EDR lexicons; the chapter also directs readers to the infrastructure already in place to support the creation of large-scale NLP systems, including references and email addresses for the Linguistic Data Consortium and the Consortium for Lexical Research.

Conclusion

I would highly recommend the book as an educational and pedagogical text as it gives an extensive review of most of the literature in the subject, including an accessible discussion of some fundamental issues in meaning representation and of many of the implicit assumptions

made by computational lexicographers. Most of the algorithms are given as clear prose descriptions, and in such sufficient detail that one feels they could form the basis of actual implementations.

Interestingly enough, the book briefly mentions the legal and financial aspects hampering the widespread use of large lexicons in NLP. One could conclude from these observations that the visible and successful use of WordNet may be partly attributed to its painless availability; by contrast, and despite large resources having been devoted to the construction of computational lexicons from MRDs, their deployment is relatively rare.

Electric Words competes with other books in computational lexicography, but while most other texts are collections of articles, this book tries to give a coherent and motivated view of various aspects of the subject.

The book is not without its deficiencies. Some problems with its editing are listed in Michiels (1996) and will not be repeated here. One could add, however, that the approaches to meaning, as discussed in the earlier chapters, could have been better integrated with the more practical descriptions later on in the book. For example, when the treatment of empty semantic heads (e.g. 'kind' as in 'a kind of lily') in definitions is discussed (pp 164–168), one hopes for a motivated discussion on the relationship between the headword and the proper semantic head, rooted in the discussion on meaning in the early chapters. However, the authors only give a brief description of the technical solutions adopted by various researchers.

The uninitiated may question the whole computational lexicography enterprise and wonder why the resources spent on analysing and processing MRDs were not spent on building a computational lexicon by hand. The answer is that if we develop tools for MRD analysis, we can use them for other dictionaries and languages and at the same time learn more about one particular way in which humans describe meaning. However, this answer is not clear from the book: the prominence of LDOCE and the exploitation of its peculiar properties does not inform us about the effect of applying the same techniques to other dictionaries in other languages, especially in terms of the quality of the resulting lexicon.

One would have expected to see mention of some of the possible areas in NLP where MRDs would at best only partially overcome the lexical bottleneck. Natural language interfacing, where one translates from the linguistic to the non-linguistic realm, is one such area, since the non-linguistic knowledge of lexical entries in such systems is not usually present in MRDs. One can use the syntactic and semantic information in MRDs to help language analysis, but the bulk of the translation of lexical items into say, SQL or some graphics package commands, will still have to be coded through other means.

Finally, while it could be said that the book will be of primary benefit to those about to embark on research in computational lexicography, those in the broader field of Natural Language Engineering can also benefit from the thorough survey of the literature provided.

Acknowledgements

Thanks to Blaise Nkwenti-Azeh for useful comments; errors and omissions remain my own.

References

- Amsler, R. A. (1980) The structure of the Merriam-Webster Pocket Dictionary, *Technical Report, TR-164*. Austin, TX: University of Texas at Austin.
- Sinclair, J. (editor) (1987) *Collins Cobuild English Language Dictionary (1st ed.)*. London: Collins.
- Procter, P. (editor) (1978) *Longman Dictionary of Contemporary English*. Harrow, UK: Longman.
- Michiels, A. (1982) Exploiting a large dictionary database, *PhD Thesis*, Liege, Belgium: University of Liege.
- Michiels, A. (1996) Review of *Electric Words: Dictionaries, Computers, and Meanings*, *Computational Linguistics* 22(3): 435–440.

Sparck Jones, K. (1986) *Synonymy and Semantic Classification*. Edinburgh: Edinburgh University Press.

ARTURO TRUJILLO
UMIST, UK

email: iat@ccl.umist.ac.uk

Nancy Ide and Jean Veronis, *Text Encoding Initiative: Background and Context*. Dordrecht: Kluwer Academic, 1995. ISBN 0 7923 3704 2, US\$45.00 (paperback), US\$99.00 (hardback). (Reprinted from *Computers and the Humanities* 29, 1–3, 1995.)

The publication of this book is good news for those who want to learn about the work of the Text Encoding Initiative (TEI) but who might find the current two-volume, 1290-page *Guidelines* too detailed to serve as a friendly introduction. (TEI Proposal number 3, or TEI P3, is formally known as the *Guidelines for Electronic Text Encoding and Interchange*, Sperberg-McQueen and Burnard, 1994. Information on the complete guidelines is available at <http://www.uic.edu/orgs/tei/>, inspected on June 30, 1997.)

The present volume is the work of members of the TEI who decided to document the history of the TEI, the special problems faced by different workgroups, the tradeoffs that led to the guidelines taking their current form, and the work remaining to be done in order for the TEI to realize its grand aim of meeting the multilingual processing and analysis needs of a wide range of electronic text users: researchers in the humanities, social sciences, sciences, librarians, publishers and others who face the challenges of document search and retrieval as well as document storage and retrieval.

The TEI came to life following a conference on the standardization of encoded text at Vassar College in November 1987 when the Association for Literary and Linguistic Computing and the Association for Computational Linguistics joined with the Association for Computers and the Humanities (ACH) to form a major international effort to establish guidelines for the encoding and interchange of literary and linguistic data. From the chapter entitled 'The TEI: History, Goals, and Future' by Ide and Sperberg-McQueen, we learn that after the publication of the first draft of the Guidelines (TEI P1) in July, 1990, about 20 small, specialized work groups were formed to extend and revise the work of the original four committees: text documentation (TD), text representation (TR), text analysis and interpretation (AI) and metalanguage issues (ML).

These workgroups have tried to express the best way to encode consistently and comprehensively a wide range of text types and features, including character set issues, text criticism, hypertext, mathematical formulae, language corpora, verse, drama, spoken texts, historical studies, dictionaries, etc. This work has resulted in the definition and modification of a core DTD (Document Type Definition) of SGML (Standard Generalized Markup Language) and several auxiliary DTDs.

For those new to SGML, a DTD is separate from any tagged document and is used to store the definitions of the different tags. The DTD serves to ensure conformance with the encoding guidelines. For those for whom 'DTD' is an unfamiliar term, it may help to know that HTML, the familiar data format of the World Wide Web, is also a DTD of SGML. The function of the HTML DTD is to ensure conformance to HTML guidelines; similarly, the function of the TEI DTD is to aid the user in automating conformance to the TEI guidelines. Sperberg-McQueen and Burnard's chapter 'The Design of the TEI Encoding Scheme' offers details on the different components of the TEI DTD, including the header, the writing system declaration, the feature system declaration, and the tag set declaration. Because it is anticipated that researchers from different disciplines will want to look at different types of information in a given text, the TEI DTD resembles a complex database from which 'different views may be derived for different applications' (p. 27).

It is not necessary to read the chapters of the Ide and Veronis book in order, especially if you are familiar with SGML or HTML. If you are not, the chapter by Lou Burnard entitled 'What is SGML and How Does It Help' is a clear introduction. It emphasizes SGML's strengths as a tool for tagging a document's content and structure independent of time, place and any specific text processing formatting software.

Gaylord further discussed the topic of character representation in his chapter. Codes are provided for alphabetic languages only in the current Guidelines, but the work of this committee continues. Giordano and Dunlop discuss the TEI header and practical application of it in the British National Corpus in chapters. Both the strengths and weaknesses of the TEI default header are explained forthrightly.

There are ten chapters on the encoding challenges of specific text types, including verse, performance text, textual criticism, history, spoken text, terminology, dictionaries, hypertext, linguistic features, and the encoding of text that does not lend itself to hierarchical representation. Many of the workgroups struggled with the challenge of keeping the coding 'general purpose' and avoiding theory-specific feature tags. Each chapter covers the trade-offs that led to the recommended encoding conventions for each type of text as well as the areas where more work remains to be done. A major theme that emerges from the work of each committee as described in this volume is the tension between generality and descriptive power. It is hard to come up with a DTD (or auxiliary DTD) that is general enough to be broadly applicable, yet specific enough to ensure the correct encoding of a very specific structure, such as a play or a dictionary entry.

Almost every author mentions the fact that much of the work remaining is in making these standards accessible to potential users by means of tutorials, introductions, good communication, and well-designed user interfaces. 'P3 is only imperfectly suited to the role of a reassuring introduction to the TEI encoding scheme. It is therefore a matter of some priority for the TEI to produce a set of introductory manuals for the Guidelines' (p. 20). 'There is a need for supporting software which connects the systematic underlying representation with a workable format for input and display' (p. 157). 'What is required now is the software to implement it' (p. 148). Specific areas that remain to be addressed in the *Guidelines* include letters and memoranda, computational lexica and analytic bibliography of early printed books. The authors look forward to large-scale use of the *Guidelines* so that they can learn what works and what areas are in need of refinement to be useful.

I recommend this book highly. It succeeds both in serving as an informative and readable introduction to the TEI and in addressing many more technical and detailed topics that will appeal to potential TEI users from many different fields. The references and bibliographies are of the highest academic quality. The chapters are well written, edited and proofread.

References

Sperberg-McQueen, C. M. and Burnard, L., editors (1994) *Guidelines for Electronic Text Encoding and Interchange*. Chicago and Oxford.

CANDACE MCKENNA
Microsoft

email: candacem@microsoft.com

Jennifer Cole, Georgia M. Green and Jerry L. Morgan, editors, *Linguistics and Computation*. CLSI Lecture Notes no. 52. Stanford, CA: CLSI, 1995.
ISBN 1 881526 81 X, £14.95 paperback, xi + 296 pages.

This book consists of the papers that were presented at a workshop held in June 1991 at the University of Illinois at Urbana-Champaign to celebrate the 25th anniversary of the UIUC Department of Linguistics. The purpose of the workshop was to bring together both

computational and more traditional linguists, e.g., syntacticians, morphologists and phonologists, to discuss important issues of common interest. The ten papers in this book are split over three areas: syntax, automated parsing and generation, and phonology. Though the areas of linguistic research addressed in these papers are diverse, the editors note in the Preface that these papers are nonetheless united by the notion that the principles of grammatical well-formedness and language processing are interdependent, and that research in theoretical and computational linguistics should thus proceed hand in hand.

I will say up front that I enjoyed this book very much, and think that the papers therein make a good case for such a joint research effort. Below I will give brief summaries of the papers in each section, followed by a very short list of the book's flaws and a much longer list of its strengths.

The first section of the book focuses on syntax and how it is represented and manipulated within various formal frameworks. Annie Zaenen and Ronald M. Kaplan's 'Formal devices for linguistic generalizations: West Germanic word order in LFG' (pp 3–27) shows how non-local dependencies in the infinitival patterns of Dutch and Swiss German can be described within Lexical Functional Grammar (LFG) by invoking two mechanisms (the principles of functional uncertainty and functional precedence) that were originally developed for very different purposes. David E. Johnson and Lawrence S. Moss's 'Stratified feature structures for multistratal relational analyses' (pp 29–84) gives a self-contained overview of Stratified Feature Grammar (SFG), a logic-based extension of relational and metagraph grammar which generalizes the notion of an atomic feature to a sequence of 'stratified' features, each of which encodes a distinct primitive syntactic relation, e.g. direct object. This paper also describes restricted versions of SFG whose weak generative capacities are more appropriate than unrestricted SFG for describing natural languages. The final paper in this section is Alan M. Frisch's 'Feature-based grammars as constraint grammars' (pp 85–100), which shows that the parsing of feature-based grammars can be simplified by specifying feature-structures implicitly via collections of constraints, and proves the correctness of a generalized context-free parsing strategy for feature-based grammars that operates directly on these collections of constraints.

The second section of the book is concerned with various approaches to the parsing and generation of natural language. The first half of Robert C. Berwick and Sandiway Fong's 'A quarter century of computation with transformational grammar' (pp 103–143) is a concise historical overview of the technical reasons behind the late 1960s decline in and subsequent 1980s resurgence of research on Transformational Generative Grammar (TGG) parsers. The second half of this paper describes both a parameterized TGG parser developed by the authors and how this parser has been configured to handle English and Japanese. Steven Abney's 'Chunks and dependencies: Bringing processing evidence to bear on syntax' (pp 145–164) suggests how the apparent mismatch of prosodic and syntactic structures in utterances may be reconciled by basing prosodic structure on a decomposition of the full syntactic structure into clause subtrees (chunks), and shows that this scheme accounts for various experimental results in the psycholinguistic literature. Dale Gerdemann and Erhard Hinrichs's 'Some open problems in head-driven generation' (pp 165–197) shows how the problems caused by non-canonical, e.g. empty, heads when unification grammars are used generate speech under a head-driven control strategy can be resolved by distinguishing carefully between the metalanguage and the object language variables in logical forms. The final paper in this section, Tsuneko Nakazawa's 'Construction of LR parsing tables for grammars using feature-based syntactic categories' (pp 199–219), describes in detail how Tomita's extended LR parsing algorithm can be adapted to handle grammars with complex feature-value systems.

The third and final section of the book focuses on phonological computations. John Coleman's 'Phonology and computational linguistics – a personal overview' (pp 223–254) has two independent parts: the first half is a selective but fascinating overview of various systems and theories that have been developed over the last 30 years to parse and generate phonological structures, and the second half is a description of YorkTalk, a text-to-speech system based on declarative constraints which has been able to avoid much of the computational and technical

complexity underlying rule-based phonological systems by using non-segmental phonological representations. Jennifer S. Cole's 'Eliminating cyclicity as a source of complexity in phonology' (pp 255–279) shows how the same complexity associated with cyclic rule application in rule-based solutions for word stress placement in Chamorro and English can be avoided in the declarative framework by an ordering of the types of domains in which the constraints apply. This solution is also of historical interest, as it independently proposes elements of the constraint ranking strategy underlying another declarative grammatical framework, Optimality Theory (Prince and Smolensky, 1993). This section and the book concludes with Julia Hirschberg and Richard Sproat's 'Pitch accent prediction from text analysis' (pp 281–296), which discusses the technical problems associated with reconstructing pitch accents, i.e. discourse-motivated emphasis and de-emphasis of particular words in an utterance, from text. The paper describes how these problems are being handled in the Bell Laboratories Text-to-Speech (TTS) system via an algorithm that consults a hierarchical representation of the attentional structure of the discourse in addition to purely syntactic information.

Three possible criticisms of this book are that many of the articles are not introductory reviews, that many of the papers are out-of-date (the full versions having since appeared elsewhere), and that the list of topics treated is not comprehensive. These criticisms are unwarranted; this book is, after all, a collection of papers from a meeting that occurred five years ago, not a cutting-edge handbook on computation and linguistics as the whole. Moreover, given the implicit limitations of such workshop volumes, it is a very well-constructed one. Though the papers therein are not on the whole either introductory or as cutting-edge as when they were originally presented at the workshop, they are clearly written with good literature reviews and bibliographies which should be adequate starting points for those wanting leads to more recent work. Though the list of topics covered is not comprehensive, those that are covered are covered well. Until a handbook of computation and linguistics does appear, those looking for a wide variety of papers on computational linguistics could do much worse than read this book (whose weaknesses, oddly enough, are complemented by the papers in Ristad (1994), which focus on computational aspects of the phonetics/phonology interface and morphology – see Wareham, 1996, for details).

The last remark above highlights the main strength of this book – namely, that the variety of papers contained in this book is successful on three levels. First, there is a good balance in the contributors between computational linguists and grammar theorists, and the papers by each group take care to address the concerns of the other. Second, there is a good mix in the approaches to problems, from theory to experiments to algorithms to working systems, which shows the full range of research methodologies at work in the computational linguistics community. Third, and perhaps most important, there is a good assortment of research topics covered in this book that one might not typically see grouped together, but whose inter-relationships become apparent when they are placed next to one another. Perhaps the most intriguing relations so formed will be those with areas familiar to the reader but not addressed in this book. Two such connections noticed by this reviewer are the uncanny manner in which current efforts at implementing parsers based on Optimality Theory seem to be recapitulating early work on TGG parsers (as described by Fong and Berwick), and the similarity between the chunking strategy proposed by Abney to account for the prosodic/syntactic disparity in adult speech and various chunk-based strategies that seem to be used by children acquiring language (Peters, 1983). Readers with other interests will undoubtedly notice other relationships. Such serendipitous discoveries and the new research that they suggest may, in the long run, be one of the greatest benefits one can derive from reading this book.

In conclusion, then, this is an attractive reasonably-priced book whose content and spirit should be useful to researchers of all levels and kinds of expertise within computational linguistics, from grammatical theorists to systems engineers. In the Preface, the editors say that the original workshop was highly successful in its goal of bringing people together for an exchange of ideas, and that it was their hope that this book would extend that opportunity to

a wider audience. That it most certainly has; hopefully, many computational linguists will take advantage of this opportunity.

References

- Peters, A. M. (1983) *The Units of Language Acquisition*. Cambridge Monographs and Texts in Applied Psycholinguistics. Cambridge University Press.
- Prince, A. and Smolensky, P. (1993) Optimality Theory: Constraint Interaction in Generative Grammar. *Technical Report RuCCS TR-2*, Rutgers University Center for Cognitive Science.
- Ristad, E. (ed.) (1994) *Language Computations*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 17. Providence, RI: American Mathematical Society.
- Wareham, H. T. (1996) Book notice: Eric Sven Ristad (ed.) *Language Computations*. *Language* 72(1): 186–187.

H. TODD WAREHAM
University of Victoria

e-mail: harold@csr.uvic.ca