ELSEVIER

Journal of Cognitive Systems Research 2 (2001) 167–172

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Book review

Review of: Speech and language processing

Action Editor: Stefan Wermter

Sheila Garfield*

University of Sunderland, Hybrid Intelligent Systems Group, Centre for Informatics, SCET, Room 201, Sunderland SR6 0DD, UK

Received 1 December 2000; accepted 1 February 2001

The book ''Speech and Language Processing'', Daniel Jurafsky and James Martin, Prentice Hall 2000, will be of interest not only to those who are working within the fields of speech and language processing but also to those who are new to the fields of natural language processing, computational linguistics and speech recognition. It covers the major areas related to speech and language processing and provides a comprehensive review of the field.

The text is divided into four sections covering the main themes of Words, Syntax, Semantics and Pragmatics. These sections are divided into chapters covering particular areas of the main themes such as Morphology and Finite-State Transducers and Dialogue and Conversational Agents. In addition there are a number of useful appendices as well as an extensive bibliography which is in excess of 20 pages. Throughout the text there are numerous pointers to other sources of reading. Each chapter begins with an outline of the material to be covered and ends with a summary, bibliographical and historical notes and exercises. The summary outlines the main points covered with emphasis on key words or phrases whilst the bibliographical and historical notes section draws the reader's attention to additional material or ideas not covered in the main text with suggestions for further reading for those interested in particular aspects of the topic. The exercises provide the reader with the opportunity to practice the techniques and/or methods introduced in the chapter. Where appropriate, points of theory introduced in each chapter are supported by illustrations to enhance the written text. For a reader unfamiliar with the subject matter the text is well-written, clear and concise. The book begins with a good introductory chapter that outlines the areas that will be covered in the remainder of the book.

Because the field of speech and language processing is encompassed by different but overlapping areas such as computational linguistics, natural language processing, speech recognition and computational psycholinguistics, a brief history is given of the main motivational factors in these areas from the 1940s until 1999.

The book introduces the algorithms that are the basis on which each particular field has developed. This is particularly advantageous for those new to this field of study but can be skipped by those who already have some knowledge in this field. Examples are also used to illustrate aspects of theory.

The book has been written in such a way that it is possible for readers to start at any point that is of particular interest or relevance to them; there appears

*Tel.: +44-191-515-3268.

E-mail address: sheila.garfield@sunderland.ac.uk (S. Garfield).

1389-0417/01/$ – see front matter
PII: S1389-0417(01)00022-5

to be no need to read each chapter or section in order as each area is relatively independent of each other.

Part I – Words

Following the introduction the opening chapter in Part I provides coverage of **Regular Expressions and Automata**. This has been approached in a straightforward, clear and concise manner. The coverage given by Jurafsky and Martin clearly spells out what regular expressions are and how, where and what they can be used for. The chapter progresses on to Finite-State Automata (FSA) which again is clearly explained and supported with illustrations to help the reader understand.

The initial ideas exposited in the chapter on Regular Expressions and Automata are further expanded in the chapter on Morphology and Finite-State Transducers. The broad term 'parsing' is introduced and the reasons why we need to be able to parse an input are explained with particular reference to morphological parsing. There is coverage of English morphology with emphasis on inflectional morphology and derivational morphology which leads into an account of morphological parsing with Finite-State Transducers (FST). Again the theory is supported with figures and tables.

For anyone looking for a good introduction to phonology there is an excellent introduction in this text which includes the vocal organs and articulatory phonetics covering consonants, vowels and syllables. This leads to the implementation of phonological rules as transducers (Kaplan & Kay, 1981; Kaplan & Kay, 1994) where finite-state transducers are used to model phonological rules, which is one of the most significant methods in the field of computational phonology. The chapter progresses on to advanced issues such as harmony, templatic morphology and optimality theory.

As this text covers the statistical methods used in the field of speech and language processing there is thorough coverage of the Bayes Rule as applied to spelling and pronunciation as well as the minimum edit distance algorithm for spelling error correction. Again this discussion of the topic covers the use of automata and in this case weighted finite-state automaton/transducer or probabilistic FSA/FST. There is extensive analysis of weighted automata and both the forward and Viterbi algorithms.

Within this book there is an entire chapter devoted to N-grams and their affinity with Markov models. N-grams can be used to model word prediction based on probability and another probabilistic model, the Markov model, is also introduced. Algorithms for smoothing are also highlighted such as backoff or depleted interpolation and Witten-Bell and Good-Turing discounting, which can be used to improve the estimation probability of N-grams. There is also discussion of entropy and perplexity which are metrics commonly used to evaluate N-gram systems.

The chapter on Hidden Markov Models (HMMs) and Speech Recognition concludes Part I of the book. This chapter covers speech recognition architecture, in particular the HMM supported by illustrations, the notion of spectral features, the forward-backward algorithm for training HMMs and the Viterbi and stack decoding or A* decoding algorithms for solving decoding problems. There is discussion of the acoustic processing of speech, the computation of acoustic probabilities, the training of a speech recognizer and human speech recognition. The reader is also introduced to the idea that aspects of human lexical access such as frequency, parallelism, neighbourhood effects and cue-based processing correspond to features of automatic speech recognition models.

Part II – Syntax

Part II begins with the topic of **Word Classes and Part-of-Speech Tagging**. There is an in-depth study of word classes, i.e., which words belong to a closed class type and which belong to an open class type with a comprehensive description of the lexical classes of words we use daily, e.g., nouns, verbs, adjectives, prepositions etc. The chapter also investigates tagsets and part-of-speech (POS) tagging referring to the two main techniques used for part-of-speech tagging: rule-based and stochastic, as well as discussing Transformation-Based Tagging (or Brill tagging) after Brill (Brill, 1995) which makes use of concepts from rule-based and stochastic taggers.

Despite extensive coverage of two of the main techniques for part-of-speech tagging there is only brief mention of the use of other tools for the automatic training of taggers in the chapter summary. One such tool is the Maximum Entropy Algorithm (Ratnaparkhi, 1996) which estimates probabilities as well as outputting tags. This algorithm is discussed in greater detail by Brill in the *Handbook of Natural Language Processing* (Brill, 2000) along with Transformation-Based Learning and Learning without a tagged text.

This section on Syntax moves on to discuss Context-Free Grammars for English in Chapter 9 and proposes three new concepts: constituency, grammatical relations and subcategorization and dependencies. However, despite these new concepts being introduced in this chapter they are not discussed in detail at this point. Two of the concepts, grammatical relations and subcategorization and dependencies, are discussed in greater detail in Chapter 11 on Features and Unification. A greater theoretical overview of the notion of constituents is provided by Steedman (Steedman, 2000) in his chapter on Rules, Constituents and Fragments. However, this syntactic knowledge can be modelled using grammars based on context-free grammars (CFG) and CFGs are introduced in this chapter. The description of a context-free grammar is both comprehensive and clear. The ideas relating to lexicons, grammars and parse trees are supported diagrammatically. Other topics discussed are sentence-level construction, agreement and finite-state automata.

The discussion of CFGs is continued in the next chapter, Parsing with Context-Free Grammars, which begins with parsing as search and presents the two main search strategies: top-down or goal-directed search and bottom-up or data-directed search. Both are covered in detail with problems highlighted such as left-recursion, ambiguity and inefficient reparsing of subtrees. Again examination of these areas is supported by diagrammatic illustrations of the subject matter. At this juncture the Earley Algorithm, which uses a dynammic programming approach, is introduced and finite-state parsing methods are also presented. However, one parsing technique that is not discussed is Augmented Transition Networks (ATN) which was a leading model during the early 1980s; brief coverage of this topic is given by Samuelsson and Wirén (Samuelsson & Wirén, 2000).

The next topic for analysis is Features and Unification. The idea is that complex sets of properties can be associated with grammatical categories and grammatical rules. Beginning with a description of feature structures and their forms of representation, e.g., an attribute-value matrix (AVM) or a directed-acyclic graph (DAG) it then moves on to the idea of unification of feature structures before introducing the idea of feature structures and grammars. The chapter continues with the implementation of unification and discusses the unification algorithm and follows this with the integration of unification into an Earley parser. The chapter finishes with a review of types and inheritance. This places restrictions on the type of values a feature can have and presents the notion that types are arranged in a type hierarchy in which specific types inherit the properties of more abstract types.

Lexicalized and Probabilistic Parsing is the next topic which raises issues concerned with the integration of structural and probabilistic models of syntax. As noted by Jurafsky and Martin this is at ''the very cutting edge of the field'' and therefore there is as yet no standardisation as there is with the CFG which has become the standard for non-probabilistic syntax. Also under discussion are probabilistic lexicalized grammars which are lexicalized grammars, i.e., a CFG with more rules, augmented with probabilities. There is however, only brief mention of probabilistic lexicalised representation in parsers (Charniak, 1997) although there is coverage of human parsing and in particular the resolution of ambiguity and the use of subcategorization information.

The final chapter in this section deals with Language and Complexity and considers the issues of the complexity of a formal language and the complexity of a human sentence and the need to identify these when deciding which formal model to use. It begins with an account of the hierarchy of grammars based on the restrictions placed on the production rules, known as the Chomsky hierarchy (Chomsky, 1959). The chapter concludes with an investigation of complexity and human processing and why certain

sentences are hard to understand perhaps because of their syntactic structure and the limitations of human memory.

Part III – Semantics

Coverage of the area of Semantics begins with **Representing Meaning**. The ideas outlined cover meaning representations using formal structures; meaning representation languages in relation to syntax and semantics used in meaning representations; semantic analysis covering the process of creating and assigning commonsense knowledge to linguistic inputs; and literal meaning dealing with the conventional meanings of words and not necessarily the context in which they occur. First Order Predicate Calculus (FOPC) which is the principal technique for investigating issues connected with semantics is introduced.

The book progresses naturally on to the topic of Semantic Analysis. Here the meaning representations addressed in the previous chapter are created and attributed to linguistic inputs. This investigates the notion that the meaning of a sentence is partially based on its syntactic structure and not solely on the words it contains. This leads to the idea of syntax-driven semantic analysis and the introduction of lambda notation or lambda calculus developed by Church (Church, 1940) as an extension to FOPC. This allows meaning or semantic attachment to be assigned to natural language expressions, i.e., information that states how to figure out what the meaning representation of a particular sentence structure is based on the meaning of its components. The chapter concludes with an examination of the topic of robust semantic analysis from a practical systems viewpoint and describes the two primary methods of implementing a syntax-driven approach: semantic grammars and cascades of finite-state automata.

The structure of words, what these words mean and how they can be used are some of the topics discussed in the chapter on Lexical Semantics. Issues covered in particular detail are homonymy, polysemy, synonymy and hyponymy. Also discussed is the internal structure of words looking at thematic roles and their use in computational systems, selectional restriction and the ability to impose constraints on lexemes and phrases.

To conclude Part III discussion focuses on Word Sense Disambiguation and Information Retrieval. It begins with an analysis of selectional restriction-based disambiguation using a rule-to-rule approach that attempts to reduce the amount of ambiguity by ruling out inappropriate senses. It continues with a discussion of machine learning approaches namely feature vectors, supervised learning, bootstrapping and unsupervised methods to produce more robust systems for sense disambiguation. The area of information retrieval is examined with the focus on the vector space model which represents documents as a vector in a multi-dimensional space.

Part IV – Pragmatics

This is the final section of the book and the opening chapter written by Andrew Kehler is an investigation into **Discourse** and covers issues relating to reference resolution, text coherence and discourse structure. The chapter begins with the area of reference resolution and introduces terms such as ''referring expression'' – an expression used to achieve reference, and ''referent'' – the individual/object being referred to, and presents the idea that we use expressions to refer to something, e.g., she or Susan to indicate a person named Susan. This leads into exploration of reference phenomena starting with the different types of referring expression and then different types of referent. Remaining with the theme of reference resolution Kehler proposes factors that need to be considered when trying to develop algorithms that are able to pinpoint the referents of referring expressions. Kehler also describes algorithms for pronoun resolution. The coverage given to the topic of Discourse takes a different approach to that presented by Lochbaum et al. (Lochbaum, Grosz & Sidner, 2000). They discuss discourse structure in the same chapter as intention recognition with greater coverage of discourse structure in terms of the two main approaches: informational and intentional.

The discussion moves to text coherence and the theme of coherence in discourse, outlining that there is a need for connections between utterances for the

communication to be coherent. This introduces the notion of inference, both deduction and abduction and an inference based resolution algorithm. The question of coherence within a longer discourse is posed and dealt with by discussion of the structure of discourse.

The content of the next chapter Dialogue and Conversational Agents is a natural progression from the preceding chapter on discourse. There are programs, known as conversational agents, that can communicate with users in natural language and it seems appropriate therefore to analyse conversation. However, conversation is distinct from other types of speech primarily because of turn-taking, grounding and implicature. Each of these factors is examined in detail and comprehensive examples given to support the theory. The analysis moves on to dialogue acts which can be viewed as being an action performed by the speaker and the understanding of dialogue acts presenting two models for dialogue interpretation: the plan inference model and the cue model. Again the question of coherence is raised and this is handled by describing an approach to coherence referred to as ''intentional'' that arises from the views of the BDI (belief-desire-intention) approach.

As the chapter is both about dialogue and conversational agents the final section looks at dialogue managers in conversational agents and in particular at issues relating to the design of dialogue managers. The issues, problems and influences associated with building a dialogue manager are discussed at length.

Natural Language Generation is the subject examined by Keith Vander Linden in his chapter by the same title. Natural Language Generation means building natural language outputs from inputs that are non-linguistic. This according to Vander Linden is a relatively new field of study. However, he presents an architecture for natural language generation systems and goes on to describe in detail two of the components of this architecture, the Discourse Planner and the Surface Realizer. During discussion of Surface Realization Vander Linden introduces Systemic Grammar, a method for choosing the language construct based on the context (Halliday, 1985) and Functional Unification Grammar that takes features from the input and merges them with the grammar to produce the output. For the reader wishing to explore Systemic Grammar and the develoment of MultiLingual Grammars further the KPML (Komet-Penman MultiLingual) development environment (Bateman, 1997) may be of interest. Vander Linden ends the chapter by discussing among other things microplanning, lexical selection and generating speech.

The final chapter was written mostly by Nigel Ward and deals with the area of Machine Translation. The chapter presents approaches used for machine translation: transfer, interlingua, direct and statistical. However, before these techniques are discussed there is investigation into the similarities and differences in language. One of the issues covered is the fact that the basic word order of subjects, verbs and objects can differ across languages. This and other factors have a bearing on how difficult the translation from one language to another will be.

This book not only focuses on the fundamental elements of this field such as Regular Expressions and Automata but also introduces other fields of study such as Natural Language Generation. Therefore it provides coverage of the major areas related to speech and language processing and would be useful both as a textbook and as a reference source. However, whilst automata and grammars are given extensive coverage and other techniques such as the Bayesian model for speech recognition and the Viterbi algorithm are revisited, artificial neural networks, another major area of natural language processing (Wermter, 2000; Mayberry & Miikkulainen, 2000; Morris, Cottrell & Elman, 2000), is only briefly mentioned. By contrast the area of artificial neural network approaches to natural language processing is given extensive coverage in the *Handbook of Natural Language Processing* edited by Dale, Moisl and Somers.

The approach adopted covers the area from what appears to be a linguistics-based point of view rather than an empirical one; hence perhaps the reason for the lack of coverage of topics related to artificial neural network approaches. The book also lacks discussion in other areas, e.g., alignment of texts and corpus creation. As it is an introductory text it would appear that it was a conscious decision not to cover these. However, some reference to these subject areas may have been beneficial to guide the reader further into the field.

Despite this it is a 'must' for those new to this field. Overall, it is very well written and presented apart from some typographical errors.[1] In summary, the book is highly recommended for all involved in the field of speech and language processing. This will be a very important reference for speech and language processing for quite some time.

References

Bateman, J. A. (1997). Enabling technology for multilingual natural language: the KPML development environment. Journal of Natural Language Engineering, pp. 15–55.

Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics 32*(4), 543–566.

Brill, E. (2000). Part-of-speech tagging. In: Dale, R., Moisl, H., & Somers, H. (Eds.), Handbook of Natural Language Processing, Marcel Dekker, Inc., New York, pp. 403–414.

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In Proceedings of AAAI-97.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control 2*, 136–167.

Church, A. (1940). A formulation of a simple theory of types. *Journal of Symbolic Logic 5*, 56–88.

Halliday, M. A. K. (1985). An Introduction to Functional Grammar. Edward Arnold, London.

Kaplan, R. M., & Kay, M. (1981). Phonological rules and finite-state transducers. Paper presented at the Annual Meeting of the Linguistic Society of America.

Kaplan, R. M., & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics 20*(3), 331–378.

Lochbaum, K. E., Grosz, B. J., & Sidner, C. L. (2000). Discourse structure and intention recognition. In: Dale, R., Moisl, H., & Somers, H. (Eds.), Handbook of Natural Language Processing, Marcel Dekker, Inc., New York, pp. 123–146.

Mayberry, M. R. & Miikkulainen, R. (2000). Combining maps and distributed representations for shift-reduce parsing. In: Wermter, S., & Sun, R. (Eds.), Hybrid Neural Systems, Springer Verlag, Heidelberg, pp. 144–157.

William C. Morris, William C. Cottrell, Garrison W., & Elman, J. (2000). A connectionist simulation of the empirical acquisition of grammatical relations. In: Wermter, S., & Sun, R. (Eds.), Hybrid Neural Systems, Springer Verlag, Heidelberg, pp. 175–193.

Ratnaparkhi, A. (1996). A maximum entropy part-of-speech tagger. In: Conference on Empirical Methods in Natural Language Processing, Philadelphia, PA, 1996, pp. 133–142.

Samuelsson, C., & Wirén, M. (2000). Parsing techniques. In: Dale, R., Moisl, H., & Somers, H. (Eds.), Handbook of Natural Language Processing, Marcel Dekker, Inc., New York, pp. 59–91.

Steedman, M. (2000). The syntactic process. MIT Press, Cambridge, MA.

Wermter, S. (2000). The hybrid approach to ANN-based natural language processing. In: Dale, R., Moisl, H., & Somers, H. (Eds.), Handbook of Natural Language Processing, Marcel Dekker, Inc., New York, pp. 823–846.

[1] Examples of these are page 53, line 14 'of' missing; page 135, line 4 'represent' should be 'represented'; page 254, line 22 'represent' should be 'represents'; page 374, line 16 'parsers' should be 'parses'; page 491 line 28 the insertion of an extra word 'a'; page 494, line 19 'confused' should be 'confuse'; and page 521, line 17 the insertion of an extra word 'such'.