siderations that here speak to my concerns in providing my utterance with reasons also to speak to yours. Of course, that is an appeal, rather in the spirit of F himself, to the pragmatics of normative utterance, but there is no reason why, if the end-relational theorist can address so many problems with such appeals, the expressivist cannot do the same.

Matters become complicated for F's theory when we have, as we of course very often do, MANY ends that are relevant to the choices we employ normative judgment to make. Often our normative concern is to balance these ends and determine what we should do 'all things considered'. Here the salient end for whoever is the salient person—speaker, auditor, agent—will be, F argues, 'the attainable conjunction of … [her] desired ends that she most prefers' (158). Things get more complicated still when we throw uncertainty into the mix, at which point F suggests that the default end will be whatever maximizes the expected utility of the salient person, though this has to be read de re rather than de dicto, since, of course, no sane person is only interested in whatever maximizes her utility AS SUCH. It is natural to wonder how this account, already to my philosophical ear a little strained at this point, is to extend to those cooperative cases where the salient persons are PLURAL. Do we then take the salient end as their aggregate utilities? But that would seem to build in too much substantive normative theory, and it is hard to see what alternative proposal would avoid doing the same.

But don't we think there are CATEGORICAL normative demands? Certainly we make demands on each other, F suggests, that have what he calls 'pragmatic categoricity'. Here we address each other with normative claims without regard for whether they have any desire for the relevant end. In general, F suggests, normative claims where we fail to explicitly specify the relevant end pragmatically presuppose that we are in a cooperative context where the end is shared. In cases of categorical use, however, this presupposition is false, but we speak as if it were not, so as to endow our utterance with a kind of 'rhetorical objectivity' that puts maximal pressure on the auditor to align her motivation with what we are demanding of them. The last substantial chapter (before a brief concluding one) offers an account of normative disagreement that again leans heavily on pragmatic considerations. F argues that there can be normative disagreement without semantic inconsistency where the desires that his 'quasi-expressivism' allows to be expressed by normative judgments are in conflict.

F's book is a hard and intricate read that will tax any reader. However, it is abundantly worth the effort, easily living up to the very high expectations that those of us who have followed F's earlier writings in the journal literature are bound to bring to it. This is one of the richest, most sophisticated, and most impressive books on metaethics to have been published in my lifetime. Everyone with any interest in normative language ought to read it. Those who would seek to defend reductive naturalist views of the sort F develops here will find it a treasure trove of dialectical resources that they will want to plunder repeatedly. Those who seek to attack such views or to defend rival views will find it a challenge it would be shameful to ignore.

Department of Philosophy
University of Sheffield
45 Victoria St
Sheffield 53 7BQ, United Kingdom
[j.lenman@sheffield.ac.uk]

Automaton theories of human sentence comprehension. By JOHN T. HALE. (CSLI studies in computational linguistics.) Stanford, CA: CSLI Publications, 2014. Pp. 204. ISBN 9781575867472. $27.50.

Reviewed by SASHANK VARMA, *University of Minnesota*

Linguistics and psychology have been intertwined since at least the beginning of the cognitive revolution. In the early 1950s, the psychologist George Miller was busy applying mathematical formalisms—probabilistic (Markov) models and information theory—to understand the sequential structure of sentences (Miller 1951). He abandoned these formalisms under the force of Chomsky's (1957) arguments that sentences are hierarchically structured and joined with him to launch psycholinguistics: the experimental study of how people understand sentences. By the early 1970s, however, linguists and cognitive psychologists had largely gone their separate ways, with the former focusing on competence and the latter pursuing performance.

Beginning in the early 1990s, linguists and cognitive psychologists began to talk once again. A driving force was the maturation of computational linguistics, which gave rise to a number of computational models of human sentence comprehension (Elman 1990, Jurafsky 1996, Just & Carpenter 1992, Lewis 1993, Lewis & Vasishth 2005). These models bridged the divide between linguistic competence (because the models took grammars as 'inputs') and psycholinguistic performance (because the models were evaluated against behavioral data).

In *Automaton theories of human sentence comprehension*, John Hale offers a next step in the intertwined evolution of linguistic and psychological theory. To preview my evaluation, his book is perhaps the most efficient (measured in ideas per page) argument for the development of computational models of sentence comprehension and the evaluation of these models against experimental data. I highly recommend it to linguists and cognitive scientists alike.

Ch. 1 articulates the overall framework of H's approach. It begins with Marr's (1982) explication of the highest level of cognitive science theorizing, the 'computational' level. This is the abstract characterization of a cognitive ability in terms of the function it computes, that is, from inputs to outputs. Marr observed that this level is absent in most cognitive science theories, which focus on the lower levels where the mechanisms that implement this function reside. He identified Chomsky's competence theory of language as one of the few exceptions. The computational level has evolved over the years to include the additional claim that cognitive abilities are OPTIMAL or rational (Anderson 1990). H presents these arguments—and then argues for a return to lower levels. For Marr, one problem with theorizing exclusively at lower levels is that mechanisms are forever underconstrained by data, and this impedes scientific progress. To avoid this problem, H proposes applying optimality to lower levels, taking as his goal the identification of a small number of mechanisms that are individually optimal and collectively minimal and sufficient. Of course, scientists have long used parsimony to guard against unwieldy theories. What differentiates H's approach is his wide-ranging search across disciplines for a set of tightly interlocking mechanisms.

Ch. 2 introduces context-free grammars (CFGs) and argues that they are sufficient for expressing a large subset of sentence structures.[1] H argues that movement and other phenomena that some have analyzed using more powerful formalisms can be handled by CFGs by multiplying the number of internal symbols, as in, for example, generalized phrase structure grammar's 'slash categories'. But he holds the line there, declining to lexicalize the grammar or adopt complex feature structures. This is consistent with his inclination toward parsimony: by keeping the grammatical component simple, he avoids propagating complexity to other parts of his model.

Ch. 3 takes up the parsing component, which is understood as an automaton. H motivates moving beyond simple top-down and bottom-up parsing strategies by considering the implications of three constraints on human sentence comprehension. These are: (i) parsing is incremental—people build up an understanding of a sentence as they read and do not consistently wait to begin until after the last word; (ii) parsing is nondeterministic—there are points during incremental parsing where multiple actions can be taken next (i.e. the parse can be extended in multiple directions). One must be chosen, and if the wrong choice is initially made, the parser must later 'backtrack'; (iii) working-memory (WM) limitations must be respected when handling nondeterminism—not all actions can be taken, nor all parses be constructed. H argues that these three constraints are best satisfied by GENERALIZED LEFT CORNER PARSING—a kind of optimality argument. He adopts a special case of this strategy, a mixed parser that augments a shift-reduce parser with a few top-down elements, going forward.

[1] This is a working hypothesis, and elsewhere H has used mildly context-sensitive formalisms to analyze complex sentence structures (Yun et al. 2015).

Ch. 4 turns to the experimental data on human sentence comprehension. The goal is not to evaluate the grammatical formalism and parser that have been advanced. Rather, it is to show that these mechanisms can naturally exhibit classic psycholinguistic principles for how people handle the incremental, nondeterministic, and WM-limited nature of sentence comprehension. At an ambiguous point during incremental parsing, a nondeterministic choice must be made. Garden pathing is when the incorrect choice is made, and when later input is inconsistent with this choice, the parser must 'backtrack' and make a different choice. This phenomenon is exemplified by the famous sentence *The horse raced past the barn fell*. Most readers initially choose the past-tense interpretation of *raced* and are boggled when they encounter the sentence-final verb *fell*. Explaining this phenomenon requires explaining how the parser chooses, at ambiguous points, which of multiple actions to take (i.e. extensions of the current parse to pursue). H first introduces two classic psycholinguistic principles for making this choice, MINIMAL ATTACHMENT and LATE CLOSURE. He then shows how they can be expressed as policies within the mixed parser developed in Ch. 3 for choosing between shifting and reducing when both operations are possible. In other words, he shows how these higher-level theoretical principles are consequences of lower-level CONTROL decisions of the parser.

The account offered thus far is computational—CFGs are formal languages and parsers are automata. Ch. 5 evaluates its cognitive validity by exploring its implementation within cognitive architectures. These are cognitive science theories of the basic computational mechanisms of the mind—its representations, processes, and control structures. In other words, they are theories at Marr's lowest level. H focuses on the implementation of the mixed parser developed in Ch. 3 within Soar, the cognitive architecture of Newell (1990). That this is possible is perhaps not surprising given the common intellectual origin of CFG rules and Soar's 'production' rules in Post's (1943) theory of computation. What is surprising is that the control policies of the mixed parser find a parsimonious implementation in the control structure of Soar—in the 'preference' representations it uses to decide which of multiple production rules to execute next.

Ch. 6 takes a fresh look at the control structure of the parser through the lens of AI. It begins by considering the failure of minimal attachment, late closure, and other psycholinguistic principles, applied in a top-down manner, to account for the flexibility with which people incrementally comprehend ambiguous sentences. It proposes instead that the control structure of the parser is bottom-up and learned from experience. In AI, problem solving is formalized as searching from an initial state through a sequence of intermediate states to a final state, where each step is taken by applying an operator that transforms the current state into the next state. H reconceptualizes parsing as problem solving where the initial state is null, the intermediate states represent incomplete parses, the final state represents a complete parse (i.e. a tree headed by S that covers all of the words of the sentence), and the operators are CFG rules. He introduces A*, an AI algorithm for efficiently finding near-optimal paths. A* uses a heuristic function to choose the next step to take based on (i) the cost of the path from the initial state to the current state, which is known, and (ii) the cost of the path from the current state to the final state, which must be estimated. He proposes that these estimated costs can be learned from experience, that is, from parsing through a Treebank. Critically, these estimated costs dynamically control the parser in a bottom-up manner. When the heuristic function is relatively certain about which path is optimal, the parser behaves as if late closure is applied in a serial manner. However, when the heuristic function is relatively uncertain, the parser smoothly shifts to parallel processing and develops multiple paths/parses simultaneously.

Ch. 7 shifts from the lower-level, mechanistic account of Chs. 2–6 to a higher-level, mathematical analysis of what makes sentence comprehension difficult. At this level, sentence comprehension can be viewed as optimal. To describe this, H turns to information theory, which was developed in the late 1940s to guide the design of optimal codes for communicating messages. As noted above, the view that sentences are sequentially structured was abandoned early in the cognitive revolution in the face of Chomsky's (1957) arguments that they are hierarchically structured. Probabilities were reintroduced into linguistics in the 1990s, in the form of probabilistic CFGs (PCFGs) where each rule is annotated with a probability estimated from a Treebank (Charniak 1993). Information theory was reintroduced soon thereafter (Hale 2001). Here, H shows how information theory can be applied to the mechanistic operation of a probabilistic parser to explain sentence comprehension difficulties. At the current word, there is a set of possible parses, each with a probability computed from the parsing operations (ultimately the PCFG rules) that generated it. SURPRISAL at the current word reflects the change in the sum of these probabilities (which is not necessarily 1) from the prior word, expressed in information-theoretic bits. ENTROPY REDUCTION at the current word reflects the change in the distribution of these probabilities from the prior word, again expressed in bits. H evaluates both metrics against reading-time data, and tentatively concludes that entropy-reduction offers the superior account. The chapter concludes by noting that these information-theoretic measures of comprehension difficulty are offered at Marr's highest level and are ultimately independent of the lower-level parsing mechanisms developed in prior chapters.

The information-theoretic measures in Ch. 7 require some way to assign probabilities to parses. Ch. 8 develops a novel account of where these probabilities come from. It begins with the POWER LAW OF PRACTICE, which governs the speed-up of cognitive skills (such as parsing) with experience (Newell & Rosenbloom 1981). One explanation of this law is that cognitive skills can be decomposed into sequences of operations, and that frequently occurring subsequences of operations (which take multiple time steps to perform) can be CHUNKED into single macro-operations (which takes only a single time step to perform). H proposes chunking as the mechanism by which frequency/probabilistic information makes its way into sentence comprehension. He applies parsers to treebanks, chunks frequently occurring subsequences of parser operations into parser macro-operations, and evaluates these expanded parsers against eye-tracking data. He shows that the more likely it is that a word is parsed using a learned macro-operation, the faster it is read (as measured by eye-tracking data). He speculates that chunking bridges between generative and cognitive linguistics—although the operations of the parser originate in generative grammar, the chunked macro-operations are closer to constructions.

Ch. 9 concludes by briefly reviewing the argument developed in the book and by sketching directions for future research.

Automaton theories of human sentence comprehension provides a computational account of human sentence comprehension. In doing so, it follows in the tradition of researchers such as Berwick and Weinberg (1984), Pritchett (1992), and Townsend and Bever (2001). It is formal, elegant, and thoroughly modern. It cuts a clean line through multiple disciplines—linguistics, psychology, computer science, and statistics. The result is a unified account that will be valuable to linguists and cognitive scientists interested in the architecture of sentence comprehension.

Although it is not a stand-alone textbook, it could form the basis for a course on computational linguistics for linguistics or psychology students. It is clearly written, provides an excellent glossary, and includes a well-curated set of references. If used in this role, it would have to be supplemented by tutorial readings on the mathematical and computational formalisms utilized by H's model, and by experimental papers reporting the data that are its target.

REFERENCES

ANDERSON, JOHN R. 1990. *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

BERWICK, ROBERT C., and AMY S. WEINBERG. 1984. *The grammatical basis of linguistic performance: Language use and acquisition*. Cambridge, MA: MIT Press.

CHARNIAK, EUGENE. 1993. *Statistical language learning*. Cambridge, MA: MIT Press.

CHOMSKY, NOAM. 1957. *Syntactic structures*. The Hague: Mouton.

ELMAN, JEFFREY L. 1990. Finding structure in time. *Cognitive Science* 14.179–211. DOI: 10.1207/s15516709cog1402_1.

HALE, JOHN. 2001. A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the 2nd meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 159–66. Online: http://www.aclweb.org/anthology/N/N01/N01-1021.pdf.

JURAFSKY, DANIEL. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20.137–94. DOI: 10.1207/s15516709cog2002_1.

JUST, MARCEL A., and PATRICIA A. CARPENTER. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99.122–49. DOI: 10.1037/0033-295X.99.1.122.

LEWIS, RICHARD L. 1993. *An architecturally-based theory of human sentence comprehension*. Pittsburgh: Carnegie Mellon University dissertation.

LEWIS, RICHARD L., and SHRAVAN VASISHTH. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29.375–419.

MARR, DAVID. 1982. *Vision: A computational investigation into the human representation and presentation of visual information*. New York: W. H. Freeman.

MILLER, GEORGE A. 1951. *Language and communication*. New York: McGraw-Hill.

NEWELL, ALLEN. 1990. *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

NEWELL, ALLEN, and PAUL S. ROSENBLOOM. 1981. Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, ed. by John R. Anderson, 1–55. Hillsdale, NJ: Lawrence Erlbaum.

POST, EMIL L. 1943. Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics* 65.197–215. DOI: 10.2307/2371809.

PRITCHETT, BRADLEY L. 1992. *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

TOWNSEND, DAVID J., and THOMAS G. BEVER. 2001. *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

YUN, JIWON; ZHONG CHEN; TIM HUNTER; JOHN WHITMAN; and JOHN HALE. 2015. Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics* 24.113–48. DOI: 10.1007/s10831-014-9126-6.

56 East River Road
250 Education Sciences Building
Minneapolis, MN 55455
[sashank@umn.edu]

New perspectives on Chinese syntax. By WALTRAUD PAUL. (Trends in linguistics, studies and monographs 271.) Berlin: De Gruyter Mouton, 2015. Pp. xvi, 357. ISBN 9783110338683. $154 (Hb).

Reviewed by JINGLIAN LI, *Beijing Institute of Technology*, and
FANJUN MENG, *Tsinghua University*[1]

Mandarin Chinese, as an isolating language, presents considerable challenges to the syntactic study of word-order typology, the inventory of lexical categories, sentence-peripheral elements like topic, sentence-final particles (SFPs), and so on. Waltraud Paul's book *New perspectives on Chinese syntax* boldly takes on these controversial and recalcitrant issues and aims to reaccommodate them into a broader linguistic landscape. In short, P has managed to reconcile some noisy facts in Chinese and has cogently falsified some pervasive, long-term conceptions in Chinese syntax.

The book is organized into eight chapters. The first chapter sets the stage for the book and provides a synopsis of the subsequent chapters. In Ch. 2, word order in Chinese is submitted to typological scrutiny from a historical perspective. Chinese is argued to have undergone two major word-order changes: OV > VO > OV (see Li & Thompson 1974). Specifically, pre-Archaic Chinese originated as an SOV language and then changed to SVO between the tenth and the third centuries BC, further followed by a shift back to SOV, which is still in progress in Modern Mandarin. P argues against this and, based on the Shang inscriptions (see Djamouri 1988), shows that even in pre-Archaic Chinese, the dominant word order is SVO instead of SOV. The so-called SOV in pre-Archaic Chinese, involving focalization of the object and object pronouns in negated sentences, turns out to be de facto the head-complement configuration, thus consistent with the SVO order. The core evidence for Li and Thompson's (1974) hypothesis of SOV order in Modern Chinese is

[1] This work is supported by 10BYY003 (National Social Science Fund of China).