This volume consists of a collection of solicited articles, presented as a tribute to Gösta Bruce, one of the leading researchers in the field of prosody. The papers are for the most part review articles (as perhaps befits a volume of this kind), and the volume thus serves an important function in prosody research in giving a broad snapshot of the field as it stands at the dawn of the new millennium. The coverage of the volume is indeed broad, including such areas as intonation inventory (Pierrehumbert, Ladd, Hirst et al., Gussenhoven), tune-text alignment (Pierrehumbert, Ladd, Gussenhoven), acoustic correlates of accent and stress (Terken et al., Beckman et al.), prominence (Terken et al., Ladd), prosodic structure and juncture (Hirst et al., Shattuck-Hufnagel, Selkirk, Ostendorf), timing (Beckman et al., Campbell), and differences between speaking styles (Hirschberg). The book is successful in providing readers with a reasonably clear perspective on what the relevant questions are, what kind of evidence is available to answer those questions, and who has done what. The authors hold in common the view that to understand prosody one needs to investigate abstract representations: one studies intonational categories and prosodic structures, rather than, for instance, vectors of f0 values.

The book starts with Horne’s introduction, which introduces the articles, and puts them in perspective, especially with respect to the work of Gösta Bruce.

Pierrehumbert’s article is an excellent survey of her own model of intonation, and would be very suitable for an introductory course on intonation: in forty pages one gets a condensed version of the historical development of a major intonational framework. It outlines the major issues that have changed the views of intonation researchers from the structuralists to today, with a section describing Gösta Bruce’s influential work on Swedish intonation, as well as his contribution to the development of intonation theory embodied in the concept that sentence intonation can be represented as a sequence of tones.

In Pierrehumbert’s intonation system, one of the central theoretical issues is how to interpret continuous phonetic data and link it to discrete phonological categories. She reviews arguments for the two-level representation of tones, using only H (high) and
L (low), as well as the arguments surrounding the assignment of accent categories. Downstep data provides a crucial argument against multiple levels of tonal representation. The argument goes as follows: Each successive downstepped tone clearly sounds distinct, and in some languages the distinctions are phonological. It can also be demonstrated that one can have as many downstepped pitch levels as are allowed by performance factors such as utterance length. If one were to represent each distinct level phonologically, then the conventional four levels proposed early on by structuralists, or any predetermined number of levels for that matter, are not enough. A more coherent treatment is to use minimal phonological levels—in this case, H and L—to implement the phonemic distinctions and phonetic models to predict gradient pitch height. The downstep data has been successfully treated in this way.

The same philosophy applies to the classification of accent inventories. Nuclear accent and pre-nuclear accent in English declarative sentences have different peak alignment patterns. Is it possible that they belong to the same underlying accent category? Pierrehumbert suggests that the answer is yes and reviews arguments supporting this treatment. Many factors affect the alignment pattern of the f0 peak with the stressed syllable. But there is no need to posit a different category whenever one sees surface variations, so long as the surface variations can be predicted by a phonetic model from a unique phonological representation.

It is interesting to read Ladd’s paper right after Pierrehumbert’s. Ladd plays devil’s advocate and raises three questions directly addressing the issue of intonation representation in Pierrehumbert’s framework:

1. What is a tone? There is a level of abstraction in Pierrehumbert’s tonal inventory involving mismatches in the assignment of H and L on the one hand, and turning points in surface pitch contours on the other hand. Should one take the turning points in surface intonation contour more seriously?

2. What is the meaning of the starred tone, as in the “*” of H*+L or L*+H? Ladd suggests that the starred tone should be the landing site of emphasis, and he expects H* to go higher and L* to go lower under emphasis. Under this assumption, he raises the issue of Bruce’s (1987) re-analysis of his Swedish data, which assigned H+L* to accent I and H*+L to accent II. Fant and Krukenberg (1994) reported that, under focus, the pitch of the low tone in accent II (H*+L) is lower, while the pitch of all targets in accent I (H+L*) are raised.

3. Is the tune-text association convention as mediated by the star really meaningful? Ladd raises the question of Greek L+H tone, where L is aligned right before the stressed syllable and H is aligned right after the stressed syllable. Ladd’s concern is that, in this case, nothing is actually aligned with the stressed syllable. It is also implied that in languages such as English or Greek, the tune-text association may not be as rigid as in a tone language such as Yoruba (see below).

Ladd’s questions reiterate the theme raised by Pierrehumbert on the difficulty of separating phonetic variation from phonological representations. In a sense, the Greek L+H case remains true to the autosegmental spirit of tune-text association: the star of the text provides an anchor, and alignment of the tune is made with reference to this anchor. As long as the facts are clear, an alignment model can be built successfully, referring to the whole contour or to any number of points along the contour. The
alignment model can be quite complicated (see van Santen and Möbius [2000]) and the predicted tone landing site can be quite far from the anchor.

In addition, it turns out that the tune-text misalignment in a tone language can be substantially more dramatic than the Greek case described by Ladd. In Mandarin Chinese (Shih and Kochanski 2000; Xu 2001), it is not uncommon to have a tone target shifted completely off the syllable it originates from. In Yoruba (Laniran 1992), a target can be delayed by several syllables. In a tone language, both the tonal inventory and the ideal, phonological tune-text alignment are known. So when the tune and text alignment are off by a few syllables, one has no choice but to acknowledge the misalignment and to zero in on an alignment model predicting the alignment pattern. When dealing with a non-tone language, it is not as easy to ascertain a case of long-distance misalignment and one often ends up with a phonological analysis that is closer to surface observables.

The paper of Hirst, Di Cristo, and Espesser is an overview of their intonation model, providing a contrast to the ToBI-based articles in the book. Theirs is an automated intonation model used for speech analysis and synthesis. The MOMEL algorithm analyzes f0 curves, smoothing out some of the micro-prosody and finding target points in the f0 contours. The target points are converted to the INTSINT transcription system. Basically, INTSINT defines the highest and lowest targets in the utterance as H (high) and B (bottom), respectively. Other phrase-initial targets are labeled as M (mid). The rest of the targets in the utterance are assigned labels reflecting relative relations with preceding and following targets: L (low), U (up), D (down), S (same). It should be apparent from this description that the analysis part of the system is language-independent: it depends on speech signals rather than on language knowledge. For the purpose of synthesis, some level of language knowledge is needed to write or to train a language-dependent intonation grammar. The f0 contours will then be generated from the targets predicted by the grammar.

Terken and Hermes provide a comprehensive review of the literature on prominence perception; readers get a clear sense of what questions should be asked and have been asked, even though the answers may not always be clear. One central issue addressed in this article is the question of what makes two accents sound equally prominent. Interesting experiments have been done comparing different pitch accents, as well as pitch accents in different pitch registers (e.g., male vs. female), in different positions of a sentence, and with or without a declining baseline. Technical questions revolve around what the correct thing is to measure (pitch excursion, or pitch-level difference), and in which scale (Hz, BART, Mel, or ERB). The data support a general declination model, in which early accents are bigger and higher than later, equally prominent accents. The tilt of the baseline has an effect on the perception of prominence. Among different pitch accents with equal excursion size, falling pitch accents lend more prominence than rising or rise-fall accents.

Gussenhoven’s contribution presents new work on a lexical tone contrast in the Roermond dialect of Dutch. Roermond Dutch, like other Limburg dialects, has a lexical accentual system reminiscent of that of Swedish. There are two types of accent. Words with Accent I have no lexically prespecified tone. Accent II words, following Gussenhoven’s analysis, have a high (H) tone linked to the second mora of the accented syllable: note, therefore, that Accent II can only occur in words where the syllable that would bear the accent is bimoraic. Accent II displays a number of interesting features. First, if no boundary tone and no pitch accent is associated with the accented syllable, the H does not surface. Second, if a low (L) pitch accent is associated with the syllable, the H transmutes to a L so that you get a sequence L*H. Third, and most interesting, a boundary tone—L or H1L1—apparently shows up before the Accent II high. So an
“underlying” sequence such as H*HL₁—where H* is an intonationally assigned pitch accent, H is the Accent II tone, and L₁ is the boundary tone—surfaces as an HLH tone sequence rather than the expected HHHL. As Gussenhoven notes, this is the first documented instance of a boundary tone being anything other than peripheral, and it has somewhat the flavor of morphological infixation.

Gussenhoven argues that a derivational account—what he terms an “SPE” (Chomsky and Halle 1968) account—would lead to an ordering paradox between two rules, both of which one would seemingly need: the rule that transmutes L*H to L*L, and the metathesis rule that reorders the boundary tone before an Accent II H. Note that in this derivational account, Gussenhoven assumes that the metathesis rule is a transformational rule that refers to the tone sequence and the boundary, and nothing else. Having rejected such an account, Gussenhoven presents a constraint-based analysis within the Optimality Theory (OT) framework, making use of about ten ranked constraints. Central to the boundary tone reordering is the assumption of two constraints, one (ALIGNT₁RT) that states that the boundary tone wants to align to the right of its phrasal domain, and the other (ALIGNLEXRT) which states that the Accent II H wants to align to the right of its syllable, which of course coincides with the right of the phrasal domain if that syllable is final. The ranking ALIGNLEXRT ≻ ALIGNT₁RT achieves the desired result that the Accent II H comes after the boundary tone.

Though Gussenhoven presents what seems to be a particularly compelling argument for OT, the Roermond data are actually grist for any number of theoretical mills. For example, taking Gussenhoven’s proposal for a lexical H tone at face value, one could explore the possibility of a more traditional autosegmental analysis: the straw man derivational analysis that Gussenhoven presents is hardly fair to the quarter of a century of phonology between SPE and the advent of OT. Then there is the possibility of not taking Gussenhoven’s analysis at face value. Indeed, two properties of Accent II—the transmutation of the H to L after L*, and the complete loss of H in accent-free non-boundary contrasts—suggests the possibility that Accent II may not involve a lexically specified H at all, but rather merely a different timing specification for whatever accent (if any) gets associated with the syllable. Such an approach would not be without complications, but it seems nonetheless worth exploring.

Beckman and Cohen’s is the second of the articles in this book to report new data. This article is a follow up of earlier work by Beckman and Edwards (1994) on the differences in the jaw-opening movements of two types of lengthening: lengthening for accent, and phrase-final lengthening. The data consist of the syllable pop, stressed/accented, stressed/unaccented, and unstressed. These stimuli are embedded in phrase-final as well as non-final positions. Beckman and Cohen consider three articulatory models to account for the jaw-tracing differences between full and reduced vowels: a truncation model, a rescaling model, and a hybrid model. The preliminary analysis suggests that the hybrid model works best. This study supports earlier findings that not all lengthenings are accomplished in the same way. Contrasting with an unstressed syllable, a stressed syllable is longer, and has a more extreme displacement of the jaw, but with higher velocity in the movement; this is not accounted for by the truncation model, which predicts the velocity of the movement to be the same. In phrase-final position, lengthening is accompanied by slower movement.

Shattuck-Hufnagel’s paper presents three types of arguments, based on stress shift, glottalization, and rhythmic pattern in speech—with data obtained from a speech corpus—to support the prosodic planning hypothesis, namely that speech production is planned with reference to prosodic structure. She suggests that the so-called stress-shift rule is not really a rule that shifts stress to avoid stress clash. The main argument is that the “stress shift” effect may be achieved by the addition of an accent without
shifting the original stress; it may also occur when there is no stress clash. There is a
tendency for speakers to use a pair of accents to frame a prosodic phrase. The first
landing site is the earliest full vowel in the phrase, and the last one is the nuclear
accent. The stress-shift rule can be subsumed under this mechanism. Glottalization
of a vowel-initial word such as *apple* is more common in phrase-initial position. Also,
phrase-final position is frequently marked by glottalization. The use of accents and
glottalization both have the effect of framing a prosodic phrase. Further evidence for
prosodic structure comes from the preference for the use of alternating stress. Although
sentences in natural speech often do not show strictly alternating stress, comparisons
of sentences with and without rhythmic patterns show that sentences with rhythmic
patterns are easier to produce and less prone to speech errors.

Selkirk presents an interesting analysis of English phonological phrasing in terms of
OT. She starts by reviewing evidence from Bantu languages for the universality
of two constraints, namely ALIGNR XP, which states that “the right edge of any XP
(maximal projection) in syntactic structure must be aligned with the right edge of a
MaP (major projection) in prosodic structure”, and a constraint proposed by Trucken-
brodt (1995), WRAP XP, which states that “the elements of an input morphosyntactic
constituent of type XP must be contained within a prosodic constituent of type MaP
in output representation”. Since the first constraint requires the right edges of MaPs
to align with the right edge of each XP, whereas the second requires all XPs to be con-
tained within a (single) MaP, the two constraints are inherently in conflict. In English,
it seems that there is no evidence for ranking between ALIGNR XP and WRAP XP.
This is because a sentence such as *She loaned her rollerblades to Robin*—where, crucially,
each of the words *loaned*, *rollerblades* and *Robin* are accented—can be phrased as ei-
ther (*She loaned her rollerblades*)MaP *(to Robin)*MaP, with two MaPs, or (*She loaned her
rollerblades to Robin*)MaP, with one. The first case violates WRAP XP, the second violates
ALIGNR XP, but both violations seem to be equal. What is not accounted for is the
failure of an “overphrased” version, namely (*She loaned*)MaP *(her rollerblades)*MaP *(to Robin)*MaP,
which is a violation of WRAP XP, but not ALIGNR XP. This motivates the
introduction of a third, lower-ranked constraint BINMAP, which requires MaPs with
just two accentual phrases; the overphrased version then has three violations of this
constraint.

Selkirk then turns to ALIGNR FOCUS, which requires the alignment of a MaP
with a focused constituent, and she argues that it is higher ranked than the other con-
straints. Thus a focused version of *She loaned her rollerblades to Robin* has one optimal
candidate: (*She loaned*)MaP *(her rollerblades)*MaP *(to Robin)*MaP. Here, the boundary af-
s ter *loaned* is favored by ALIGNR FOCUS, and the boundary after *rollerblades* is favored by
ALIGNR XP.

Selkirk’s data are based on intuitive judgments concerning the putative presence
or absence of a MaP boundary tone, the diagnostic she adopts (following Beckman and
Pierrehumbert [1986]) for deciding whether a phrase boundary is present. It would be
interesting to see to what extent these data hold up under experimental conditions.

Ostendorf presents a brief review of linguistic and engineering issues related to
the automatic detection of prosodic boundaries. As she points out, automatic bound-
ary detection is desirable in a number of areas of speech technology. For example,
in speech recognition and understanding, prosodic information can, in principle, be
used to prune the search space (since some hypotheses are likely to be incompatible
with a given phrasing) and to score linguistic hypotheses. Automatic detection is also
desirable in text-to-speech synthesis to aid in the rapid development of prosodically
labeled databases. A holy grail of this enterprise, as Ostendorf notes, is an approach
that is robust enough to work on spontaneous speech; current automatic phrase de-
tection methods work well only on read speech and perform considerably less well on the kind of conversational speech found in the Switchboard corpus.

Campbell presents an overview of work on duration modeling, starting with a rather detailed account of Klatt’s (1973) rule-based model, and covering various statistical approaches such as Riley’s (1990) CART-based models and van Santen’s (1994) sum-of-products models. Campbell’s own view is that segment-based models of duration are misguided because they are based on the notion of a segment’s “inherent duration,” and that instead one should model higher levels of prosodic structure (syllables, feet, or even prosodic phrases), deriving segmental durations once the higher-level durations are set up. The second half of the article describes earlier work of Campbell that provides support for a syllable-based approach from English and Japanese.¹

Hirschberg concludes the book with a review of earlier work on prosodic cues that differentiate speaking style—or, more properly, two particular speaking styles, namely read speech and spontaneous speech. She catalogs differences in rate (read speech is faster), differences in the distribution of different boundary tones, and differences in the rates of disfluencies, as well as a few other factors. Disfluencies, while more common in spontaneous speech, as one might expect, are nonetheless sufficiently rare in both styles that they are not a particularly useful cue to distinguishing the two.

Finally, a word on production quality, which unfortunately is mediocre. There are a variety of problems, particularly with the presentation of some of the figures and the equations. So, the shaded areas of the tableaux in Selkirk’s paper are too dark, though the ones in Gussenhoven’s paper are fine. In at least a couple of the papers—Gussenhoven, Campbell—there are some quite annoying changes in font size between successive linguistic examples or equations. On the whole the production quality is not what you would expect for a volume that lists at over US$150. But there is presumably nothing to do here but lament the fact that as academic publishers continue to up the prices of their wares, they also seem to be taking less and less care in their production.

References


¹ In work published about the same time as this book, van Santen and Shih (2000) present arguments against the syllable-based approach as espoused by Campbell.


Chilin Shih is a Research Scientist at Bell Laboratories, Lucent Technologies. She has built text-to-speech systems for many languages. Her current research focuses on tone and intonation modeling. Shih’s address is 600 Mountain Avenue, Murray Hill, NJ 07974; e-mail: cls@research.bell-labs.com. Richard Sproat is a Technology Consultant at AT&T Labs – Research. He has done research in computational morphology, writing systems, and text-analysis for text-to-speech synthesis, where he has worked on accent prediction and multilingual text analysis. Sproat’s address is 180 Park Avenue, Florham Park, NJ 07922; e-mail: rws@research.att.com.
Lexicon Development for Speech and Language Processing

Frank Van Eynde and Dafydd Gibbon (editors)
(University of Leuven and University of Bielefeld)

Dordrecht: Kluwer Academic
Publishers (Text, speech and language
technology series, edited by Nancy Ide
and Jean Véronis, volume 12), 2000,
xi+298 pp; hardbound, ISBN
0-7923-6368-X, $128.00, £79.00, €109.00

Reviewed by
Ken Litkowski
CL Research

As a computational lexicologist with little background in speech technology, I approached this book as an opportunity to gain a basic understanding of how lexicons are used in speech and to see whether semantically oriented lexicons from my work could bring something to this field. The first objective was met, but the second was not. This book arose from the 1997 Fifth European Summer School on Language and Speech Communication under the auspices of the European Language and Speech Network (ELSNET) and is the report of nine of the ten lectures. The book offers a “survey of methods and techniques for structuring, acquiring and maintaining lexical resources for speech and language processing” (p. ix).

As part of my first objective, I learned that speech technology is greatly hampered by limitations on the size of the lexicon that can be handled efficiently in real systems. The amount of data is overwhelmed by the need to use signal data in analog or digital form. The speech community has to resort to and develop many tricks for working with the prodigious amount of data. This book provides an appropriate overview of the complexity of the problems and will serve well as background reading in an introductory computational linguistics course. Each of the papers is well referenced, with the authors providing their (important) guides to further details.

The organization of the book is not optimal, however. It consists of an overview, two papers on lexical formalisms, two research papers, three database papers, and a final research paper. Logically, the database papers should be first, followed by the research papers, the lexical formalisms, and finally the overview. The quality of the papers and the editing is high; the contributors clearly worked to turn their lectures into readable form. I will present comments based on the logical order, rather than the physical order.

Christoph Draxler (“Speech databases,” Chapter 6) provides an overview of what kind of database technology is required in speech technology, albeit with little reference to the lexicon. This paper shows the impressive range of speech material, describing recording mechanisms and annotation levels. Silvia Quazza and Henk van den Heuvel (“The use of lexica in text-to-speech systems,” Chapter 7) outline practical steps for putting together the components of a text-to-speech system and for building the lexicon, showing quite well where the lexicon comes into play. This chapter has the most detailed information about what lexical information is used in speech systems and how. Martine Adda-Decker and Lori Lamel (“The use of lexica in automatic speech recognition,” Chapter 8) carry us through the design and development of the lexical information used in recognition, describing what is required of the lexicon for de-
coding speech. It is here that the size of the lexicon evinces the complexities and is likely the place where ingenious (semantic) solutions are most needed for recognizers to “understand” speech. These three papers provide the backbone of the book.

The three research papers represent examples of ongoing research; all are presented quite well, giving the reader a sense of active problems. Walter Daelemans and Gert Durieux (“Inductive lexica,” Chapter 4) describe their work on inducing regularities implicit in phonological lexical representations using “memory-based learning” machine learning techniques. The methods followed are clearly laid out, providing an introduction to the authors’ research, which may be followed via their references. The other two research papers do not have a close tie to speech. R. Harald Baayen, Robert Schreuder, and Richard Sproat (“Morphology in the mental lexicon: A computational model for visual word recognition,” Chapter 9) present investigations on the relative time of lexical processing for morphologically complex words; such research may eventually help in the design of computational lexicons for speech processing, via activation of potentially matching lexical candidates. Gregory Grefenstette, Anne Schiller, and Salah Ait-Mokhtar (“Recognizing Lexical Patterns in Text”, Chapter 5) provide an introduction to finite-state automata for recognizing compound noun patterns in building a lexicon; in addition to its introductory pedagogic value, the paper describes a fully developed system.

Gosse Bouma, Frank Van Eynde, and Dan Flickinger (“Constraint-based lexica,” Chapter 2) provide an introduction to the HPSG formalism. Although the presentation is clear, there is no tie to speech or to speech systems. Lynne Cahill, Julie Carson-Berndsen, and Gerald Gazdar (“Phonology-based lexical knowledge representation,” Chapter 3) present a tutorial on the DATR formalism, valuable in itself, but more importantly they demonstrate how this formalism can be used for phonetic or phonological representations. There are 34 “exercises” in the tutorial; however, these may not be as useful in the book as they would have been in the summer school.

The introduction and overview to the book (“Computational lexicography,” by Dafydd Gibbon), which needs to be read both first and again after reading the other chapters, brings together the papers in the volume under the author’s development of an “integrated lexical sign model.” This is a little forced, but useful for providing an overview of the lexicon in speech processing. However, there is a strong element of Unix hacking, without a comprehensive view of traditional computational lexicology and no reference to computational lexicography as may be practiced by dictionary publishers (e.g., corpus evidence).

As to the second objective stated above, the incorporation of semantics into speech seems a long way off. However, the dialogue can begin.

Ken Litkowski is a computational lexicologist with CL Research (http://www.clres.com). He is currently performing computational lexicography tasks for dictionary publishers and performing computational lexicology research on how computational lexicons from machine-readable dictionaries and thesauruses can be used in NLP applications, including word-sense disambiguation and question answering. Litkowski’s address is: CL Research, 9208 Gue Road, Damascus, MD 20872; e-mail: ken@clres.com.
Advances in Probabilistic and Other Parsing Technologies

Harry Bunt and Anton Nijholt (editors)
(Tilburg University and University of Twente)

Dordrecht: Kluwer Academic
Publishers (Text, speech and language
technology series, edited by Nancy Ide
and Jean Véronis, volume 16), 2000,
xv+267 pp; hardbound, ISBN
0-7923-6616-6, $112.00, £71.00,
Dfl 230.00

Reviewed by
Chris Brew
The Ohio State University

This book is an edited selection of papers presented at the Fifth International Workshop on Parsing Technologies, held at MIT in September 1997. Several of the papers are already well-known and others should be. The book could easily be used as the basis for a graduate-level advanced course on parsing. The title is unwieldy, but appropriate: most but not all of the papers have a strong probabilistic flavor.

My favorite papers are Erik Hektoen on “Probabilistic parse selection based on semantic co-occurrences,” Jason Eisner on “Bilexical grammars and their cubic-time parsing algorithms,” and Chris Manning and Bob Carpenter on “Probabilistic parsing using left corner language models.” I like these papers because they step back from the details of parsing technology and consider its wider significance.

Manning and Carpenter offer both detail and overview. They provide a series of probabilistic models that relax the context-freeness assumption of probabilistic context-free grammars, measure performance in the usual way, draw appropriate conclusions, then provide the kicker in the form of a brief section explaining “Why parsing the Penn Treebank is easy.” As Manning and Carpenter point out, in the particular case of the Penn Treebank, the currently accepted PARSEVAL metrics (Grishman, Macleod, and Sterling 1992) are actually quite easy to do well on, even if the system makes systematic errors on such things as prepositional-phrase attachment. If systems are to be deployed into situations where such deficiencies might matter, it might be necessary to find more appropriate evaluation methods. This issue has subsequently been addressed by others (Carroll, Briscoe, and Sanfilippo 1998; Carroll, Minnen, and Briscoe 1999), who argue for more obviously task-related evaluation schemes involving predicate argument structure and/or dependency information.

Hektoen’s contribution is in the same vein; it takes seriously the notion that parsing is often simply a device for getting at an underlying semantics. Under his scheme, parse selection relies on the ability to collect statistics over semantic forms. Following this path leads Hektoen into a careful exposition of a Bayesian-estimation approach to parse selection, which appears to be “a sufficient response to the high degree of sparseness in the lexical co-occurrence data without the blurring associated with smoothing and clustering” (p. 162). Hektoen’s approach appears to work well; of course, it does require a broad-coverage parser capable of generating semantic representations, which may be an obstacle for many. The exposition of the method is very clear and the comparison with previous approaches is enlightening.
Mark-Jan Nederhof’s “Regular approximation of CFLs: A grammatical view” is similar to Eisner’s contribution in that its focus is primarily mathematical. It describes an attractive approach to finite-state approximation of regular grammars. The essential idea is to characterize properties that make grammars non-regular, and to develop schemes for systematically removing such properties. This helps to keep the approximation process perspicuous. Experimental work with this approximation scheme is absent from the current article, but is reported elsewhere (Nederhof 2000).

In “Probabilistic GLR parsing,” Kentaro Inui, Virach Sornlertlamvanich, Hozumi Tanaka, and Takenobu Tokunaga provide a careful analysis of the process of LR parsing. This leads to a probabilistic parsing scheme having the desirable property, not previously achieved for LR parsers, that the sum over all parses of the probability is unity. Once again experimental work is not present here but is reported elsewhere (Sornlertlamvanich, Inui, Tokunaga, Tanaka, and Takezawa 1999).

Eisner’s paper does not report experiments either, but addresses a problem with profound practical significance. It analyses the computational properties of grammars in which potentially idiosyncratic word-to-word relationships play a key role. The framework used is general enough to capture the essence of many recent statistical parsers and clean enough to make it easy (and interesting) to compare one with another. I like Eisner’s paper for the insight it provides into the options available to the lexically minded probabilistic modeler. This aspect is also present in “Encoding frequency information in lexicalized grammars,” where John Carroll and David Weir, using lexicalized tree adjoining grammar (LTAG) as an example, analyze the problem of providing practically useful estimates of the large number of parameters that are potentially present in lexicalized grammars. Similarly, in “Towards a reduced commitment, D-theory style TAG parser,” John Chen and K. Vijay-Shanker describe an approach to TAG parsing whose goal is to delay attachment decisions. This is a design sketch, not an implemented parser, but the design is well fleshed out, and looks worth testing.

Several articles do have extensive evaluation data. Joshua Goodman contributes “Probabilistic feature grammars,” developing an implemented and efficient stochastic feature-based grammar formalism. The key idea, prefigured in, for example, Stolcke’s (1994) doctoral dissertation, is to choose a feature formalism that does not impede dynamic programming implementations of the usual inside, outside, and Viterbi probability calculations. Goodman includes extensive quantitative evaluation, which is greatly to be welcomed. “A new parsing method using a global association table” by Juntae Yoon, Seonho Kim, and Mansuk Song, is a description and evaluation of a semi-deterministic parsing algorithm designed to exploit the fact that Korean is an SOV language with many surface cues to syntactic dependency. Extensive evaluation is provided. Bangalore Srinavas’s “Performance evaluation of SuperTagging for partial parsing” exploits the author’s SuperTagging idea (i.e., employing part-of-speech–tagger technology to “almost parse,” using the elementary trees of lexicalized tree adjoining grammar) for the now-standard task of partial parsing. Given the title, the plethora of interesting performance figures is to be expected. For example, connecting to the discussion of the Penn Treebank above, Bangalore reports that 35% of the sentences tested have no dependency-link errors, while 89.8% have three errors or less.

Two papers give evaluations that are based on the measurement of run-time behavior. In “Parsing by successive approximation,” Helmut Schmid describes an efficient parsing technology that is nonetheless able to process grammars that make significant use of features. The efficiency of this algorithm is demonstrated by appeal to a range of empirical performance statistics. Udo Hahn, Norbert Bröker, and Peter Neuhaus
take a similar approach to evaluation. Their contribution describes “Message-passing protocols for object-oriented parsing,” and shows how to derive different heuristically guided parsing algorithms from variations in the communication patterns in an object-oriented parser. They report a variety of performance statistics for a set of 41 challenging-looking sentences from German computer magazines.

Since a version of the material of the book has already been presented at a workshop with proceedings (Bunt and Nijholt 1997), it is relevant to ask what has been gained (or lost) in the transition to (an expensive) book form. The articles average 20 pages—longer than the original conference presentation—and several authors have made good use of the opportunity to update and revise their work. The editors have selected an interesting group of papers, and provide a clear introduction with useful summaries of the chapters, pointing out some interesting relationships between the different lines of research.1 On the other hand, despite the high price of the book, there is no evidence that a competent professional copy editor was involved in the process of publication. This is a shame, since several of the contributions (especially Hektoen’s) deserve to be more widely known.

References


Chris Brew is an assistant professor of computational linguistics and language technology at the Ohio State University. His recent research has concerned the use of corpus-based methods in psycholinguistics and in natural language generation. Brew’s address is: Department of Linguistics, Oxley Hall, 1712 Neil Avenue, Columbus, OH 43210; e-mail: cbrew@ling.ohio-state.edu.

1 In some cases, Bunt and Nijholt seem to be going out of their way to convince themselves that essentially symbolic work is founded on a probabilistic approach. The papers by Chen and Vijay-Shanker, and by Hahn, Bröker and Neuhaus, in spite of the editorial claim that they fall under “the development of strategies for efficient probabilistic parsing”, do not go into detail on this issue.
Presumptive Meanings: The theory of generalized conversational implicature

Stephen C. Levinson
(Max Planck Institute for Psycholinguistics, Nijmegen)

Cambridge, MA: The MIT Press
(Language, speech, and communication series), 2000, xxxiii+480 pp;
paperbound, ISBN 0-262-62130-4, $35.00

Reviewed by
Nancy Green
University of North Carolina at Greensboro

Levinson’s book presents a theory of generalized conversational implicature (GCI), and makes the central claim that this theory necessitates a “new view of the architecture of the theory of meaning” (p. 9). Levinson claims that to account for GCI (and other types of presumptive meanings, or preferred interpretations), it is necessary to distinguish a new level of utterance-type meaning from sentence-meaning and speaker-meaning: “This level is to capture the suggestions that the use of an expression of a certain type generally or normally carries, by default” (p. 71). The book belongs to the genre of linguistic argumentation. Expanding upon the Gricean notion of GCI (Grice 1975), the author provides numerous examples of GCI and classifies them into three categories, each category representing a different licensing heuristic. Then he discusses the implications of the theory: first, for the interface between semantics and pragmatics, and second, for syntactic theory. Throughout the presentation, the author addresses in great detail potential objections and counterarguments from alternative theories of meaning.

According to the author, GCIs are defeasible inferences triggered by the speaker’s choice of utterance form and lexical items because of three heuristics mutually assumed by speaker and hearer. The heuristics, which can be related to Grice’s maxims, are these:

- The First (Q) Heuristic: “What isn’t said, isn’t.” For example, in the context of a blocks world where there are salient oppositions of objects {cones, pyramids, cubes} and colors {red, blue}, from the assertion “There’s a blue pyramid on the red cube”, this heuristic triggers the following inferences: ‘There is not a cone on the red cube’; ‘There is not a red pyramid on the red cube’ (p. 31).

- The Second (I) Heuristic: “What is simply described is stereotypically exemplified.” For example, from the assertion “The blue pyramid is on the red cube,” in the context described above one is licensed to infer ‘The pyramid is a stereotypical one . . . ,’ ‘The pyramid is directly supported by the cube . . . ,’ etc. (p. 32).

- The Third (M) Heuristic: “What’s said in an abnormal way, isn’t normal; or Marked message indicates marked situation.” For example, from the assertion “The blue cuboid block is supported by the red cube,” in the context described above one is licensed to infer ‘The blue block is not, strictly, a cube,’ ‘The blue block is not directly or centrally or stably supported by the red cube,’ etc. (p. 33).
In addition, the theory provides a refinement to Gazdar’s (1979) projection mechanism; GCIs licensed by Heuristic 1 are preferred to those licensed by Heuristic 3, which in turn are preferred to those licensed by Heuristic 2.

The book argues against the traditional view of the roles of semantics and pragmatics, according to which the output of semantics is the input to pragmatics. Instead, it argues for a more complex relationship in which GCI can play a role in truth conditions. In this model, two distinct types of semantic processes and two distinct types of pragmatic processes are involved. First, the semantic representation derived from the syntactic structure and lexical items of a sentence may be underspecified. The output of this semantic process is the input to a pragmatic process (“Gricean pragmatics I”), in which default, defeasible pragmatic inferences such as GCIs may contribute to determining the proposition expressed, for example, by helping to disambiguate lexical ambiguity, “generality narrowing” (i.e., narrowing word sense), and determining reference. The output of this process is the input to model-theoretic semantics. After sentence meaning has been determined by this semantic process, another pragmatic process (“Gricean pragmatics II”) is responsible for deriving other inferences such as particularized conversational implicatures; this final process yields speaker meaning.

Certainly, the debate on the role of pragmatics in the linguistic “architecture” is of significance to computational linguistics. In addition, the book provides a wealth of descriptive information on GCI as well as many pointers to related work in theoretical linguistics. However, unfortunately, the book lacks a computational or formal orientation. For example, much theoretical work would remain to build a computational model of GCI based just upon the information presented in the book. Also, its coverage of potentially relevant work in computational linguistics is not up to date; for example, there is no discussion of recent lexical-pragmatics-oriented approaches such as those of Di Eugenio and Webber (1996), Elhadad, McKeown, and Robin (1997), and Stone and Webber (1998). Despite these limitations, this book will be of interest to language researchers, computationally-oriented or not, with an interest in theories of meaning.

References

Nancy Green is an Assistant Professor of Computer Science at the University of North Carolina at Greensboro. Her dissertation research was on a computational model of conversational implicature for interpreting and generating indirect answers (Green and Carberry 1999). Green’s address is: Department of Mathematical Sciences, 383 Bryan Building, University of North Carolina at Greensboro, Greensboro, NC 27402-6170. e-mail: nlgreen@uncg.edu.
A Computational Theory of Writing Systems

Richard Sproat
(AT&T Laboratories)


Reviewed by
Kenneth R. Beesley
Xerox Research Centre Europe

1. Introduction

Interest in the history, theory, and classification of writing systems has never been higher, and the last decade saw the publication of several worthy books on the subject, including the formidable one by Daniels and Bright (1996). At the same time, under the Unicode initiative, there has been solid progress in the definition of and, finally, the implementation of standards for computer encoding and rendering of scripts used around the world. However, the implications of this multi-lingual revolution for computational linguistics beyond the level of word-processing have not been well explored, and Richard Sproat’s book A Computational Theory of Writing Systems is a welcome contribution.

In particular, Sproat’s observations and theories are motivated and tested by years of work at AT&T on text-to-speech systems. This is not the first or the last time that the rigor of computational application, and the massive practical testing that it allows, will come back to shape theory.

2. Derivational geometry

Sproat provides very little in the way of an introduction to writing systems; after pointing the reader to reliable sources, he jumps boldly into his model of reading devices. The geometry of his model is traditional derivational phonology, involving a mapping through various levels from an underlying representation $U$ to the surface phonology. Sproat makes two major claims:

1. Consistency: Every orthography corresponds to a single Orthographically Relevant Level (ORL) in the derivation; and

2. Regularity: $M_{\text{ORL}} \rightarrow T$, the mapping from the ORL to $T$, the spelling itself, is regular.

In the current academic world, derivational approaches to phonology are out of fashion, overshadowed by Optimality Theory (Prince and Smolensky 1993) and other mono-stratal models. The Two-Level finite-state model of computational morphology (Koskenniemi 1983) also eschews derivation. Nevertheless, a great deal of computational phonology and morphology continues to be done with derivational cascades of rewrite rules, which Johnson (1972) and Kaplan and Kay (1994) showed to be only
finite-state in power. Computational linguists from this tradition have challenged OT (Karttunen 1998; Gerdesmann and van Noord 2000), assuming that as long as linguists stay within finite-state power, they can construct their grammars flatly or derivationally as they find most convenient and perspicuous.

The interesting contribution of Sproat to this ongoing debate is his argument that the derivation has explanatory power, modeling “orthographical depth.” In this model, English orthography reflects a fairly deep ORL, Russian a shallower one, Belarusian a level slightly shallower than Russian, Spanish quite shallow, and so on.

Sproat’s claim that his derivations are regular is testable and potentially disprovable; he is forthright in discussing challenges and apparent counterexamples. The second claim, that an orthography always represents a consistent level of representation, and not sometimes one level and sometimes another, is a stronger claim, especially given the messy history of borrowing and adapting scripts. Scripts, as Sproat himself points out, are contingent “artifacts” or technologies, not something inherently human like phonology. He argues credibly, however, that a viable everyday writing system must bear a “sensible relationship” to the language it represents, that we can expect a natural pressure in the direction of consistency.

3. Derivational breakdown

Sproat’s $M_{ORL\rightarrow I}$ rules are divided into two main subgroups: $M_{Encode}$ which are (morpho)phonological mappings, and $M_{Spell}$ which are “autonomous spelling rules” or rules reflecting the conventions of the orthography itself. As the rules are regular, they can simply be composed together, and the result can be encoded as a finite-state transducer and applied bidirectionally.

$$M_{ORL\rightarrow I} = M_{Encode} \circ M_{Spell}$$

This distinction is easy to defend. $M_{Spell}$ covers phenomena such as the conventions for representing phonologically long vowels in Dutch orthography, and parallel examples in other orthographies are easy to find.

Sproat then makes another, less-obvious, distinction, splitting up $M_{Spell}$, so that $M_{Spell_{map}}$ is a mapping, encoded as a finite-state transducer, but $M_{Spell_{constrain}}$ is encoded as a regular-language filter. Again, the two subsystems are regular and are composed together.

$$M_{Spell} = M_{Spell_{map}} \circ Id(M_{Spell_{constrain}})$$

As composition is defined only for transducers, the composition must technically involve the identity relation on the filter as shown.

The examples of $M_{Spell_{constrain}}$ involve alternate representations that appear in complementary distribution in the surface orthography. In Malagasy, the vowel /i/ is represented as either ⟨i⟩ or ⟨y⟩, with ⟨i⟩ occurring only in nonfinal position and ⟨y⟩ occurring only at the end of words. If $M_{Encode}$ contains the rule

$$i \rightarrow ⟨i⟩ \ | ⟨y⟩$$

that is, vowel /i/ is realized as either orthographical ⟨i⟩ or as ⟨y⟩, then $M_{Spell_{constrain}}$ would include the following regular filter to constrain the variants to appear only in appropriate contexts.

$$-[(\Sigma^* ⟨i⟩ \ #) \ | \ (\Sigma^* ⟨y⟩ -#)]$$
Grammars that overgenerate and then filter in this way have an obvious OT flavor. However, while this approach would definitely seem to work, it is difficult to see how it differs substantially from the following two mapping rules.

\[ i \rightarrow \langle y \rangle \# \circ i \rightarrow \langle i \rangle \]

Here the phoneme /i/ is first mapped to /y/ at the end of words, and elsewhere, any leftover /i/ is simply mapped to /i/. If orthographical /y/ and /i/ are excluded from the domain, then the following equivalence holds.

\[ \overline{[\Sigma^*(\langle y \rangle \mid \langle i \rangle) \Sigma^*]} \circ i \rightarrow \langle i \rangle \mid \langle y \rangle \circ \overline{[\Sigma^*(\langle i \rangle) \mid \langle y \rangle \rightarrow \#]} \]

\[ \equiv \overline{[\Sigma^*(\langle y \rangle \mid \langle i \rangle) \Sigma^*]} \circ i \rightarrow \langle y \rangle \circ \overline{\#} \circ i \rightarrow \langle i \rangle \]

The filter component \( M_{\text{Spell}_{\text{min}} \text{im}} \) is also invoked for handling the alternation of Greek nonfinal \( \varsigma \) vs. final \( \varsigma \), and for the contextual variant shapes of Arabic (what Unicode calls the character vs. glyph distinction). Again, it is difficult to see why these same phenomena could not be handled with mappings or transducers rather than filters, according to the taste and convenience of the linguist.

4. Planar regular language

Sproat expands the normal notion of regular language, consisting of strings of linearly concatenated symbols, to planar regular languages, which allow a richer set of concatenation operations, including left concatenation (\( \bullet \)), right concatenation (\( \bullet \)), downwards concatenation (\( \# \)), upwards concatenation (\( \bar{\#} \)), and surrounding concatenation (\( \bigcirc \)). Illustrated on Chinese, this notation allows for the grouping of semantic radicals and (semi-reliable) phonetic elements into traditional Chinese morphograms. The same mechanism is applied, even more successfully, to Korean Hankul (Hangul), where letter units are arranged into syllable-sized glyphs, and to Devanagari and Pahawh Hmong. Planar regular expressions are therefore the mechanism proposed for notating the relative placement of glyphs in all “Small Linguistic Units,” where variation from the macroscopic order of the script is possible.

5. Conclusion

Any theory of writing systems takes on a huge task, ultimately beyond the ability of any single human being. Sproat’s references are intimidatingly wide, and his examples and experiments include Chinese, Korean, English, Russian, Belarusian, Croatian, Mayan, Manx Gaelic, Egyptian Hieroglyphics, Syriac, Malagasy, and others. He applies his theory to classification, adaption of writing systems, spelling reform, and the psycholinguistics of reading. Yet he’s repeatedly candid about the challenges to and limitations of his theory (e.g., calligraphy is beyond the pale), and he is to be commended for pointing out the areas needing further research and confirmation.

Individual readers will of course want to see the theory tested on their own favorite writing systems; I, for one, wanted to see how the model might apply to Arabic orthography, and how the model might shed light on the debates about rival proposed orthographies in the field, for example, for Bantu languages. Critics will want to test the claims of Regularity on reduplicating languages, and the claims of Consistency against the amount of ad hoc “lexical marking” needed to support it. But Sproat has laid out a testable theory, and in the best scientific spirit he has even provided listings of lexicons and derivation rules for English, contrasting a deep ORL solution à la
Chomsky and Halle (1968) with a shallower solution. To complete the picture, the AT&T finite-state libraries and a set of programming formalisms called Lextools are now available on the Web, which further facilitates the reproduction of his English examples and testing on other languages.\footnote{http://www.research.att.com/sw/tools/fsm/, http://www.research.att.com/sw/tools/lextools/}

All in all, this book is commendable, and it is highly recommended for any serious student of writing systems.

References


Kenneth R. Beesley is a Principal Scientist at the Xerox Research Centre Europe. He works mostly in finite-state morphological analysis and generation. Beesley’s address is Xerox Research Centre Europe, 6, chemin de Maupertuis, 38240 Meylan, France; e-mail: Ken.Beesley@xrce.xerox.com.
The study of prosody is perhaps the area of speech research which has undergone the most noticeable development during the past ten to fifteen years. As an indication of this, one can note, for example, that at the latest International Conference on Spoken Language Processing in Philadelphia (October 1996), there were more sessions devoted to prosody than to any other area. This focus on prosodic research is partly due to the fact that developments in speech technology have made it possible to examine the acoustic parameters associated with prosodic phenomena (in particular fundamental frequency and duration) to an extent which has not been possible in other domains of speech research.