Errors and Intelligence in Computer-Assisted Language Learning. Parsers and Pedagogues. Trude Heift and Mathias Schulze. 2007. Routledge: New York. (Routledge Studies in Computer-Assisted Language Learning, 2, series editor Carol Chapelle) 227pp., 10pp. notes, 25pp. bibliography, 10pp. index.

There are three excellent reasons for wanting this book in your local library. First, it contains an impressive survey of work done on computer-assisted language learning (CALL) over the past twenty-five years. As the book was written, the last major overview on CALL had been Levy's (1997) monograph, so the time seemed ripe for a follow-up (as it turns out, Levy and Stockwell (2006), too late for Heift and Schulze's bibliography, also updates Levy (1997)). The twenty-five page bibliography alone is a substantial contribution, and the review of "more than one-hundred" projects involving natural language processing (NLP, see below) and CALL is simply invaluable (pp.52-82). I was surprised at the omission of more recent literature on second-language learning (Doughty and Long, 2003; de Bot et al. 2005), but these are slightly off the book's focus..

Second, Heift and Schulze's book has added value for its focus on work attempting to exploit natural language processing (NLP), the discipline which aims to understand how natural languages such as English or French can be processed using the computer. "Processing" is understood broadly in NLP to include e.g. problems of indexing, searching, (alphabetic) sorting, recovering linguistic structure, generating (producing) language, or translating. So the particular perspective seems quite natural: let's use the best knowledge on how to process language to support programs intended for assisting language learners. There are many potential uses for NLP in CALL (Nerbonne 2002), but *Errors and Intelligence in CALL* (E&I-CALL) focuses specifically on uses of NLP to spot errors in learner input, to classify those errors, and, by intelligently drawing learners' attention to selected problems, to contribute to learners' improvement. The authors are more than conscious of the need to be selective in reporting errors back to language learners.

It is difficult to detect errors automatically as many readers will know. Microsoft *Word* offers "grammar checking," but many of its suggestions are irrelevant or simply wrong. This is not a great problem for competent native-language authors, for whom the grammar checkers serve a purpose, but language learners would quickly become confused. And *Word* is close to the state-of-the-art in error detection, suggesting that the problem is non-trivial. E&I-CALL illustrates one difficulty nicely in the following brief example (p.207), which appears to have two nominative pronouns contending for the subject role:

Sie	liebt	er
She-nom.	love-3s	he-nom

correction:

Sie	liebt	ihn
She-nom.	love-3s	he-acc
'She loves him'		

The difficulty with the simple correction is that the original is not absolutely ill-formed; *sie liebt er* is well-formed e.g. if the *sie* (which, incidentally, can also be accusative case, and therefore an object) is contrastively stressed. The correction is nonetheless appropriate for beginning or intermediate learners who need to master pronominal case – not only to keep the exercise going, but also since this construal only makes sense in very particular circumstances. The example illustrates the more general point that wide-coverage grammars (or parsers) will need to be tuned to the target grammars that learners are aiming for. This implies that special-purpose grammars will need to be developed for error detection.

A further, general difficulty in detecting errors must be seen against the background of the ambiguity problem in computational linguistics. Natural language parsers normally assign too many analysis trees to well-formed sentences, including those that native speakers perceive as unambiguous. As Abney (1996) argues, the problem lies not in the quality of the grammars but rather in the fact that natural language structures may be used for multiple purposes. The potential ambiguity does not impede communication because interlocutors focus effortlessly on each other's intended meanings – i.e. they DISAMBIGUATE. If we now attempt to additionally recognize not only the legitimate (target) structures of the language being learned but also the ill-formed structures produced by learners, we recognize additional structures and exacerbate the disambiguation problem. To make matters more difficult, one also needs to classify the errors detected if one is to offer effective feedback to users. In his respect, parsing for CALL applications is more difficult than parsing in applications in which well-formed text dominates (newspaper text). Chap. 2 reviews various approaches to solving this problem, but does not, to my mind, add a great deal of clarity to the discussion. While Menzel and colleagues are sanguine about error detection (see references in E&I-CALL), Vandeventer Faltin (2003) is explicitly pessimistic about the same techniques after evaluating them practically and extensively on a corpus of learner language, and obtaining error recognition rates of only about 50% (p.180).

When two competent practitioners disagree about an issue as simple as whether one can or cannot detect errors, then the question may need to be posed more subtly. Perhaps it is useful to study the circumstances under which errors may or may not be found, e.g. as a function of sentence length or target grammar complexity, or of error sorts, or learning situation (direct response vs. free composition).   Such suggestions are in the literature, and are referred to throughout the book, but they are not developed systematically. This reflects a fundamental weakness in E&I-CALL, namely that it is largely structured as a report on literature rather than as a systematic development. This also means that the authors do not attempt to extend a systematic description to an agenda for future research, something the field could definitely use.

A third reason for valuing this book indeed provides some of the systematization just found lacking: E&I-CALL is valuable for the integrated presentation it offers of the substantial research line that Heift has developed for investigating learner errors in CALL. Heift has evaluated components for their utility in CALL applications (Ch.2), and also studied the degree to which students are willing and able to use feedback from a CALL system (149ff), incorporating a modest number of learner variables. She has also shown how existing grammatical description frameworks allow parsing of ill-formed input together with an indication of the source of errors (Ch. 4.6), and

she has evaluated that system vis-à-vis genuine student errors, including multiple errors (145ff, in a misnumbered section). This section illustrates the role of error detection and classification in supporting student exercises, a nice way to introduce the novel techniques in a limited fashion. Finally, she has researched the opportunities for employing user models to track student responses and thereby optimize system responses. E&I-CALL is valuable for pulling these various studies together.

But there are many points for which another book or a follow-up to this one is needed. In the ten years since Levy (1997), CALL (albeit NLP-less CALL) has taken over the role the older analogue language labs once played, meaning that the technical infrastructure for the sorts of applications E&I-CALL promotes is no longer a problem. This might suggest a strategy of developing increasingly ambitious applications for error detection and classification which might be integrated straightforwardly into existing systems, and Heift and Schulze would be the right researchers to sketch this research agenda in more detail (see remarks about Heift's deployment of error detection within simple exercises). In that case we should also prefer to see more attention paid to how these sorts of application should best be evaluated, perhaps using learner corpora of the sorts discussed in the book.

It would also be rewarding to pay more attention to what students *like* to do, and not only to what they learn when they are active. Given the general emphasis in education in guaranteeing "time on task," it is bothersome that CALL is sometimes accused of making language learning less social and less enjoyable, since this will depress the time students will spend learning, making CALL-based instruction less successful. Perhaps game-based or more socially oriented learning forms would be a more enjoyable embedding or accompaniment to the "focus on form" that is central in E&I-CALL.

I would also have appreciated some consideration of the techniques that have been developed for processing ill-formed input in spoken natural language understanding (NLU). Granted that the primary problem in NLU is not error detection and classification but rather the recovery of the intended message, NLU nonetheless remains an area in which the processing of ill-formed input is central, and from which techniques might be available. Finally, I would suggest that follow-up texts might focus more sharply on central issues in CALL and avoid digressions on general issues such as the choice of programming language for implementation or the appropriate background theory for understanding human-computer interaction (4.1-4.3).

The book has been produced somewhat carelessly. In Ch. 1, Sec. 2 announces that the book has four main parts (p.5), and then lists five, beginning with Part 2. Sec. 1.1.1 follows Sec. 1.2, and is followed by two sections numbered 1.1.3. "Buggy" rules, those used to process learner errors, are introduced on p.39, but are referred to later as "mal-rules" (p.196, p.202) without explanation. Neither term appears in the index. In general the prose is fine, but a copy editor would have caught many minor errors, such as the agreement error in the second sentence of the book (p.1). Naturally, authors are responsible for their books, but competent editing would have prevented many of these mishaps.

References

Abney, S. (1996) Statistical methods and linguistics. In: J.Klavans and P. Resnick (eds.) *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge: MIT Press.

de Bot, C. L. J., Lowie, W. and Verspoor, M. (2005) *Second Language Acquisition: An Advanced Resource Book*. London: Routledge.

Doughty, C. and Long, M. (eds) (2003) *Handbook of Second Language Acquisition*. Oxford: Blackwell.

Levy, M. (1997) *Computer-Assisted Language Learning. Concept and Conceptualisation.* Oxford: Clarendon.

Levy, M. and Stockwell, G. (2006) *CALL Dimensions: Options and Issues in Computer-Assisted Language Learning*. Mahwah, NJ: Lawrence Erlbaum.

Nerbonne, J. (2002) Computer-Assisted Language Learning and Natural Language Processing. In: R. Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*. Oxford: OUP. 670-698.

Vandeventer Faltin, A. (2003) *Syntactic Error Diagnosis in the Context of Computer-Assisted Language Learning*. Diss., University of Geneva.

John Nerbonne, University of Groningen, The Netherlands
Email: j.nerbonne@rug.nl