P.S.F van Keulen and W.Th. van Peursen, eds., *Corpus Linguistics and Textual History: A Computer-Assisted Interdisciplinary Approach to the Peshitta*, Koninklijke Van Gorcum BV, 2006

DERYLE LONSDALE, BRIGHAM YOUNG UNIVERSITY, PROVO, USA

The invitation to review this book was accepted with some trepidation; half of the topics mentioned in the title were quite familiar, but the others less so. Fortunately, the editors' goal was to build a conceptual and methodological bridge between two disciplines, and reading the book was very much an experience in leaving behind comfortable territory and crossing over into unfamiliar realms. Ultimately the traversal was successful, enlightening, and perspective-enhancing, and was punctuated with only occasional desires to climb over the railing and jump or otherwise escape.

This review is written from the perspective of one who is a computational linguist with extensive experience in corpus development as well as translation theory and practice, having a good knowledge of biblical Hebrew and a growing acquaintance with Syriac, but with very little knowledge of (albeit increasing interest in) textual criticism and exegesis. Anyone who shares any subset (or superset) of these interests will find the book compelling, though those with narrow specialties will find themselves correspondingly stretched.

The content includes, and extends, material presented at a seminar held in the Netherlands in 2003 that focused on the Computer-Assisted Linguistic Analysis of the Peshitta (CALAP) project. Evidently the effort combines two teams, one specializing in the development of computational tools for Hebrew, then Syriac, linguistic analysis (WIVU), and the other (PIL) with a history of analyzing the Peshitta from text historical, critical, exegetical, and translation-theoretic perspectives. The attempt to unify these two traditionally separate undertakings under one umbrella effort seemed initially to this reader an intriguing but Herculean (pardon the pagan reference!) task. The body of the text is intended to convince the skeptical, and for this reader it did.

The first chapter is a wonderful 30-page survey of the motivating factors for the project: to create a truly interdisciplinary approach—complete with the requisite tools—to linguistic and textual analysis, and to illustrate its usefulness with a nontrivial

application (no less than the Peshitta). A broad discussion of such topics as the document itself, linguistic factors in Old Testament exegesis, synchronic vs. diachronic analyses, translation theory, the cultural context of language(s), language use, and stylistics lays the linguistic groundwork for this effort. A brief overview of relevant manuscripts and other texts sets the focus on the target of the approach. This enjoyable chapter could serve very well as a standalone tutorial on this constellation of topics.

The next 40-page chapter includes a technical description of the computational approach that WIVU used in annotating the Biblical Hebrew data, and how their methods were adapted for processing the linguistic content of Syriac text. The discussion is replete with data listings that may be intimidating to some, but the narrative is expertly crafted to help initiate the non-computational to the myriad of levels of analysis, descriptive labels and features, and processing stages that the text is subjected to. The formatting and layout of the data examples is impressive and very readable, and the technology described is noteworthy, if not well documented so far in the usual computational linguistics publication venues elsewhere.

The balance of the first third of the book consists of several chapters laying out the technical issues involving the tools development effort, linguistic analysis conventions, and annotation schemes. These will be of interest to anyone undertaking similar linguistic annotation projects, or specialists who will someday use such tools. After accessible discussions by project members on these topics, responses by others raise issues about coverage, evaluation, ambiguity, overall project goals, assumptions about linguistic theory, and the tensions about empirical versus rational analyses. The discussion is informative and interesting.

The next third of the book involves two back-and-forth dialogues and highlights the role that CALAP's offerings can play in these discussions. The first centers on a syntactic issue, that of nominal clauses and the role of the enclitic personal pronoun. An introductory chapter summarizes three prevailing approaches, and then the proponent of each responds in subsequent chapters. The discussion is interesting for its linguistic implications, but too involved to mention further here. The topics involve such current issues as predication, clitics, and definiteness. Linguists looking for language data to sound out various theoretical approaches to

morphosyntax will find a treasure trove in this exchange. The second discussion illustrates questions germane to the other side of the "bridge", that of analyzing textual variants. The issue at hand is where the Targum and the Peshitta agree and diverge, with respect to each other and to the MT. The investigation was, again, very carefully written and was perfectly tractable to this reader, who was now in largely unfamiliar territory. One of the respondents points out the problem—independently apparent to this reader as well—that no mention was made in this latter study of how the CALAP material was used, though clearly it was to some degree.

The last third of the book answers the question so often posed to corpus developers by potential end-users: "Now that I have the corpus, what can I do with it?" To this end a nine-verse passage in 1 Kings 2 is strategically chosen to illustrate the possibilities that an extensively annotated corpus provides to researchers. A wide array of perspectives is applied in viewing the contents of the text in these verses from formalist and functionalist angles, and the result is an impressive illustration of CALAP's capabilities (as well as a few of its shortcomings). An epilogue serves to reiterate how well the interdisciplinary approach bridges the interests of a wide range of researchers.

One has the impression that in some aspects of CALAP's technology, the 2003 snapshot we are given can be updated (and perhaps it has): there is no mention of current topics such as best practices in corpus annotation, morphological parsing tools could be analyzed in a more versatile way using finite-state technologies, statistical analyses could be a little more developed, and machine learning is more viable in corpus annotation work today. Still, the theoretical and methodological work is sound, even solid, and the demonstrations of its effectiveness are impressive.

Finally, some low-level remarks are perhaps in order. The text is replete with examples, quotes, data, and footnotes in several languages, and therefore assumes some familiarity on the part of the reader with French, German, Italian, Greek, Latin, Slavic, and (naturally) Semitic languages. A superb work was performed in editing such a complicated text; only a dozen or so errors, mostly in English spelling, were detected. Reflecting the bipartite nature of this text, some chapters had their extensive citations, footnotes, and textual apparata notated at the bottom of each page, which to

this reader was a little unwieldy; the others had this material at the end of each chapter.

Overall, this book is a remarkable work and will stand as one of the memorable examples of how to design, implement, successfully realize, and document a large-scale, multi-layered linguistic development project. It also serves as a model of how to build an interdisciplinary bridge across theoretical and methodological gaps that need to be addressed if we are to better appreciate language and its use.