1. Introduction

This entry presents a brief overview of the development of computational information systems for Biblical Hebrew. Highlighting major trends and research topics that have received attention in the impressive body of literature in the field since Hughes (1987) and through Poswick (2004), Tov (2003; 2006), and beyond, this article also points out deficiencies in existing studies and discusses opportunities for future research.

2. Levels of Analysis

Biblical Hebrew (BH) grammar has been studied from many different angles, and much of this body of knowledge has been preserved in various computer software systems and databases. The most basic systems entail a digital reproduction of the text in Hebrew characters alongside an exact transliteration in the Roman alphabet. The transliteration may be used to reconstruct the text in the Hebrew alphabet, while the inclusion of a separate phonological transcription serves to convey the pronunciation of the Hebrew text, although the letters or signs used do not necessarily correspond exactly to the Hebrew spelling. The next levels of analysis are the morphological, morpho-syntactic, and syntactic levels. Less attention has been paid to the development of more advanced linguistic levels dealing with the semantic and pragmatic ‘modules’ (Van der Merwe 2002:89) of the language, and one can only hope that knowledge databases and expert systems dealing with these levels will become more readily available (See, e.g. Bothma 1992a:163–172).

There are currently 14 software applications containing the Codex Leningradensis version of the Hebrew Bible and one application containing the Aleppo Codex (Tov 2006:343). Eleven of these tools offer morphological analyses (Tov 2006:356). Only a select few contain syntactic data, for example the database of the Werkgroep Informatica of the Free University in Amsterdam (WIVU). In addition to this grammatically oriented software, ancient and modern translations are available, as well as critical apparatuses and other tools, such as dictionaries and even one reference grammar (Tov 2006). The most advanced programs allow searches for morphological characteristics and for combinations of lexical and grammatical details (Tov 2006:346). Five main groups of morphological analysis exist:

- Westminster Hebrew Old Testament Morphology (Groves-Wheeler)
- Werkgroep Informatica (Talstra)
- Bar-Ilan analysis
- Academy of the Hebrew Language
- Additional commercial and private morphological analyses

The latest version of the Hebrew Old Testament linguistic database, developed over the past three decades by Werkgroep Informatica at the Free University in Amsterdam (WIVU), has been included in the Stuttgart Electronic Study Bible (SESB), published on the Libronix/Logos platform (SESB 2002–2010). This tool allows researchers to perform advanced syntactic queries such as searching for examples of clauses having a conjunction and proper name as subject preceding an imperfect verb (Talstra 2007:93). According to Talstra (ibid.:96), “[t]he search for syntactic data offers one way to get a better handle on the function of ordinary and extraordinary constructions in a literary composition” for the sake of discovering grammatical tensions built into a literary text. The search engine operates mainly on formalistic characteristics and the researcher must ‘translate’ his or her query into these terms (cf. Talstra 2007:91, 93, 95). That being said, the SESB does provide a user interface that allows users to use buttons and checkboxes to select various “combinations of syntactic or grammatical features and functional categories”, either on word, phrase, or clause level (Kummerow 2005:2–3). Although Gómez (2004) regards the search function as one of the ‘jewels of the crown’ of the SESB, and while plain morphological searches are simple to execute, he cites what he calls a ‘steep learning curve’ that must be overcome for the implementation of more complex queries. The search engine was developed in cooperation with Prof. C. Hardmeier of the University of Greifswald (2001). The WIVU team is currently working on a more advanced version of the database, which captures pragmatic information. The ability of the tool to produce phrase parsing, clause level parsing, and clause hierarchies distinguishes it from other Biblical Hebrew
information systems. There is, however, no clear distinction between the various linguistic modules in the tagging system. For an overview of other commercially available products such as Bibleworks and Accordance, see Harris (2008).

Another project that offers analysis on an advanced syntactic level is the private database developed by Andersen and Forbes. Their project analyzes the elements of each clause up to the most atomic elements (morphemes). Syntactic information and semantic-role information are presented as horizontal trees/graphs. Andersen and Forbes (2003) trust that their proposal will make a contribution to the field of Biblical Hebrew linguistic information systems by moving beyond the limits of single clauses. Their work is without a doubt extremely useful for the study of syntax, but since they reject the autonomy of syntax, it neither differentiates between the various linguistic modules nor is it intended to facilitate multi-modular linguistic studies. Moreover, one needs extensive knowledge of the symbols used in order to make sense of the myriad of labels used to tag the nodes and leaves of their representations. Their tagging of semantics at word level rests on ‘naïve semantics’, though their repertoire of semantic roles is extensive, involving forty-four categories. In Andersen and Forbes (2010), the authors describe and discuss the theory and practice of their database in detail.

Richter’s database, which is not yet available publicly, does assign semantic values to syntactical functions, though it, too, does not create a clear distinction between syntactic and semantic modules; for example, its ‘syntagmemes’ include categories from both modules, such as subject, predicate, and direct object (syntactic functions), as well as locative, advantage, and agent (semantic roles) (Rechenmacher and Van der Merwe 2005:71–72).

Holmstedt’s (University of Toronto) and Abegg’s (Trinity Western University) Syntactic Database of Ancient Hebrew is aimed at producing a syntactically tagged database for all ancient Hebrew texts written in the first millennium B.C.E. These texts include all Hebrew inscriptions found in recent collocations, all biblical texts, all non-biblical Qumran texts according to the published editions, and the text of Ben Sira as re-transcribed from the photographs and facsimiles and compared to all previous editions. Excluding semantic, pragmatic, and discourse information, the goal of the database is to capture the relationship among the various constituents in the domain of the clause. The linguistic principles upon which the tagging scheme is built are: (1) hierarchical phrase structure; and (2) non-binary branching. In other words, a given clause is divided into the core constituents of subject and predicate, while all modifiers are nested within the domains of the head of the subject (e.g., noun) or the head of the predicate (e.g., verb). Although every clause is binary by virtue of being attributed a subject and predicate (whether these components be syntactically overt or absent), the various
remaining components (e.g., complement, adjuncts, determiners) need not be binary (e.g., a verb may have just one or even three or more structurally ‘equal’ adjuncts).

The endeavors of the J. Alan Groves Center for Advanced Biblical Research currently still focus on text, morphology, and syntax, and is responsible for three major digital Biblical Hebrew projects: the Westminster Leningrad Codex (WLC) consists of the biblical text (consonants, vowels, accents, ketiv-qere, but not the masoretic notes); the Westminster Hebrew Morphology (WHM) seeks to provide a lemma—when possible, according to HALAT/HALOT (Richardson 1994–2000), part of speech, and other parsing for each of the approximately 500,000 morphemes of the Hebrew Bible, including the Aramaic sections; and the Westminster Hebrew Syntax (WHS), the most recent project of the Groves Center, seeks to tag the text for clause boundary, clause constituent, and hierarchical relations between phrases, as well as main and dependent clauses. This project also includes some additional labeling for clause types (verbal, nominal, etc.). The database is intended to be a ‘modern’ syntax using accepted methods in computational linguistics and natural-language processing. The data are generated by a set of rules (grammar) that is submitted to a computer program (parser), which creates the ‘treebank’. The goal is a formal, ‘theory neutral’ syntax that allows the user maximum freedom in conceptualizing and using the data. Syntax labels are primarily taken from the part of speech of the head of phrase, while the clause-level labels are based upon the formal characteristics of the clause.

3. Underutilization of Existing Tools

Poswick (2004) gives an overview of Biblical Hebrew information system projects conducted between 1985 and 2004, coming to the conclusion that, although various tools provide morphological analyses and even other levels of analysis, “classical Biblical exegesis would not appear to be benefiting as yet from the results of this type of analysis.” Tov (2006:337) agrees with Poswick that biblical scholars still do not make optimal use of these tools. This may be because scholars have been so focused on creating and improving the tools themselves that they have not yet maximized in-depth exploration of the huge amounts of data that have been made available by these tools (ibid.:338).

Nonetheless, one must concede that many exegetical articles have been produced as a result of Werkgroep Informatica’s databases. A recent example of the use of their syntactic hierarchies database can be found in Talstra (2006:231–232), an investigation of the use of yiqtol verbs in the narrative prose found in Exodus, shedding new light on the exegesis of sentences where these verbal forms follow wayyiqtol forms. In her review of the SESB, which houses this database, Conybeare (2005) states that “[t]he student who was most excited by the possibilities of the SESB was
the one most closely engaged already with biblical exegesis.” In order to exploit this tool fully, a user would probably have to make a careful study of the essay in the manual that explains “how the Hebrew text was analyzed to facilitate more complex syntactical searches” (ibid.).

Winther-Nielsen (forthcoming) also uses Werkgroep Informatica’s database as a basis for developing the theory of Persuasive Learning Objects and Technologies (PLOT) for learning and teaching introductory Biblical Hebrew. He focuses on accessible and low-cost tools, and develops learning objects, experiences, and quizzes in Moodle and in screen-capture videos. The database is also used in a second project, the Role-Lexical Module (RLM).

A feeling of information overload may be another reason for the underutilization of Biblical Hebrew information systems. Claassen and Bothma (1988:83) highlight the problem of information overload that already existed in electronic Biblical Hebrew research twenty years ago. According to Bothma (1992b) hypermedia may be a solution for minimizing problems of information overload, as the network of hyperlinks allows the user to access only relevant information.

A third reason for the underutilization of these systems may be the lack of user-friendliness. According to Tov (2006:338), Biblical Hebrew research software is underused because “[t]here remains a wide gap between the knowledge of the experts creating the tools and that of the scholars for whom the tools are intended.” Indeed, many of these tools are not easy to use and are in desperate need of user-friendly interfaces.

4. Integration as a Solution to Enhance Utilization

Despite the remarkable utility of electronic aids for the study of the Hebrew Bible, they can also be overwhelming and even frustrating, due to the fact that several tools must often be used concurrently in order to study different linguistic levels and perspectives. Bothma (1990) proposes the use of integrated Biblical Hebrew information systems in order to enhance the process of computer-based education and to solve the problem of the often mutually isolated study of biblical languages. These systems should integrate introductory grammars, reference grammars, sources on the cultural background of the Bible, and research databases. Various levels of granularity of data should be available for users with different levels of knowledge and requirements. Poswick (2004) also advocates the use of hypermedia to take biblical research to a new level, “from the accumulation of electronic texts to the construction of hypertextual links between them with all the cultural data which permit their interpretation.” Hypermedia also provides a way to link electronic information systems with printed material, provided that these texts are made available in electronic Unicode format. There is some irony in this, given that many of the recent tools published in print are based on
electronic files and programs, one example being the *Concordance de la Traduction Oecuménique de la Bible*.

Systems have been suggested, and at least one has already been developed (the Lexham Hebrew-English Interlinear Bible; see Van der Merwe 2005), to display multi-level analyses of Hebrew clauses, integrating the various dimensions of clausal analysis into an interlinear table format on one screen. These tables resemble those found in relational databases, which gives birth to the wish of being able to conduct ad hoc queries on the stored data. However, these tables cannot simply be transformed into relational database tables, as each record (or clause) has a separate table, and the rows do not represent unique records. A closer inspection reveals that the rows actually represent various dimensions or levels of data-analysis that are strongly linked to the elements in the upper row. This type of interlinear table is in fact a two-dimensional representation of three- (or multi-) dimensional linguistic data structures. Bothma (1992a) proposes and successfully tests the use of SGML, of which XML is a derivative subset, to provide a platform-independent databank of linguistic and other related data.

Paratext from UBS (United Bible Society) is a useful information system that offers integrated functionalities for Bible translators. It is not specifically focused on Hebrew, but on all biblical languages. Since it offers a wide range of translation related functions, it can be used very effectively by all kinds of Biblical scholars. This program is made available to all Bible translators free of charge, but it is also available to other interested people. It offers unique features, some of which are highlighted here. The system includes the Hebrew and Greek texts, interlinear word-for-word parsing, the basic un-inflexed word and English glosses. When one clicks on the base form of the word, a choice of various dictionaries is offered. It also offers adaptable search facilities, even in the original languages, as well as a comprehensive 'back translation' tool which forms an interlinear text consisting of the new translation and existing translations. This function also suggests possible translation glosses. Spelling and consistency checks can be done. A very handy check, for instance, is the 'Sound alike words’. An exegete can also place notes into the text in the form of little flags for the translators to be seen. When changes are made the text can be compared and changes can be accepted all at once or one by one. UBS also offers a version control facility on their website. Program updates, as well as other downloads like new text resources, can be downloaded from the UBS website. Bothma (1992b) expresses the need for syntactic and semantic databases of Biblical Hebrew. Such databases may enhance grammatical research because “manual searching for complex syntactic examples is extremely difficult and inadequate in that retrieved information is very often incomplete due to the size of the corpora of texts” (Bothma 1992b:340). Although various syntactic databases have become available, the authors of this entry are unaware of any existing databases containing a separate module of
semantic functions that may help users understand the logical relations between the constituent elements of clauses and sentences, but it should be noted that Winther-Nielsen (2009) is currently working on a project that adds semantic roles to the WIVU database.

An XML data structure may provide an “appropriate information model for presenting Biblical information in an electronic form”, with reference to integrating and storing information from various linguistic modules (cf. Bothma 1992b:345). The advantages of XML, however, are not limited to the creation of a database structure. According to Van der Merwe (1995:419) the purpose of an electronic reference grammar “plays a major role in determining its structure and content”. The extensibility and adaptability of advanced mark-up languages such as XML make them ideal for implementing a custom-made macro-structure, which should, for example, fulfil the following requirements: “An electronic BH [reference grammar – JHK] should serve as a cheap up-to-date, as well as updateable, source of easily retrievable information on BH for readers of the BH text of the OT. These readers may have various degrees of receptive competency of BH” (Van der Merwe 1995:420).

The use of XML as a mark-up language to tag the data in a bank of biblical data may also enable learners to move between teaching and reference textbooks and to emulate deductive grammars (Bothma 1992a). Furthermore, it could also facilitate a shift in focus in biblical research from textual aspects to communicative aspects (Poswick 2004).

The combination of hypermedia, such as XML, and database concepts, forms a strong and promising alliance of techniques, which facilitates solutions to cater to a diversity of domains, users and applications, including integrated Biblical Hebrew information systems (Claassen and Bothma 1988:84). The use of an extensible, multidimensional data structure could facilitate the accommodation of other types of linguistic and non-linguistic data and may, therefore, be a step in the right direction toward solving the problem of new requirements that may be laid down by the “shift of paradigm from exegesis based on a philological approach, to hermeneutic[s] based on a linguistic and socio-linguistic approach” (Poswick 2004). Marked-up texts could be used to allow users access to these multidimensional databases, since hypertext not only holds the promise of creating multi-level translations (Van der Merwe 2004:10), but also of tagging texts in terms of divergent theoretical models (Van der Merwe 2006:276).

5. Visualization and Flexibility
Adding visualization techniques to the mixture of XML and databases could provide even more exciting possibilities. Claassen & Bothma (1988:88–89) suggest the use of visualization to direct users in finding their way through the convoluted sets of paths in hyperspace. Advanced processing
and visualization techniques may also contribute toward the development of user-friendly interfaces (Bothma 1992b:348). Kroeeze (2008) aims to contribute to the attainment of this goal by proposing a macro-structure for the integration and packaging of Biblical Hebrew linguistic information and by experimenting with some visualization techniques to render captured data in innovative ways.

According to Andersen and Forbes (2003:44), one of the requirements of a proper rendering of syntactic structures of Biblical Hebrew is for it to be pictorial, meaning, “clearly and concisely diagrammed”. They use graphs and trees to visualize (“represent”) the hierarchical syntactic structures of Biblical Hebrew clauses.

Scalability is a serious issue that needs to be addressed if one wishes to represent aggregate, linguistic information on a lateral level across the single units of the textual corpus. According to Andersen and Forbes (2003:45) the text of the Hebrew Bible consists of approximately 59,000 main clauses and 13,000 embedded clauses. Visualizing lateral information of larger sections, books, or the entire Hebrew Bible will surely pose new and difficult challenges for researchers.

Tov (2006:337) differentiates between non-flexible and flexible Biblical Hebrew software. Non-flexible tools, now often obsolete, reflect only the result of computer-assisted software in textual format, not giving the reader access to the original data or tool itself. Flexible tools, however, allow the interactive use of both tool and data. They may be used as “an extension of our own thinking” and to “improve and expand the areas of our research”. The tools that are already available can be categorized according to their intended purpose, serving, for example, as aids in authorship studies, analyses of stylistics and linguistics, or statistical and text-critical studies (Tov 2006:338–342). Making use of interactive visualization tools could pave the way to more flexible Biblical Hebrew linguistic software.

In addition to the representation of linguistic data, a comprehensive Biblical information system should, according to Bothma (1995), include images of textual-critical material and cultural-historical objects in order to facilitate the preservation, publication, and research of the ancient manuscripts. Multiple disciplines and teamwork are necessary for the creation of such a system, as no researcher could have all the skills needed to construct the various building blocks. Bothma (1992b:348) highlights the cooperation between linguists, theologians, and IT specialists necessary to build well-designed Biblical Hebrew information systems. Although there might not be many researchers who have an in-depth command of all of these disciplines, the members of the team should have a basic understanding of the complex nature of one another’s abilities and fields.
6. Current and Future Projects

An international conference is organized every three or four years by the Association Internationale Bible et Informatique (AIBI). This conference and its proceedings of selected papers provide a comprehensive view of past, current, and future projects (see AIBI 1986 through AIBI, forthcoming). The latest (2008) conference’s proceedings are currently in press (Vegas et al., forthcoming) and aims to give a critical evaluation of past and present tools used in computer-assisted Bible research.

The electronic databank of Wolfgang Richter (the Biblia Hebraica transcripta project [BHt]) consists of four levels (word, word group, clause, and clause combinations) and will be made available by Logos (Rechenmacher and Van der Merwe 2005). The third and fourth levels are still under construction. Winther-Nielsen (2009, forthcoming) plans to integrate his two projects, PLOT and RLM, referred to above, into a single tool for technology-enhanced learning. Andersen and Forbes’s (2010) ongoing work involves refinement of their syntactic/semantic mark-up and the introduction of a rigorous discourse analysis. With reference to Holmstedt’s and Abegg’s project, as of April 2010, the programming of the search engine and the tree diagramming display, done by Oaktree Software for their Accordance Bible program, have been completed. The Hebrew epigraphic texts and various biblical and Qumran texts have been tagged. The search engine and the first stage of the tagged texts were released at the National SBL meeting in Atlanta, GA, in November 2010. The projected completion date for the entire project is the summer of 2012. The Groves Center would like to expand the WLC to the entire codex, the WHM to XML format with additional annotation for each morpheme, and add case role and other semantic tagging to the WHS. The Center intends to extend the WHS to the text level and to explore data-visualization techniques to ‘teach’ the computer to discover and display patterns of data that exist in more than the three or four dimensions that humans can easily perceive.

7. Conclusion

Recognition should be given to the excellent, in-depth work that has already been done by Biblical Hebrew scholars in regard to computer projects. Some gaps in the current body of work suggest possibilities for further research. Although some of the projects, discussed above, do facilitate rather advanced searches, they do not clearly differentiate between the linguistic levels of syntax and structural semantics; neither do they facilitate comparative studies on and between these levels (with the exception of Winther-Nielsen’s pilot project). A trans-disciplinary approach is needed to combine expertise from the fields of Information Systems and Biblical Hebrew studies to build further on the vast body of research that has been done in this area. Systems that integrate the results of divergent computational linguistic projects in Biblical Hebrew
could promote the use of electronic data and analyses, providing a solution for the under-utilization of existing tools. New developments that tend to make use of more flexible functionalities and user-friendly visualizations may facilitate the creation and use of advanced Biblical Hebrew information systems in the next decade.

*Computational information systems (referenced)*


Syntactic Database of Ancient Hebrew. http://individual.utoronto.ca/holmstedt/Ancient_Hebrew_Syntax_Database.html


References


Bible in the Age of Technology”, Rome, April 2004. (Copy on file with authors.)


—. Forthcoming. “Persuasive Hebrew exercises: The wit of technology enhanced language learning”. (Forthcoming in a Festschrift.)