The Accordance Hebrew Syntactic Database Project

Robert D. Holmstedt
University of Toronto; Canada
University of the Free State,
South Africa

John A. Cook
University of the Free State,
South Africa
john.cook@asburyseminary.edu

Abstract

The Accordance Hebrew Syntax database is the result of a decade of collaborative planning and research. The origin of the project lies in a research grant proposal written by Robert Holmstedt (University of Toronto) in 2007. At that time, two other databases had become accessible to the public: 1) the WIVU Emdros database of the Werkgroep Informatica of the Vrije Universiteit in Amsterdam, now administered by the Eep Talstra Centre for Bible and Computer (see footnote 1) and presented as the ETCBC database (see footnote 2), and 2) the Andersen-Forbes Analyzed Text of the Hebrew Bible (see footnote 3). The initial motivation for proposing a third database was straightforward—to be able to use a database created upon a model of Hebrew syntax that differed from the two existing databases.

Keywords: Accordance Hebrew Syntactic Database; Hebrew Bible; Hebrew database

Introduction

The Accordance Hebrew Syntax database is the result of a decade of collaborative planning and research. The origin of the project lies in a research grant proposal written by Robert Holmstedt (University of Toronto) in 2007. At that time, two other databases had become accessible to the public: 1) the *WIVU Emdros* database of the Werkgroep Informatica of the Vrije Universiteit in Amsterdam, now administered by the Eep Talstra Centre for Bible and Computer[1] and presented as the *ETCBC database*[2] and 2)

---

1 https:\\www.godgeleerdheid.vu.nl/en/research/institutes-and-centres/eep-talstra-centre-for-bible-and-computer/index.aspx.

2 https://shebanq.ancient-data.org.

the *Andersen-Forbes Analyzed Text of the Hebrew Bible*.[3] The initial motivation for proposing a third database was straightforward—to be able to use a database created upon a model of Hebrew syntax that differed from the two existing databases.

No doubt because interest in Hebrew syntax is the domain of a relatively small group of scholars, even among Hebraists, the grant proposal was not initially successful.[4] The silver lining in the process was one of the external reviewers: Martin G. Abegg Jr (Trinity Western University, emeritus). After the grant cycle had ended, Abegg contacted Holmstedt to propose a larger project: a syntactic database covering not just biblical texts, but one that included all ancient Hebrew texts, from epigraphs to Dead Sea Scrolls. The full scope of ancient Hebrew texts would allow deeper research into diachronic syntactic development, thus contributing to the ongoing research in the history of Hebrew. Additionally, Abegg was also well known for database work since he had been instrumental in creating early digital texts of the Dead Sea Scrolls. This prior experience had resulted not only in a well of wisdom regarding the creation of digital texts with metalinguistic tagging, but also relationships with Bible software companies, especially with Roy B. Brown (Oaktree Software), the creator of Accordance Bible software. Thus it was that in late 2008 the Accordance Hebrew syntax project was born.

Besides including all ancient Hebrew texts, another feature of the envisioned databases was a tight focus on syntax, grounded in (but not bound by) Chomskyan generative linguistic theory. However, the levels of language (phonology, syntax, and semantics) are not discrete,[5] and it quickly became clear that we would be obliged to deal with some

---

3   www.andersen-forbes.org.
     Another syntactic database, the *Westminster Hebrew Syntax* database, has been underway since 2009 (http://www.doxologypress.org/sites/groves/B/?page_id=19 accessed June 12, 2017).
4   The project was later partially funded with a 2010-2011 grant (#410-2010-796) from the Social Sciences and Humanities Research Council of Canada.
5   For readers aware of theoretical debates, our statement on the syntax and semantics interacting does not violate the actual (often mischaracterised) generative principle, the "autonomy of syntax". On the linguistic blog "Faculty of Language", Norbert Hornstein nicely describes both what the Autonomy of Syntax is and is not: "Generative Syntax endorses the autonomy of syntax thesis (AOS). Though AOS has often been misunderstood to assert that there is no relation between the grammar and meaning, it actually means that the primitives and operations of the grammar are independent of the contents of what they are used to express. In particular, syntactic categories, principles and operations to not reduce to semantic ones.... Precisely because the syntax is autonomous it is able to combine information from different encapsulated modules. In other words, autonomy is just the flip side of not being modularly restricted. The intra modular primitives and operations cannot do this, which is what makes it impossible for rats, young kids and linguistically distracted adults from combining different kinds of information (i.e. predicates from different modules). From the present perspective, a more revealing term for the autonomy of syntax might be the inter-modularity of syntax, autonomy being precisely the property we want in a tool required to combine diverse types of thoughts and concepts, ones otherwise confined to specialized cognitively encapsulated modules" (Hornstein 2011).

semantic features of verbs, specifically to determine whether certain prepositional phrases were the "objects" of certain verbs or not. Thus entered the early guidance and ongoing participation of John A. Cook (Asbury Theological Seminary), whose research on the verb added a critical underpinning to the project, especially as he began to articulate a valency model for Biblical Hebrew verbal syntax.

The resulting first-fruits of this rich collaboration are the Accordance syntax databases for Hebrew inscriptions, the Hebrew Bible, and the Hebrew texts from Qumran. In this paper, we will focus on the design principles and research capabilities of the syntactic database on the Hebrew Bible.

Linguistic Theory

There are numerous complexities involved with an ancient language syntactic database project. They range from issues of *audience* to *theory* to *programming*. Among other questions, we asked ourselves (repeatedly, in many cases):

- Who will use this database and what will they expect to see?

- How much can we draw upon linguistic theory—and which one?—while still making the modules usable for the broadest audience?

- How much theory-internal structure can we set aside yet not produce a scientifically naive and theoretically flawed database?

Our primary goal for the creation of the database is to produce a usable research tool for the academic community. Determining syntactic relationships, however, not only requires judgment, which is necessarily subjective, but also depends on one's theory of grammar. To think that such a project can be accomplished *without* a theory would be like saying that exegesis can be performed without a methodology or that interpretation can exist in a vacuum, without a hermeneutical theory. It is simply not reality; even if an exegete or interpreter is unaware or ignorant of theories and methodologies, there is always a framework in which analysis occurs (however coherent that framework may or may not be). And yet, although some of us have situated our research on Hebrew syntax within the linguistic approach of generative grammar, specifically as it is articulated within the program of Chomskyan minimalism (Chomsky 1995; Radford 1997; Boeckx 2006, 2008), we knew that to base the database and its underlying tagging scheme on a fully articulated minimalist framework would be inappropriate. Not only would its usability be severely limited, since it is unlikely that most users of the database will subscribe to Chomskyan linguistics, but given the ever-changing nature of linguistic theory, the database would risk becoming obsolete before it was finished.

To keep our balance on a very narrow beam, we sought to develop a metalinguistic "tagging" scheme that reflected what became our motto: "**data primary, theory wise**". That is, while the project team has read broadly in linguistics, from various types of
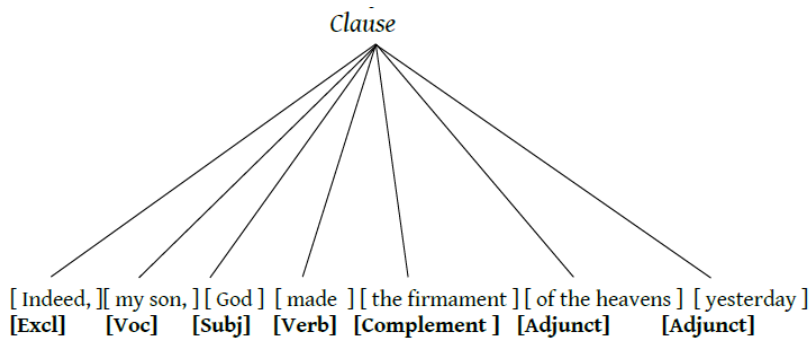
functionalism and typology to generative grammar, it was important for the project that the usability and accessibility dictated our use of linguistic theory. Four defining features of the database design illustrate our balancing act: inclusion of null constituents, hierarchical, non-binary phrase structure, non-movement approach to discontinuous constituents, and narrow syntactic focus.

Inclusion of Null Constituents

A significant feature that distinguishes our database and illustrates its interaction with linguistic theory is its inclusion of null constituents.[6] Before addressing the "null" concept, we should clarify our use of "constituent". The syntactic elements at each stage of derivation are referred to as constituents. A constituent is a single syntactic unit that has a place within the hierarchy of a larger syntactic unit. It is important to recognise that morphological words and constituents *may* overlap but are not always identical. That is, a single word may represent more than one syntactic constituent, such as English *teacher's*, in which the constituent *teacher* has a syntactic role that is distinct from the syntactic role of the possessive *'s*.

This is the case in Hebrew, too. Moreover, the converse is also true: occasionally multiple words form one syntactic constituent, as with many proper nouns, such as בֵּית לֶחֶם *Bethlehem* "House of Bread" (see illustration at left). We analyse complex prepositions similarly. For example, מֵעַל פְּנֵי is decomposable morphologically as "from.upon the.face.of". But syntactically we analyse this string of words as a single syntactic constituent, a preposition with the dominant meaning "from" (see illustration at right).

---

6   On null constituents in Hebrew, see Creason (1991); Naudé (1991, 2001); Holmstedt (2013).

Constituents within a hierarchical clause structure approach stand in some tension to an analysis based on parts of speech. Parts of speech are inadequate for syntactic analysis. Using the parts of speech labels typically used for Hebrew, some may suffice for syntactic description, so that *verb* and *adjective*, for example, may also describe the syntactic roles those words play; however, the other parts of speech labels, *noun*, *pronoun*, *preposition*, and the umbrella label *particle*, are wholly opaque concerning the syntactic relationships between these words and any others in a given clause. Therefore, syntacticians often use a different set of labels for the various constituents in a clause. The core labels are *subject*, *predicate* (or *verb*), *complement*, and *adjunct*, with the non-core constituents (in our database) *vocative*, *exclamative/interjection*, *parenthesis*, *appositive*, and *casus pendens*.

Returning to the matter of "null" constituents, we follow the generative principle that every phrase has a "head", whether a "verb" for a predicate phrase or a noun or similar nominal(ised) constituent for a subject phrase. Therefore, we have inserted a null marker (0) in every phrase that lacks an overt head. The use of null constituents is most common in the subject position, since Hebrew allows an overt subject to be omitted, as in example (1), and nearly as common in Hebrew is the use of a null copula in the predicate position, the so-called verbless clause, as in (2):

(1) ⊙יִּשְׁבֹּת ___ בַּיּוֹם הַשְּׁבִיעִי מִכָּל־מְלַאכְתּוֹ

"**(he)** rested on the seventh day from all his work" (Gen 2:2)

(2) וְחֹשֶׁךְ ___ עַל־פְּנֵי תְהוֹם

"and darkness **(was)** upon the face of

the deep" (Gen 1:2)

In addition to null subjects and predicates, Hebrew also allows null complements and null relative clause heads. All of these null items have been included and tagged appropriately in our databases. No doubt some will look through the short list of syntactic roles above and ask, "Where is the direct object? And what about the indirect object?" The answer is that they are not syntactic relationships that are explicitly tagged in our database. Why? The answer to that is more complex, but here is the beginning of an explanation.

The **complement** category essentially incorporates the concept of "object", of which there are a number of sub-types. The direct object is the *accusative* (to borrow a case term), or a nominal (non-prepositional) constituent that is the person or thing undergoing the (active, transitive) verbal action or process, i.e., the "patient". In contrast, the indirect object is limited to a small set of verbs that require a "recipient" (or "beneficiary") of the verbal action or process to be specified. There are two basic problems with encoding the concepts of direct and indirect object in a syntactic database, especially one for Hebrew. First, these concepts are not exclusively syntactic in nature; one must necessarily interact with argument structure (or thematic role) information concerning the predication, information that is explicitly outside the scope of our *syntactic* database (see further "The Narrow Syntactic Focus (+Verbal Valency)" below). Second, whereas direct objects in English are always in the accusative case (i.e., non-prepositional), verbs in Hebrew (and Greek) are varied in their selection of a syntactic constituent as their object: some select a non-prepositional constituent, while others select some type of prepositional constituent. In sum, using complement allows us to capture a greater generalisation, regardless of the type of constituent, whether non-prepositional, prepositional, or even clausal. In contrast to the complements, which are "licensed" by a verb's syntactic-semantic features, **adjuncts** are those constituents which are unlicensed and so may be recursively applied (see further "The Narrow Syntactic Focus (+Verbal Valency)" below).

Hierarchical, Non-Binary Phrase Structure

There are two basic options for clause structure: a flat clause structure and a hierarchical clause structure. The flat clause structure is based on a finite state model, the "Markov Model" (Malmkjaer 2002, 138–139), applied to language, in which it is argued that a clause is constructed word-by-word in a linear fashion; clauses in this model are also called "word chains". In this model, which is often associated with computational linguistics, it is proposed that the speaker has a simple mental system that allows him or her to make a decision about the appropriateness of each word as it is added to the clause-in-making and, when all the given words are added, the product is either accepted or rejected based on a final analysis. An example of a flat-structure clause is given in **Error! Reference source not found.**.

(3)   Flat-structure clause analysis

[Indeed,][my son,] [God] [made] [the firmament] [of the heavens] [yesterday]
[Excl]   [Voc]   [Subj] [Verb] [Complement] [Adjunct]   [Adjunct]

The central problem with this flat structure model of the clause is the inability to account for long-distance syntactic relationships, in which two syntactic elements that somehow depend on each other are separated by an arbitrary number of words. For example, in 0a–b, the subject and verb are adjacent and so the subject-verb agreement is immediate, or "local"; in 0c, however, the agreement is non-local or long distant.

(4)

    a. The [baby $_{SG}$] [cries $_{SG}$].

    b. The [babies $_{PL}$] [cry $_{PL}$].

    c. The [babies $_{PL}$] in the nursery [cry $_{PL}$].

In contrast to the flat structure, the hierarchical approach to clause structure is not primarily linear but, as its name signals, hierarchical. The syntactic elements relate to each other in terms of how they "cluster" together. For example, for the clause *she hit her sister with the teddy bear*, in 0, we might suggest that "she" and "hit" relate to each other non-hierarchically, as the two basic halves of the clause. But we would not put rest of the clause on the same level: the words "her sister", which seem to belong together, and the words "with the teddy bear", which also seem to form a group, both seem to form a group with the verb "hit". These hierarchical relationships are typically represented by brackets or trees:

This hierarchical clause structure can also account for how long-distance dependencies exist, illustrated in **Error! Reference source not found.**.

(5) [She] [hit [her sister] [with the teddy bear]].

She hit

her sister with the teddy bear

(6) The [babies PL] in the nursery [cry PL].

Subj PL◄─►PL *Verb*
cry
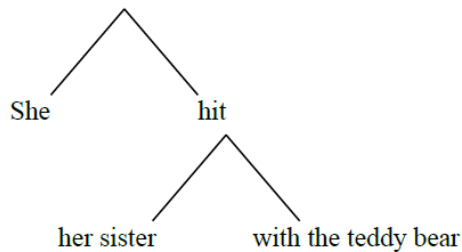
The babies in the nursery
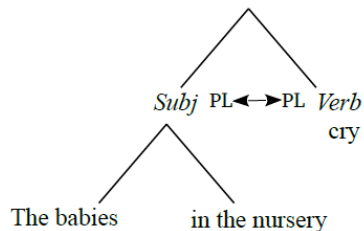
In this example, the element "in the nursery" is hierarchically dominated by "the babies". This allows the plural "the babies" to be hierarchically adjacent to the plural verb "cry", thus providing an explanation for how the subject and verb may agree even though they are separated by other words.

The process of formation is from the bottom-up, that is, as each lexical item is introduced into the "clause-in-the-making" (called a "derivation"), the lexical items merge with each other and project a larger structure, a phrase. The lexical item that gives the phrase its syntactic identity is the phrasal head. Thus, a prepositional phrase is the projection of the hierarchy around a preposition, a noun phrase is the projection of a noun, a verb phrase the projection of a verb, etc.

The highest level constituent is a clause. A clause is a single constituent consisting of a subject phrase and a verb phrase. Main clauses (or "independent") are self-contained and thus do not function within a larger *syntactic* hierarchy, while subordinate (or "dependent") clauses are contained within a phrase, typically a verb phrase in a higher clause.

The point of this discussion of hierarchical clause structure has been to explain how we designed our database on a well-known linguistic theory of phrase structure, in which it is argued that constituents *are contained within* larger constituents, all the way up to the clause level. For each word, we had to make a decision regarding the word's location in the syntactic hierarchy—within what other constituent does it reside? And for that resulting complex constituent, the same question must be answered, until there are no more constituents and one is left with a clause.

The clause itself seems to consist of **two basic parts**: a *subject phrase* (no matter how simple or complex) and a *predicate (or verb) phrase* (no matter how simple or complex), as illustrated in 0. Thus, in terms of the basic elements of a clause, the hierarchy that we have followed is *binary* in nature.

(7)   Clause structure
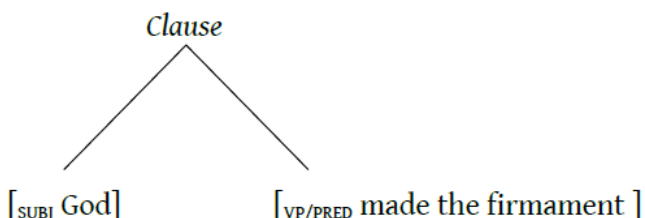
The tree diagrams in the Accordance syntax database account for both hierarchical levels in Hebrew syntax as well as the basic division of each clause into subject and predicate, as the tree in **Error! Reference source not found.** illustrates.

(8)   Accordance tree diagram for Gen 4:1
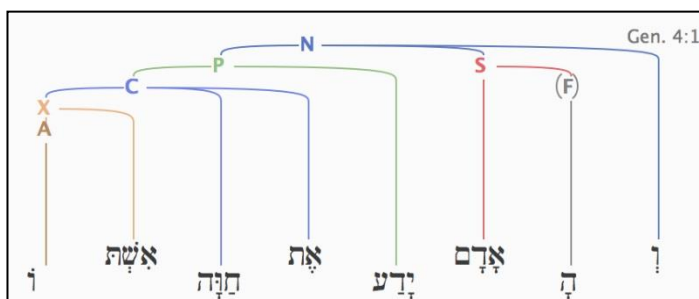
Notice that under the N node (which represents iNdependent clauses), there are two halves of the clause, the Subject (S) and the Predicate (P). The subject has a speciFier (F), i.e., the article, and within the predicate there is the verb and its Complement (C). The complement is a noun phrase that is itself modified by an appositional (X) phrase.

Earlier forms of Chomskyan generative syntax allowed for "*n*-ary" branching (i.e., whatever number of branches appear to be required) and some non-Chomskyan generative frameworks maintain an *n*-ary principle of phrase structure. However, since the mid-1980s, the Government-and-Binding model, followed by minimalist syntax, adopted a strictly binary approach to constituent structure.[7] But the addition of clause-edge constituents, such as dislocations (*casus pendens*), vocatives, and exclamatives results in a tree that is not easy to fit into a binary structure and to do so requires a good deal of theory-internal arguments, as illustrated by the non-binary hierarchical analysis of **Error! Reference source not found.** in **Error! Reference source not found.**.

(9)   Hierarchical, non-binary clause analysis; cf. (3)

[Indeed,] [my son,] [God] [made [the firmament [of the heavens]] [yesterday]]]
[Clause [Excl] [Voc]   [Subj]   [Pred [Complement [Adj to Comp]] [Adjunct]]]
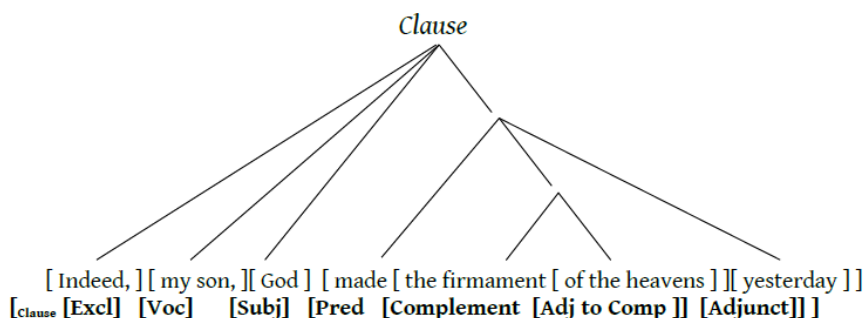
Thus, we made the decision to depart from a basic principle of this particular theory in favour of presenting hierarchical data in a manner that is not so theory dependent. Here, data-presentation for a broader audience outweighed theoretical preference.

The non-binary structure of some examples in our database is apparent in clauses in which the verb has multiple complements, a complement and adjunct(s), or when "edge" constituents such as dislocations are present. Two such examples are provided here.

---

7   Though minimalist phrase structure is explicitly hierarchical and binary, and supports this position with data and an appeal to theoretical economy, there is certainly no consensus among linguists in general on whether the syntactic structure of human language reflects a flat or hierarchical structure, and if the latter, whether the hierarchy is fundamentally binary or "*n*-ary" (Culicover and Jackendoff 2005, 112–116; Berg 2009, 33–56, 325–328; Guevara 2007). Of course, empirical support and theoretical advantage are claimed by everyone who bothers to make an explicit defense of their phrase structure(s).

(10)   Accordance tree diagram for Gen 3:22

In (10), the first prepositional phrase after the copular verb is the complement, and the second prepositional phrase (containing an infinitive clause) is an adjunct. The tree representation of this in our database results in a ternary structure for the predicate phrase, which is patently disallowed in Chomskyan generative syntax. Similarly, the subject phrase in Deut 5:25, in 0, is ternary.

(11)   Accordance tree diagram for Deut 5:25

The head of the subject phrase, the noun הָאֵשׁ, is followed by both an adjectival adjunct (הַגְּדֹלָה) and a demonstrative pronoun adjunct (הַזֹּאת). While in generative syntax both these modifiers would typically be situated hierarchically in the binary phrase structure, we have collapsed the hierarchy and allowed a non-binary structure.

Non-Movement Approach to Constituent Discontinuity

Constituent movement is a hallmark of transformational generative grammar (Brown 2010), although it has been dismissed by some non-Chomskyan generative theories (i.e., "monostratal" theories). The basic idea is that the linear order of constituents in many actual clauses cannot reflect the "original" order of those constituents. Neither

defending nor criticizing this proposal, we determined that *representing* it in our database was not desirable or necessary. Yet, we were forced to deal with the results of constituent movement, that is, discontinuous constituents—constituents that are divided into parts separated by un-related constituents. This happens less in English than in Hebrew, although it does occur with some English relative clauses, as in 0:

(12)   12[A new king] arose over Egypt, [who had not known Joseph]

In this relative clause clearly modifies the NP "a new king", and yet it is separated from this NP by the VP "arose over Egypt".

In Hebrew, discontinuity is extremely common, since many narrative clauses begin with the *wayyiqtol* narrative verb, switch to a subject, and then continue with the rest of the predicate, as in 0.

(13)   וַיַּרְא אֱלֹהִים אֶת־הָאוֹר

and-saw God DOM[8]-the-light

"and God saw the light" (Gen 1:4)

The challenge of constituent discontinuity is that, based on the hierarchy and the projection principle that a phrase contains all its complements and/or adjuncts, a verb and its modifiers together make up *a single constituent.* But how, then, can this be represented when they are broken by non-related intervening constituents, such as a subject?

To account for discontinuous constituents, we employ a system of cross-referencing, which allows us both to include discontinuous constituents in syntactic searches and display the connection in the tree display. We have used this cross-referencing system to allow us to represent more accurately three additional phenomena: dislocation (*casus pendens*), resumption in relative clauses, and ellipsis (or "gapping").

---

8   We take אֵת to be a "differential object marker" (see Bekins 2014).

The Narrow Syntactic Focus (+Verbal Valency)

A final defining principle of the Accordance syntax database that we will mention here is a *narrow focus* on syntax. That is, the tagging scheme provides phrasal, clausal, and inter-clausal information to the exclusion of semantic judgments, discourse relationships, and implicational pragmatics. This decision on the narrow syntax (+ valency) focus of our database was made for two practical reasons:

1. Every additional layer adds an increasing amount of subjectivity, and we intended this research tool to be as broadly usable as possible.

2. The additional semantic and pragmatic layers would add a disproportionate number of years to the project. Whereas we have completed all our ancient Hebrew texts within a decade, it would likely take twice that time (if not much more) to produce a multi-layered database.

This narrow syntactic focus of the database results in a powerful syntactic research tool that leaves non-syntactic judgments to the user. For example, when the particle כִּי is a subordinator, we make no distinction between its use as a temporal ("when") subordinator or a clausal ("because") subordinator. Those distinctions are left to the user to determine. What we do provide is the distinction between כִּי as an adjunct subordinator (temporal or causal), a complement subordinator ("that"), a conjunction ("but"), and an exclamative ("indeed!"). Our narrow focus on syntax is not, however, blind to the interaction between syntax and semantics (see above, n. 5). This is true particularly in one specific area verbal semantics—*verbal valency information*, which we associate with the lexical entry of a verb (see Cook 2014, 2016). The term valency derives from chemistry and has been employed in linguistics for about a half-century. Verbal valency, in particular, refers to the properties of a verb that determine the syntactic structures in which it may function grammatically. That is, the verb selects or "licenses" certain types of constituents (syntax) with specific roles (semantics) to satisfy a well-formedness constraint associated with the verbal lexeme.[9] For example, compare the structures in (14): the English verb *snored* licenses a subject, *help* licenses both a subject and an NP complement, and *returned* licenses a subject and prepositional (locative) complement:

> (14) Types of complements
>
> She *snored*.
>
> He *helped* the boy.

---

9   Looked at from the reverse direction, adjuncts are recursive whereas complements are limited by the well-formedness constraints of the verb.

They *returned* to the house.

For the Accordance syntax project, it was necessary that we employ valency information to determine whether the non-subject constituents associated with a given verb were complements or adjuncts. And yet, in the database we do not identify these complements or adjuncts by any semantic categories, such as locative, temporal, means, manner, etc. Moreover, we do not include any discourse-pragmatic judgments in the database, such as whether a complement preceding a verb has a Topic or Focus function. This strategy, as stated above, enables us to capture greater generalisations in the database (see "Inclusion of Null Constituents" above). To illustrate, compare contrastive examples in 0. Employing complement-adjunct language, we are able to generalise the grammatical well-formedness of *Qal* אחז despite the variation between a NP complement in one instance and a ב-PP complement in another. By contrast, generalizing the trivalent structure of *Qal* נתן for contrastive examples like those in c.-d. highlights the semantic variation of "give" and "place" as (typically) determined by the type of complement, whether an inanimate locative one ("in the dome of the sky") versus an animate beneficiary one ("to your offspring").

> (15)   Complement variation for the same verb

> יֹּאחֲזוּהוּ פְלִשְׁתִּים

> "The Philistines seized <u>him</u>." (Judg 16:21)

> וָאֹחֵז בְּפִילַגְשִׁי

> "And I seized <u>my concubine</u>." (Judg 20:6)

> וַיִּתֵּן אֹתָם אֱלֹהִים בִּרְקִיעַ הַשָּׁמָיִם

> "He set them in the dome of the sky." (Gen 1:17)

> לְזַרְעֲךָ אֶתֵּן אֶת־הָאָרֶץ הַזֹּאת

> "To your offspring I will give this land." (Gen 12:7)

What makes the complement-adjunct distinction particularly important—and particularly challenging—for the syntax project is the inclusion of null constituents ("The Narrow Syntactic Focus (+Verbal Valency)" above) because this necessitated that we determine when complements of the verb needed to be represented by null. Among the most frequent cases of null complements are resumptive constituents in relative clauses and in elliptical poetic couplets, as illustrated by the examples in **Error! Reference source not found.**.

Null complements

(16)

a. וַיִּשְׁבֹּת בַּיּוֹם הַשְּׁבִיעִי מִכָּל־מְלַאכְתּוֹ
אֲשֶׁר עָשָׂה

"(He) rested on the seventh day
from all his work that he had done
(it)." (Gen 2:2)

b. בְּךָ בָּטְחוּ אֲבֹתֵינוּ בָּטְחוּ וַתְּפַלְּטֵמוֹ

"In you our fathers trusted; they
trusted (in you) and you saved
them." (Ps 22:5)

Retrievable and Non-Retrievable Data

Now that we have briefly described the history, purpose, and linguistic underpinnings
of the Accordance syntax project, we shall turn to two practical questions: *What kinds
of grammatical data can be retrieved?* and, *What kinds of grammatical data remain
elusive?* The most obvious types of data that can be retrieved are those associated with
the distinctive features of the database itself: **word order** and **null elements**. For
instance, one interesting search the database allows is to find Past Narrative forms that
are preceded by their subject phrase, which turns up a number of examples with complex
subject phrases, as in 0.

(17)   Past Narrative with subject-verb word order

הַנְּבִיאִים אֲשֶׁר הָיוּ לְפָנַי וּלְפָנֶיךָ מִן־הָעוֹלָם וַיִּנָּבְאוּ אֶל־אֲרָצוֹת רַבּוֹת וְעַל־מַמְלָכוֹת גְּדֹלוֹת לְמִלְחָמָה וּלְרָעָה וּלְדָבֶר:

"The prophets who were before me and before you from ancient times <u>prophesied</u> against many lands and against great kingdoms war, famine, and pestilence." (Jer 28:8)

The ability to search for **null elements** is one of the most significant features of the database, making it possible to search for null-copula/verbless clauses **Error! Reference source not found.**a. and for elliptical structures in poetry **Error! Reference source not found.**b.

(18)   Searches for null elements

a. Null-copula/verbless clause

(excluding participles)

וְחֹשֶׁךְ עַל־פְּנֵי תְהוֹם

"And darkness (was) upon the face of the deep" (Gen 1:2)
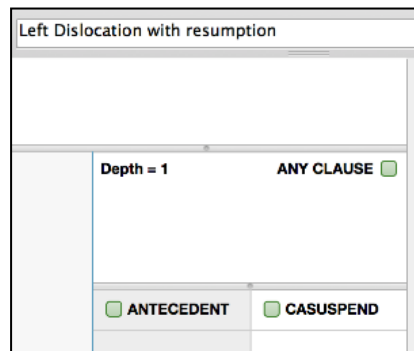
b. Verb gapping

אֱלֹהַי אֶקְרָא יוֹמָם וְלֹא תַעֲנֶה וְלַיְלָה וְלֹא־דוּמִיָּה לִי:

"My God, <u>I call</u>ᵢ by day

and you do not answer; and (<u>I call</u>ᵢ) by night I have no quiet." (Ps 22:3)
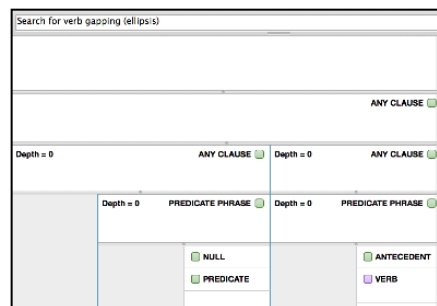
The same feature that links the verb with its associated "gap" (antecedent) in **Error! Reference source not found.**b. in the database also enables searching for **left dislocation** (aka *casus pendens*), which is the antecedent for an overt resumption within the clause.

(19)   Left dislocation (*casus pendens*)

וּבָנַי הִנָּם אִתְּכֶם

My sons_i, look they_i are with you.

(1 Sam 12:2)

Because the Accordance syntax is not a stand-alone database but is layered upon the Accordance graphemic and morphological data and integrated into the program's search interface, a rich array of searches is possible involving multiple layers of Hebrew grammar—phonology, morphology, syntax, semantics, and lexicon (and even, if desired, Masoretic טְעָמִים). Thus, by combining morphological and syntactic search parameters we can pull up substantival participles serving as subject 0a or as complement of a non-copular verb 0b.

(20)   Substantival participle

As subject

וַיָּרִיבוּ רֹעֵי גְרָר עִם־רֹעֵי יִצְחָק

"The shepherds of Gerar had a dispute with Isaac's shepherds." (Gen 26:20)
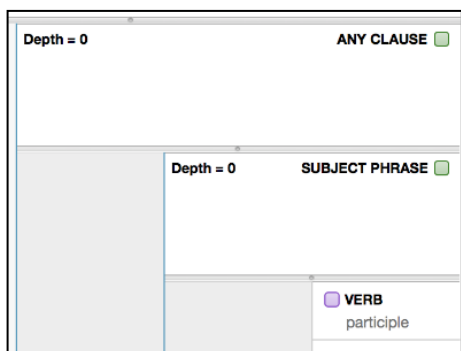
b. As complement (non-copular verb)

בְּמִבְחַר קְבָרֵינוּ קְבֹר אֶת־מֵתֶךָ

"In the choicest of our graves bury your dead." (Gen 23:6)
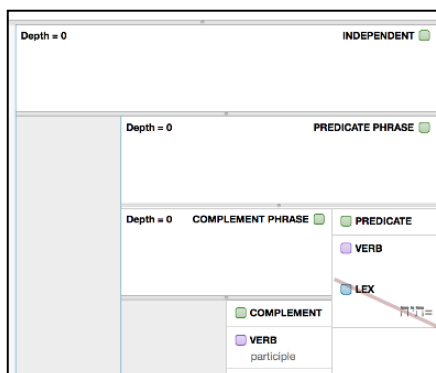
Similarly, by combining morphology and syntax we can search for subject-predicate lack of gender concord, as in **Error! Reference source not found.**, which involves both morphology and syntax.

(21)   Lack of gender concord

וְעָרֵיכֶם יִהְיוּ חָרְבָּה

"And your cities(fem) will become(masc) waste."

(Lev 26:33)
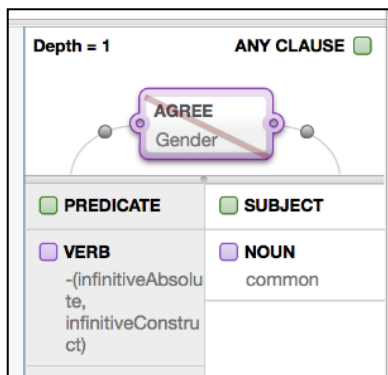
Accordance also allows syntax searches to be supplemented with its rich array of **analytic tools** for examining the results of searches. For example, we can see from the hits graph in **Error! Reference source not found.**a that lack of gender concord happens notably more frequently in Leviticus, Numbers, Job, and Proverbs. We can also discover from the chart in **Error! Reference source not found.**b a possible significant correlation between lack of gender concord and verb conjugation.[10]

(22)   Analysis of lack of gender concord.
       a. Search hits graph

---

10   The reason the cases of *yiqtol* are so much higher than *qatal* is undoubtedly due to the third-person plural common gender in *qatal* conjugation versus the *yiqtol*. However, this fact does not explain the paucity of Past Narrative (*wayyiqtol*) forms over and against *yiqtol*.
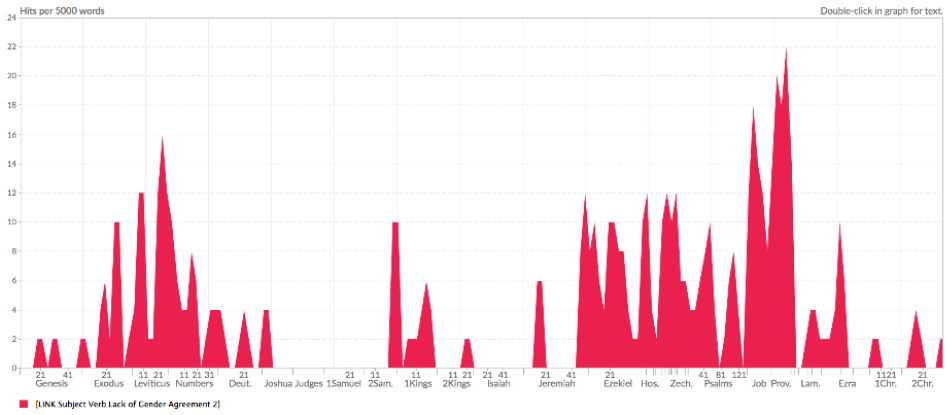
b. Correlation with verb conjugation
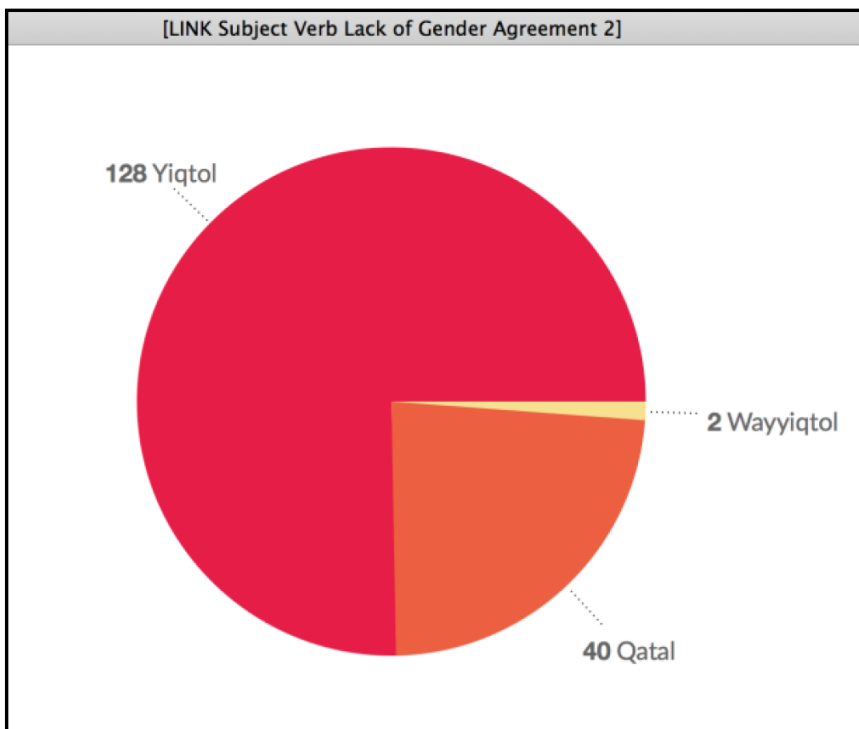
Finally, given the combination of layered information available with the syntactic focus of the database allows searching for highly complex sorts of constructions, such as an אם marked conditional clause with subject-verb word order and *yiqtol* predicate in the if clause and a *waw*-prefixed *qatal* in the then clause 0.

As already stated, the Accordance syntax is narrowly focused on syntax. That means that it does not contain semantic or discourse-pragmatic features. As such, semantic functions of constituents, topic and focus, participant reference, narrative structure, style features, and diachrony are not immediately retrievable from the database. These features can be analyzed to the extent that they intersect with the structure of the text— the syntax. For example, although the pragmatic functions of topic and focus are not distinguished in the database, dislocated structures are encoded and searchable 0. Given how subjective the determination of topic and focus functions are, we have preferred to leave that determination to the exegete once the dislocated structures have been retrieved from the database.

(23)   Conditional clause search
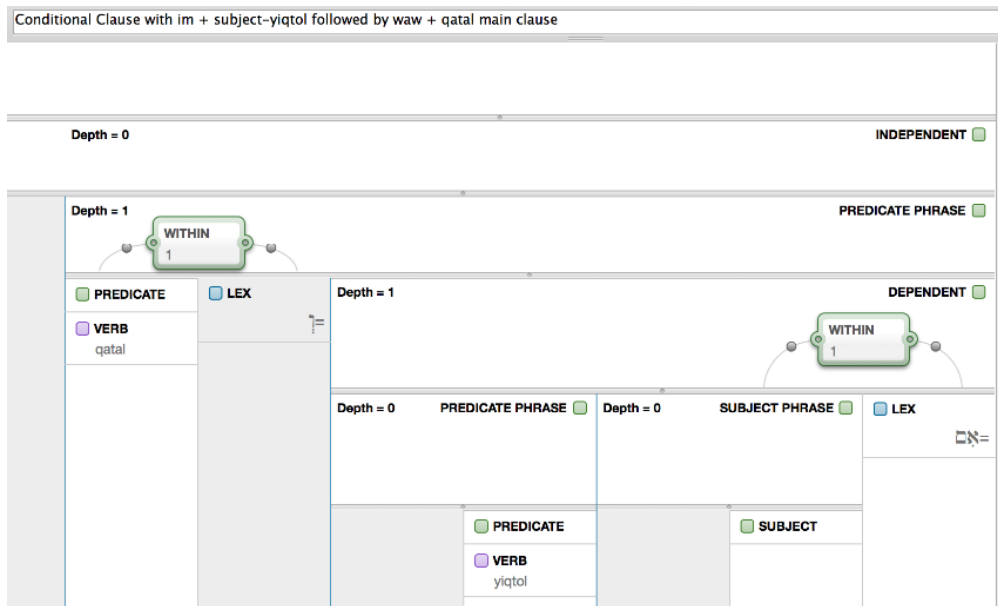
On the Place of Databases in Linguistic Research

The last two questions asked by the seminar organisers asks about the ability of the databases to retrieve features of "style" and address issues of diachrony. Our blunt answer is that the Accordance syntactic databases cannot speak directly to these issues, because they are not part of the structure of Hebrew grammar *per se*. But the posing of the questions themselves prompts some necessary reflection on the place of databases in the study of Biblical Hebrew.

We endorse linguist David Crystal's view that "style" is essentially "a descriptive convenience which summarises our awareness at any given moment of the controllable linguistic features that distinguish one use of English [or any language] from any other" (Crystal 1987, 222). Being able to accurately describe the manipulations of these linguistics features requires an understanding of "the socially-conditioned varieties a language may be shown to possess" (Crystal 1987, 202). Thus, for a database to encode style is to presume that the database creators have access to an ancient Hebrew native speaker's awareness of his or her social context, the linguistic options that may be relevant to that context, and their social implications. We do not claim to have such an awareness; indeed, we do not believe that the full array of options necessary to encode them into a comprehensive database are available to modern researchers.

So what is the use of a syntactic database for the user interested in matters of style? We see a syntax database as a concrete data research tool, providing information about the relationships between written words—words that are concrete historical artifacts due to their presence on the potsherd or manuscript page of the historical witnesses to the Hebrew language. Assessments concerning style, from identifying features of various authors in a multi-author work (which is likely for, e.g., the book of Isaiah) to formal versus colloquial registers to the manipulation of linguistic features to characterise speech in narrative (see, e.g., Holmstedt and Kirk 2016), move beyond syntax, beyond the concrete data, to formulate hypothetic profiles of ancient Hebrew language usage that may reflect conventions widely accepted at some time in ancient Israel or may reflect the idiosyncrasies of a single author.

Databases can reflect our analyses or be tools for our analyses, but they cannot form hypotheses for us. Their proper role is to test hypotheses about the grammar and its use within human discourse. Concerning style, a syntax database cannot encode stylistic differences, but the access it provides to the syntactic data may be able to test style hypotheses; similarly, such a database cannot encode diachronic differences, but it may be used to test diachronic hypotheses. To put it another way, the relationship between a linguistic database and a hypothesis concerning language use is akin to the relationship between a shovel and a garden. A shovel is undoubtedly useful for planting a garden, but the actual planting of the garden is far beyond the scope of the shovel's role as a tool.

References

Bekins, P. 2014. Transitivity and Object Marking in Biblical Hebrew: An Investigation of the Object Preposition "et". Harvard Semitic Studies 64. Winona Lake, IN: Eisenbrauns.

Berg, T. 2009. Structure in Language: A Dynamic Perspective. Routledge Studies in Linguistics. New York: Routledge.

Boeckx, C. 2006. Linguistic Minimalism: Origins, Concepts, Methods, and Aims. Oxford: Oxford University Press.

Boeckx, C. 2008. Bare Syntax. Oxford: Oxford University Press.

Brown, K. 2010. "Generative Grammar." In The Linguistics Encyclopedia 3rd ed., edited by K. Malmkjaer, 181–200. London: Routledge.

Chomsky, N. 1995. The Minimalist Program. Cambridge, MA: The MIT Press.

Cook, J.A. 2014. "Verbal Valency in Biblical Hebrew and the Case of מלא." Paper presented at the annual meeting of the Society of Biblical Literature, San Diego, CA, November 24.

Cook, J.A. 2016. "Verbal Valency: The Intersection of Syntax and Semantics." In Contemporary Examinations of Classical Languages (Hebrew, Aramaic, Syriac, and Greek): Valency, Lexicography, Grammar, and Manuscripts, edited by T.M. Lewis, A.G. Salvesen and B. Turner, 53–86. Piscataway, NJ: Gorgias.

Creason, S. 1991. "Discourse Constraints on Null Complements in Biblical Hebrew." University of Chicago Working Papers in Linguistics 7: 18–47.

Crystal, D. 1987. "Style: The Varieties of English." In The English Language, edited by W.F. Bolton and D. Crystal, 199–222. London: Penguin.

Culicover, P.W. and Jackendoff, R. 2005. Simpler Syntax. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199271092.001.0001.

Guevara, E. 2007. "Binary Branching and Linguistic Theory: Morphological Arguments." In Proceedings of the XXXII Incontro di Grammatica Generativa, Firenze, 2–4 March 2006, edited by M.C. Picchi and A. Pona, 93–106. Alessandria: Edizioni dell'Orso.

Holmstedt, R.D. 2013. "Pro-Drop." In Encyclopedia of Hebrew Language and Linguistics, edited by G. Khan, vol. 3, 265–267. Leiden: Brill.

Holmstedt, R.D. and Kirk, A.T. 2016. "Subversive Boundary Drawing in Jonah: The Variation of אשר and שׁ as Literary Code-Switching." Vetus Testamentum 66: 542-555. https://doi.org/10.1163/15685330-12341256.

Hornstein, N. 2011. "Darwin's Problem."
    http://facultyoflanguage.blogspot.com/2013/01/darwins-problem.html accessed 15 May
    2017.

Malmkjaer, K. 2002. The Linguistics Encyclopedia. London: Routledge.

Naudé, J.A. 1991. "Qumran Hebrew as a Null Subject Language." South African Journal of
    Linguistics 9 (4): 119–125. https://doi.org/10.1080/10118063.1991.9724303.

Naudé, J.A. 2001. "The Distribution of Independent Personal Pronouns in Qumran Hebrew."
    Journal of Northwest Semitic Languages 27 (2): 91–112.

Radford, A. 1997. Syntactic Theory and the Structure of English: A Minimalist Approach.
    Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
    https://doi.org/10.1017/CBO9781139166706.