



Hebrew Syntax

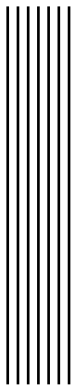
A Guide to the Westminster Hebrew Syntax Database

Kirk E. Lowery

January 14, 2009

DRAFT

DRAFT



Preface

A Word about Words

One of the difficulties in writing about language is, well, the language about language. In everyday speech, we use words about language in an ordinary, non-technical way. It's not a problem: everyone understands us. Like most scientific disciplines, when we study language we start using words in very specific, technical ways. Linguists will use the same words for language and grammar that we use in ordinary speech, but then modify the definition, usually to something more specific and precise. Because linguists have competing theories about language and grammar, they will use ordinary words in extraordinary ways and even coin new terms.

What makes things difficult for the non-specialist is that linguists will borrow technical terms from one another and then use them in ways that violate those terms definitions. This confuses not only the general reader, but the specialist (including me!) as well. Since I'm writing to readers with all the way from no knowledge of linguists to those professionally educated in linguistics, I'm now on the horns of a dilemma. If I use words in their non-technical sense, then I am limited in what I can say. I need those technical terms. But the moment I make choices about which technical term I will use, I will imply I hold to a particular theory of language when in fact I don't.

For example, I find the concept of "word class" and "slot" very helpful in describing syntax. These words are technical terms used in the theory of language known as "Tagmemics." Am I attempting to describe Hebrew syntax consistently using this

theoretical perspective? No. And some would criticize this decision, dismissing me as “eclectic,” implying my descriptions of Hebrew syntax will be somehow “inconsistent” and defective. Others will say, “I don’t agree with Tagmemics.” Assuming I do hold to this theory, they will tend to move on to other, more compatible descriptions of Hebrew.

And then there will be those who, trained and familiar with the “standard” descriptions of Hebrew syntax will be annoyed with my jargon. For example, there is no “nominative absolute” in our syntax database. Why? Because it is a 19th century term used first of all to describe Indoeuropean grammar. It does not work well with Hebrew and misses the point of the construction (**LEFT-DISLOCATION** and **TOPIC FOCUS**). It lacks explanatory value.

What is my theory of language? For the purposes of this book and the database, my approach is **DESCRIPTIVE**. My emphasis is upon **FORM**, hence modern theories of syntax such as **HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR** are attractive. Also, the fact that we used a computer to automate much of the analysis forced me into a more **CHOMSKYAN** world than I’m comfortable with, not least because we have no native speaker informants to verify our intuitions that **CHOMSKYAN** theories require. While I use the forms of language to validate my analysis, I also recognize that each form —morpheme, word, phrase, clause or larger structure —have a “meaning,” purpose or significance. Indeed, form does not exist without meaning and the only reality we experience with language is the result of the interaction of both. There’s no such thing as a “pure” formalist, but form does limit my subjective intuitions about meaning. As to the semantic side of things, I am very partial to **FUNCTIONALISM**, but cannot find a way to determine the **NUCLEAR PREDICATE** and **OPTIONAL ARGUMENTS** to the clause. This is because of the very limited sample of ancient Hebrew we have and the absence of native speakers to verify rules for us. The most we can do is make statements about distributions and consistency of occurrence of linguistic features.

For a complete, technical description of the linguistic labels and principles of analysis used to develop the database, see Appendix “x”.

Still, it is my responsibility to use technical terms in a way that does not confuse the reader. That is the purpose of the Glossary. While terms will be defined as we go, those definitions will be collected together in one convenient place.

Goals

While this book is being written, this “Preface” will serve as a place for all my meta-comments and thoughts about this writing. At the end, this section will be rewritten from scratch.

What am I trying to do here? I have several goals:

- Introduce syntax to those who have little knowledge of linguistics.
- Show users how to use the soon to be released *Westminster Hebrew Syntax Database*.
- Bring Hebrew Syntax as a (sub)discipline out of the 19th and into the 21st century.

I have several different classes of readers whom I would like to reach:

- The Bible software user who has no formal training in the biblical languages
- The average seminary student
- The formally trained Bible scholar

Now that I have these all in front of me, I can see a number of conflicts. The scholar will want and need all the academic apparatus and bibliography of technical publications. The ordinary user will be put off by this. I can assume the reader knows more if he is a scholar, and likewise the scholar will be impatient with the slow explanations needed for the ordinary user. And the poor seminary student is caught in the middle!


I am going to attempt to do what Roger Penrose, theoretical physicist and mathematician, has attempted to do in his writing: make no compromise with the subject, but write in such a way that the non-specialist can follow the exposition, albeit with effort. And at the same time, I need make the subject interesting enough for the specialist to want to review basic ideas, giving him new clarity and insight into questions long thought to be answered.

A tall order.

Context

I can see now that this present *Guide* is only the first of three books. The second should be a formal *Modern Syntax of Biblical Hebrew* and then a *Discourse Grammar of Biblical Hebrew*. This last project is one that I had originally envisioned for my Ph. D. dissertation. But when I got into it around 1980 I realized that a lot of foundational work had to be done before it could be written. That is when I changed my goal and topic to *Toward a Discourse Grammar of Biblical Hebrew*. The first two may be thought of as a “bottom-up” treatment of Hebrew grammar and the last --- by necessity --- is a “top-down” approach.

DRAFT



1 What is “Syntax”?

What is a “syntax database of the Hebrew Bible”? What do I do with it? Why should I care? This book answers these questions. I assume the reader knows some Hebrew, enough to read the Hebrew script, know what the basic grammatical features of Hebrew are and can puzzle out a sentence in Hebrew when given the parts of speech and basic meaning of each word. A person who has had at least one semester of formal instruction in Hebrew should be able to take full advantage of the help offered in this book.

In the following chapters we will look at what a “syntax” of anything is, what a syntax of Hebrew is, why these are important and how a database of these things can help answer questions about the meaning of the biblical text.

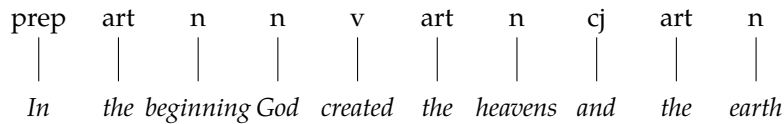
1.1 Thinking About Thought

Here is an example of a complete thought:

In the beginning God created the heavens and the earth.

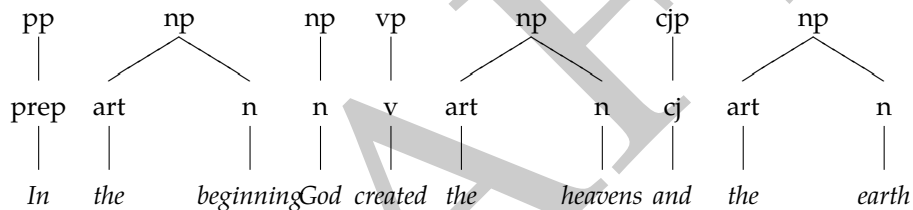
Linguists have named complete thoughts “clauses” (CL). “Sentences” are complexes of more than one clause. Different languages combine CLs by different rules, but for the moment, let’s just consider this CL from the standpoint of English. Each word belongs to one of a class of words, called its “Part of Speech” (POS). So we label each

word and arrange it graphically in an interlinear fashion (Tree 1.1).¹ This analysis is



Tree 1.1: A simple clause with parts of speech (Gen 1:1)

not very interesting. All it says is that each of these words has a label. But we can say more about this CL. Each of these words relate to other words around it. We can group these words together based upon those relationships and give that group a name (Tree 1.2). Now we have constructed "phrases", groups of words that share a



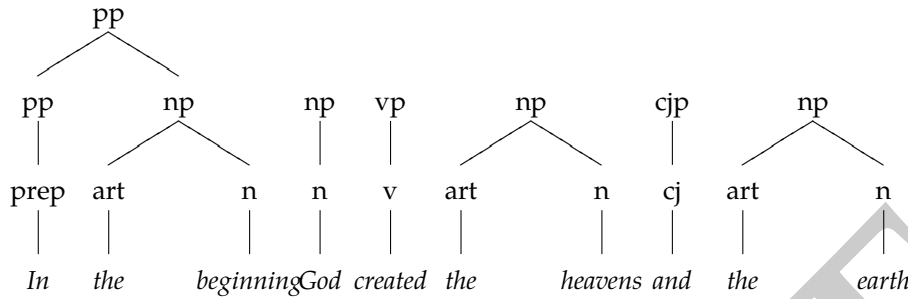
Tree 1.2: The simple clause grouped into simple phrases (Gen 1:1)

group name. That group name is taken from the "head" of that phrase. So a definite article plus a noun becomes a **NOUN PHRASE (NP)**, a noun by itself can be an **NP**, and so forth.² But we aren't finished yet. We see that one more group can be constructed in Tree 1.3, namely the formation of a **PREPOSITIONAL PHRASE (PP)**. We now have a group of phrases. What are their roles in the **CL**? Linguists have labels for them as well. We ask the question, What **ROLE** does this phrase play in relation to the verb? The next tree (Tree 1.4) connects all the phrases into one **CL** at the tree's root. Note the labels are in all capitals. This distinguishes *clause-level* names from *phrase-level* labels. This **CL** consists, in the order they occur, of a prepositional phrase (**PP**), followed by the subject (**S**) of the verb (**V**), ending with the object (**O**) of the **V**.

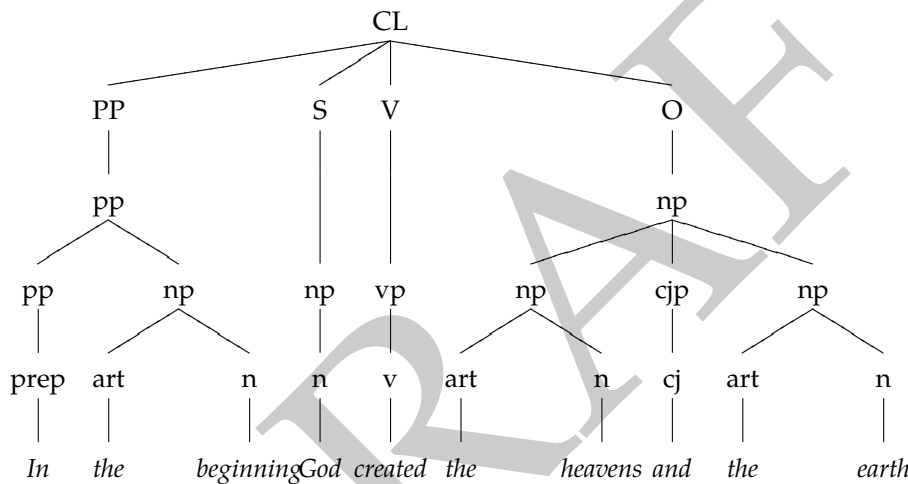
Linguists call this a syntax **TREE**. The tree is upside-down, with the root at the top, and the "leaves" at the bottom. Each line is a "branch" of course. The places where labels occur do not follow the tree metaphor, but are called "nodes". The higher node is the "parent" and the nodes directly stemming from a particular node

¹It doesn't look much like a tree now, but it will!

²The rest of the mnemonics are "conjunction phrase" (cjp), "verb phrase" (vp), and "prepositional phrase" (pp).



Tree 1.3: Phrases can be composed of other phrases



Tree 1.4: Clauses are composed of phrases that play specific roles

are called its “children”! Linguists are just as guilty of mixing metaphors as the rest of us. As we work from the leaves to the root, the act of extending a branch upward to a newly created node is “promoting” the child node to a new parent. Yes, yet another metaphor! And a final piece of jargon: the actual words themselves at the end of all the branches that have no children are “terminal” nodes. This terminology now permits us to talk about trees and the operations we make on them.

1.1.1 Syntax and Semantics

Let's step back and consider the analysis just performed above in its larger context of language. How do we go about understanding language? To most of us, language is like water to a fish. We know instinctively how language works and understand many of its pitfalls and twists. But when we come to a language other than our mother tongue, we come face to face with the need to make the sounds of that language somehow equivalent in our minds with the words and the rules of grammar about those words that we already know in our own native speech. First, we have to learn the equivalents between the meanings: "pen" is "toll" in Hungarian, unless the context of "toll" is a sentence about driving on roads that require one to pay for the privilege!

Linguists talk about "signs" of a language that are composed of a physical sound or visual glyph that points to some abstract or concrete idea.³ The sounds "pen" and "toll" point to the same object, a writing instrument. But what a sound points to depends upon a context. "Toll" in an English context points to a completely different idea than in an Hungarian context. Context matters.

When one speaks, there is a continuous stream of sound. This sound stream can be captured on an oscilloscope. If the stream is one continuous whole, how do we know where sentences and words begin or end? Somehow, as we associate meanings with segments of the sound stream, we begin to chop the stream into pieces in our minds. An individual sound "s" in English means "the previous chunk of sound refers to multiple objects", i. e., it makes a noun plural. These are called "phonemes" and are the minimum unit of meaning in a language. Phonemes group together to combine into larger constructs we call "words" and words combine into phrases. Phrases combine into complete CLs, CLs group into complex sentences and larger structures like paragraphs.

Each group, phonemes, morphemes, phrases, clauses and clause groups have a physical form and also have one or more meanings associated with that form. Words have a lexical or "referential" meaning. They point to an object or abstract concept. Clauses, however, have a different kind of meaning. Since they are complete thoughts, they represent what the speaker "means", or what the speaker *intends* to communicate.

To show how the semantics (meaning) and syntax of a clause are separate things, consider two famous sentences by Naom Chomsky:

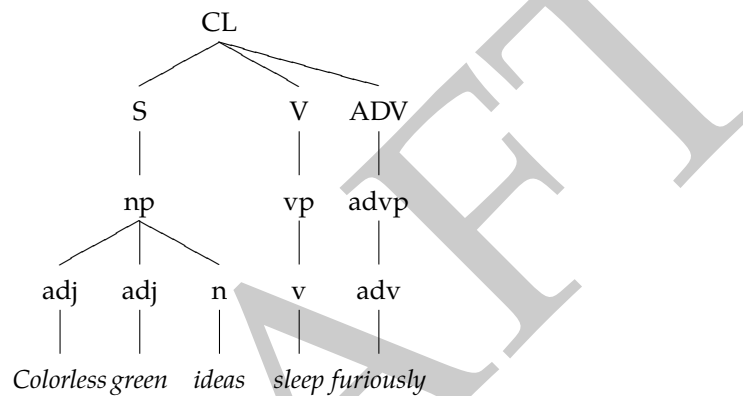
1. Colorless green ideas sleep furiously.
2. Furiously sleep ideas green colorless.

It is fair to assume that neither sentence (1) nor (2) (nor indeed any part of these sentences) had ever occurred in an English discourse. Hence, in

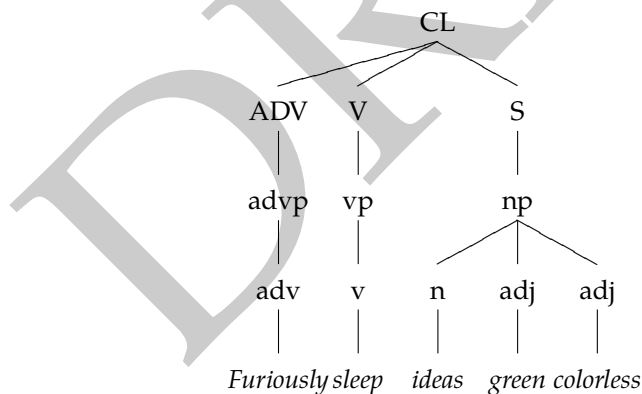
³Saussure 1998.

any statistical model for grammaticality, these sentences will be ruled out on identical grounds as equally “remote” from English. Yet (1), though nonsensical, is grammatical, while (2) is not.⁴

How can it be that (1) is correct and (2) is not? Both are “meaningless”! It is true that the *semantics* of the sentences do not make sense by putting these words together. Nevertheless, (1) “feels” right and (2) “feels” wrong. The syntax trees of these two sentences will graphically explain why.



Tree 1.5: A grammatically correct English sentence



Tree 1.6: A grammatically incorrect English sentence. Do you know why?

How are these two trees different? The word order is different. (1) is S + V + ADV while (2) is ADV + V + S. Word order (1) is more frequent than (2), but (2) is

⁴Chomsky 1957, pp. 15–6.

quite common in English. The order of the two adjectives and noun are reversed just like the clause order. In (1) the adjectives come first, and the noun comes first in (2). This is where the grammatical error occurs. In English adjectives modifying a noun never follow it (unlike in Hebrew!). Thus *ideas green colorless* is just plain wrong. That is what we sense from the second sentence.

Syntax and semantics. They are different.

1.1.2 Semantics: Clauses Have It, too

Mapping syntax between languages

English and Hungarian verb "help".

The choices translators make

[Find a Hebrew Bible verse where translators have made various choices, depending upon how they interpret the Hebrew syntax (Isa 40:3). Then use an example where translators use different choices in English to render the same Hebrew clause which is itself clear and unremarkable.]

1.2 The Anatomy of Clauses

There are many competing ways that linguists have analyzed sentences. Since this is a book about Hebrew syntax, and specifically the system of analysis used in the *Westminster Hebrew Syntax Database*, we will limit our discussion to those features that relate to Hebrew and not necessarily to all languages in general. The system of analysis used in the database is deliberately as syntax "theory neutral" as possible. Of course, this goal is not absolutely attainable. By "theory neutral" we mean that the labels used and the method of segmentation of the text into "chunks" or syntactic units are only as complex as is needed to show the structure of a clause or sentence. The labels are based as closely as possible to the POS of the head of a phrase or element of a clause. The assumption is that POS is as close to objective as is possible. Inevitably, there will be disagreement. Other choices are possible, and ought to be easily derived from the trees offered in the database. The analysis of clauses is deliberately intended to be minimal rather than maximal. Other, more complex analyses can, we believe, layered on top of the minimal analysis of the database fairly easily. Actual usage by users will prove or disprove this belief.

So far we have looked at only one clause, one thought. But human beings combine more than one thought into an utterance. In writing, sentences can get quite complex. There is an advantage in combining clauses into larger structures. The

clauses can share verbs, subjects or objects without having to tediously repeat the same words over and over again in the small space of three or four thoughts.

1.2.1 Slots and Word Order

What are the parts of a clause? A simple answer would be, all of the phrases found in a clause. But what about the prepositional phrase in Tree 1.3? Inside that phrase another phrase, a noun phrase is embedded. How do we know this phrase can be constructed? Because there is a *pattern* in English where a preposition followed by a noun phrase constitutes a *prepositional phrase*. The np is called the “object” of the preposition and its position (in English) immediately follows the preposition. The preposition is the “head” of the phrase because it and not the np directly relates to the verb of the clause. The relationship of the NP to the PREP is called a “slot”.⁵ Just as phrases have patterns or slots, so do clauses. Sometimes this is called “word order”. Words of the same class (POS) tend to occur in the same position in the clause. But slots are not always linear or consecutive. There are discontinuous slots, where one morpheme of a slot is separated. In the sentence

George went, he and all his clan.

the subject *George ... he and all his clan* are separated by the verb. Yet the subject is one slot, a discontinuous one in this case.

1.2.2 Simple Clauses

By “simple” we mean “one verb.” This distinction between the simple and complex utterances is important, because it defines the minimal unit that then can be combined into a vast array of differing sentence structures, each having their own purpose in the larger unit of text in which it occurs.

What are the constituents of a CL? In English, the traditional answer is subject (S) and predicate (P). The P includes everything but the S, with the emphasis upon the V. This division works well for English, but less so for Hebrew. And we want a fine grained analysis of each word or morpheme. Going back to Tree 1.4, let’s consider the parts of the CL there. We have PP + S + V + O. Everything revolves around the V. The PP modifies the action of the V temporally. Other PP will modify the action

⁵The term “slot” comes out of the linguistic model known as “Tagmemics.” The “tagmeme” is “the correlation of a grammatical function or slot with a class of mutually substitutable items occurring in that slot. This slot-class correlation has a distribution within the grammatical hierarchy of a language.” (Elson and Pickett 1964, p. 57) We make no attempt to consistently follow Tagmemic theory, but we do find this concept helpful in understanding how languages and especially Hebrew construct phrases, clauses and complex sentences, in particular the embedding of one structure into another.

in other ways, e. g., by manner or location. The **O** completes the action of the **V** in some way, often by being *affected* or *changed* in some way. Typically, the **S** is the agent or "doer" of the action, but not always. When the passive voice is used, as in *The ball was hit by the bat*, the **S** is the entity affected by the action, and the cause of the action is in a **PP**. The roles are reversed in the sentence *The bat hit the ball*.⁶ The parts of a **CL** that seem to be necessary in the analysis of Hebrew are listed here and briefly defined:

Verb (V) A word or phrase that points to an action, process or state. It "governs" the rest of the **CL** in that it determines the minimum number and kind of other parts of the **CL**.

Predicate (P) An arbitrarily chosen label for clauses where there is no verb and one of the phrases is chosen as the "predicate". Verbless clauses do not occur in English, but they do in many other languages including Hebrew and Hungarian.

Subject (S) Normally that phrase which "agrees" with the inflection of the verb or sits in the subject "slot". A **NP** is the usual occupant.

Object O The phrase that "complements" or "receives" the action of the **V**. It is concrete or "objective" and so is usually a **N** or **PRON**.⁷

Prepositional Phrase (PP) A phrase that modifies the action of the **V** in some manner. It consists of the **PREP** as the **HEAD** of the phrase, followed by a **NP**.

Adverb (ADV) A single word functioning in a similar way as a **PP**, but the phrase usually consists of one word, or an **NP** at most.

Conjunction Phrase (cjp) A word or phrase that organizes phrases or clauses in some sort of logical relationship.

1.2.3 Complex Clauses or Sentences

Just as a phrase such as an **NP** can be embedded inside another phrase, e. g., inside a **PP** (see Tree 1.3), so clauses can also be found embedded inside another clause. Clauses can be tied together by **cjs** or even by **PREPS**. Clauses in complex sentences

⁶The "roles" that phrases play with respect to the verb is called "case role grammar" and is a part of the underlying semantics of the clause. At the present time case roles, such as Agent, Patient, Benefactive and others, are not encoded in the database.

⁷Radford 1997, p. 519.

(containing more than one **CL**) are either **INDEPENDENT** or **DEPENDENT**. An **INDEPENDENT** clause may be **COORDINATED** with other **INDEPENDENT** clauses, using conjunctions or simple **JUXTAPOSITION** with no intervening words between the clauses. It is important to note that the definition of the **DEPENDENT** clause excludes the idea of “subordinating” conjunctions like *if ... , then*. Such situations are defined as two **INDEPENDENT COORDINATED** by conjunctions into a logical relationship; in this case the first **CL** is the **CONDITIONAL** and the second is the **CONSEQUENT**. The second is *logically* subordinate to the first **CL**, but not grammatically subordination to the first. The reason is that the second clause plays no grammatical role as a clause constituent of the first clause.

(1.1) subordinated clause example

(1.2) coordinated clause example

(1.3) “if, then” example

(1.4) juxtaposition example

adjectival or relative clauses adverbial clauses inter-clause vs. inter-phrasal or inter-word conjunctions

Dependent Clauses

The definition of a **DEPENDENT CLAUSE** is very specific: a clause that fulfills a grammatical role in the **CL**. A **CL** may function as **S**, **O**, **ADJ** modifying a **N**, **APREPOSITIONAL OBJECT** or as an **ADV** modifying the **V**. The type of **v** used --- inflected or otherwise --- is a language-specific issue. When one wishes to use a **CL** for an **ADJ**, often one uses a **RELATIVE PRONOUN**, in English *who, whom, that, which*. For **S** as a **CL**, a gerund often occurs in English. A **CL** used as a verbal object is usually called an **OBJECT COMPLEMENT** or **SMALL CLAUSE**. **ADVERBIAL CLAUSES** in English use a **PREP** and an infinitive. Hebrew uses this same method.

Subject *Going to the store* got him into trouble.

Adjective He is the one *whom you want to speak to*.

Object He won't accept *you working so late*.

Prepositional Object I want you inside *before that storm hits us*.

Adverb I came *in order to meet you*.

Here is where the concept of *slot* is helpful. As the italicized clauses show, the dependent clause fits into the same position that any nominal subject, object, adjective or adverb would. Hebrew does the exact same thing as English, but using Hebrew clause slots.

DRAFT



2 What is Hebrew Syntax?

We must remember that ancient Hebrew is only a sample of the larger language and its literature. It is a literary language and so its grammar and syntax will reflect that. The Bible is not the only remnant of ancient Hebrew, but is the only significant example of it. Since its content is religious, the words, expressions and literary genres are naturally closely associated with Israel's religious beliefs. As a result, we cannot compare the syntax of the Hebrew Bible to its larger linguistic environment. That context is lost to us. The best we can do is observe and describe what has survived. If we only have one or a few examples of a syntactic structure, we are limited in how much we can generalize from those examples. In other cases, we have a rich and numerous set of examples that allow us to hypothesize their motivation, predict their occurrence and enumerate their variations.

2.1 How to Use Syntax Information

2.2 Simple Clauses

A clause is a complete thought. Hebrew expresses complete thoughts in a number of different ways. Here I describe the forms. Questions of *when* a particular form is used and *why* that form was chosen is more of a matter for **TEXT LINGUISTICS**, which studies the **MOTIVATION** of the linguistic and especially syntactic choices a language

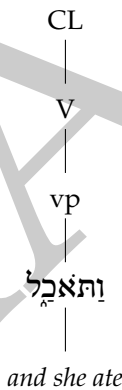
user makes. A major omission in most Hebrew studies of syntax is this business of *motivation*.

How do we classify Hebrew clauses? Classic Hebrew grammar since the Middle Ages [citation needed] begin with two types of clauses: those that begin with a verb and those that begin with a noun. This has been modified to be clauses which have a verb in any position of the clause and those lacking any verb: verbal and nominal (or verbless) clauses.[cite Waltke&O'Connor]

A preliminary examination of the text shows considerable variation in formal patterns as the type of verb or predicate varies. Examples of simple clauses classified according to type of verb or predicate follow.

2.2.1 Finite Verbal Clauses

Hebrew finite verbs are inflected for subject and so can stand alone as a one word clause.



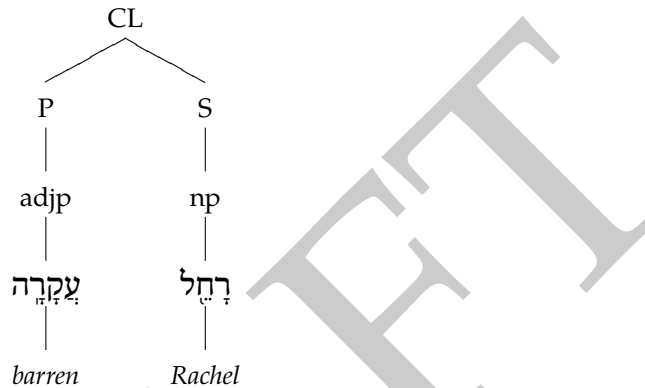
Tree 2.1: Simple finite verbal clause
Genesis 3:6

The tree is flat and vertical, being a single *vp* that is “promoted” to the *V* and is the only element in the *CL*. The verb is of the *wayyiqtol* conjugation, earlier known as the “waw-consecutive.”¹In Hebrew, an explicit *S* is not necessary and is, in itself, a complete thought or assertion.

¹Technically speaking, the *-וַ* should not be a part of this tree, since the *cj* coordinates this *CL* with another at the same level. The database actually splits the *cj* off into a separate tree. There is also some question of whether the *-וַ* plus dagesh is a *cj* at all, but is rather a morpheme belonging to this verbal conjugation. The database takes the position that it is indeed a *cj*, and future *wayyiqtol* trees will omit the *waw*.

2.2.2 Nominal Clauses

Many languages have CLs that do not have verbs. In Hebrew, this class of clause is called “nominal” or “verbless.” The example below has the simplest structure. A minimum of two constituents are necessary, an **S** and a **P**.²



Tree 2.2: Simple nominal clause
Genesis 29:31

For English speakers, who have no equivalent of the nominal clause, must add (in their heads) a “copula,” that is, the verb *to be* to sit in between the **S** and the **P**. Thus we translate the example above as *Rachel [was] barren*. But it is important to remember that there is no “understood” verb *to be*. For native speakers, the way they think about this construction is *Rachel barren*.³

2.2.3 Infinitival Clauses

2.2.4 Participial Clauses

2.2.5 Vocative Clauses

2.3 Complex Clauses

2.3.1 Prepositional Clauses

Where clauses are objects of prepositions.

²P stands for “Predicate,” meaning roughly “what is asserted.” It is used to distinguish between those assertions that are verbal and those that are some other **POS**

³As an example from a very different language, modern Hungarian has an exact equivalent to the Hebrew nominal clause and so is able to translate word-for-word. A simple example is *ő beteg* “he sick.” It does not require the verb *to be* as far as an Hungarian is concerned.

2.3.2 Relative Clauses

That is, it does not “stand for” a noun. It sits in a “slot” for adjectives and “stands for” a clause. Its function in the main clause is usually attributive, that is, a further specification of the preceding noun phrase.

It is a “dummy” nominal

What do you do when you want to use a whole clause in the place of the sentence where a noun or an adjective is expected or required? You use **וְשֵׁנִי** followed by the desired clause.

וְשֵׁנִי is not a part of the clause that follows it.

Its role is in the “upper” clause. It is not the subject or object or conjunction or anything else.

2.3.3 Direct Speech

2.3.4 How the conditional **כִּי** combines clauses

There are two parts: the condition and the consequent. How do we know how many clauses belong to the first or to the second? How do we know if the consequent is omitted? The expression **כִּי לֹא** omits the consequent when no words follow it and either the verse ends or a new clause begins.

Appendices

DRAFT

DRAFT

**A**

Labels and Principles of Analysis

Here we document the labels used in tagging Hebrew clauses, the principles used in analysis and the justifications for them.

Moving from the bottom upwards, the trees begin at a terminal node which is the Hebrew morpheme itself. The next node upwards is the **POS** node. Next is the single morpheme phrase level.¹

A.1 Parts of Speech

We begin analysis with the simplest set of labels to describe each **MORPHEME**. A morpheme can be an entire word, such as מֶלֶךְ *king*. Or it can be just a part of a word, such as a **PREP**, e. g., בְּ *in* or the **CJ** וְ *and*. We do not label inflections of the verb or noun except for pronominal suffixes such as אַתָּה *you, your*. The list of possible **POS** are taken from *The Westminster Hebrew Morphology*.

¹The definition of, for example, an **NP** is “one or more nouns”. With this rule in the parser, the result is that nearly every terminal node is promoted to a phrase, and then is combined with other morphemes to form more complex phrases. Some have commented that this creates quite a bit of redundancy, since it is clear from the **POS** what the element is. This may be so — for humans. But since it’s obvious that there are phrases which consist of only one morpheme, there must be a rule **NOUN** → **NP**. The nature of **CYK** bottom-up parsers is such that this rule is going to be applied *everywhere*. For the purposes of display, this “extra” level of “redundant” analysis could be omitted. The desirability of the omission is questionable.

POS	Abbreviation
article	ART
conjunction	CJ
adverb	ADV
interjection	IJ
object marker	OM
preposition	PREP
relative pronoun	REL
independent pronoun	PRON
noun	NOUN
adjective	ADJ
verb	VERB

A.2 Phrase-level Nodes

The creation of phrases and their labels uses the convention of lower case letters. For the most part, **POSs** have a corresponding phrase label. The label is given to the **HEAD** of the phrase, corresponding to the head's **POS**.

Phrase type	Abbreviation
article	none
conjunction	CJP
adverb	ADVP
interjection	IJP
object marker	OMP
preposition	PP
relative pronoun	RELP
independent pronoun	NP
noun	NP
adjective	ADJP
verb	VP

The **ART** never occurs in a phrase by itself, so it does not have a corresponding phrase type. The **PRON** is considered a **NOUN** since the rules for **PRON** phrases are the same as for **NOUN** phrases. **PRON** subtypes such as the suffixed pronouns are distinguished in the **NODE** element's attributes.

A.3 Clause-level Nodes

By convention, a visual distinction is made between clause- and phrase-level node labels. Clause-level labels are in all capitals.

CL constituent	Abbreviation
clause	CL
subject	S
(direct) object	O
adverb	ADV
prepositional phrase	PP
object marker	O
second object	O2
object complement	OC
indirect object	IO
verb	V
predicate	P

The purpose of the Hebrew syntax database is to provide a tool for research into higher levels of Hebrew grammar, that is, the **TEXT LINGUISTICS** of biblical Hebrew. In theory, the highest level of language is the **CL**. In practice, this is not entirely possible. **CLs** begin combining at once and the free standing, single **CL** is not the only class of **CL**. Embedding and subordination create complex sentences. However, the grouping of **CL** and complex sentences into larger structures is not marked in this database. Although one of the basic tenets of the theory of language embraced by the authors of this database is that syntactic variation is primarily motivated and explained by text-level goals of the speaker, methodologically it is one step too early for such analysis. The present tool will provide the means by which we can continue to “bootstrap” our way up the levels of language.

A.4 Database XML Format

The database format is a valid **XML** document. This format was chosen not only for its utility for an archival and data exchange purposes, but also because syntactic structures are inherently hierarchical. Here is the essential structure:

```
<?xml version="1.0" encoding="utf-8" ?>
<Sentences>
  <Sentence ID="gn1:1">
    <Trees>
      <Tree>
        <Node attribute="..." ...>
          <Node attribute="..." ...>
            <Node attribute="..." ...>
              ...
            </Node>
          <Node attribute="..." ...>

```

```

    <Node attribute="..." ...>
      <Node attribute="..." ...>
        ...
      </Node>
    ...
  </Node>
</Node>
</Tree>
<Tree>
  ...
</Tree>
</Trees>
<Sentence ID="gn1:2">
  ...
</Sentence>
...
</Sentences>

```

The **<NODE>** is where all the primary information resides. The parent elements (**SENTENCES**, **SENTENCE**, **TREES**, **TREE**) are simple containers. How the **<NODE>** nests with parent and child **<NODE>** elements represents the syntax tree structure. The attributes associated with each **<NODE>** contains additional information about the syntax and structures.

A.5 Attribute List

This list does not include all the attributes used in the database. Rather they represent those data which will be of interest to the user. Other attributes are used by the parser to help make decisions on which rules apply to a given case.

Name	Abbreviation	Description
Apply Rule	AR	A parser directive: apart from any other applicable rule, apply the rules specified in the dictionary for this morpheme at this place.
Analysis	Analysis	The morphology parse string from the <i>Westminster Hebrew Morphology database</i> .
Block Rule	BR	A parser directive: the rules specified in this attribute for this node shall not be applied at all.
Category	Cat	Essentially, the POS.
End	End	The numeric position of where the phrase ends.
Head	Head	In a phrase, morphemes are numbered from 0 to n. This attribute identifies which morpheme is the head of phrase. Usually the first morpheme of a phrase is the head, so the most common value is "0".

Continued on next page

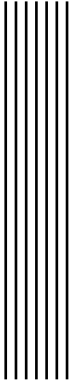
Name	Abbreviation	Description
Identification	Id	
Left Edge	LE	The beginning of the phrase according to masoretic accenting.
Lemma	Lemma	The “dictionary” form of the morpheme, according to the <i>Westminster Hebrew Morphology database</i> , in Michigan encoding.
MichiganCode	MichiganCode	The “surface text” of the Hebrew morpheme in the ASCII transliteration scheme known as the “Michigan encoding”, as it actually occurs in the text including all affixes, suffixes and inflections.
Right Edge	RE	The end of the phrase according to masoretic accenting.
Rule	Rule	The parser grammar has a string label for each rule that is applied to a <NODE>. The string enclosed within brackets, e. g., [NP2S].
Start	Start	The numeric position of where the phrase begins.
StrongNumber	StrongNumber	The number of the lemma according to the (modified) system first created by James Strong. ^a
Clause Type	Type	The kind of predicate occurring in a CL . ^b
Unicode	Unicode	The surface text in UTF-8 encoding.
UnicodeLemma	UnicodeLemma	The lemma in UTF-8 encoding.
Cantillation Error	cantErr	Whether the morpheme violates the masoretic phrase boundary. ^c
Morpheme ID	morphId	A string of numbers representing the book, chapter, verse, word-in-verse, and part-of-word.

Continued on next page

^aGet details on which modifications of Strong’s numbers are used in the database.

^bThe possible values are **NOMINAL**, **VERBAL** (that is, a finite verb), **INFINITIVE** and **PARTICIPIAL** (when the use of the participle is as a verb). This information comes from the morphology and is useful in creating parser rules.

^cThe masoretic accents have two sets. One set are “conjunctive” accents, meaning the word is to be understood with the following word. The other set are “disjunctive,” indicating there is a phrase boundary at this accented word. These accents were created for the chanting of the biblical text in liturgical contexts and so are influenced by phonological criteria as well as by syntactic considerations. Nevertheless, we have found the accents to be a very high predictor of phrase boundaries. This attribute notes if the masoretic “phrase” boundary has been crossed. The parser takes this information into account in the grammar rules.



Glossary

case role *The way in which a clause constituent functions with reference to the verb.*

clause *A complete unit of thought with one verb or, in the case of verbless clauses, with a predicate.*

dependent clause *A clause that has a grammatical role to play with respect to an independent main clause.*

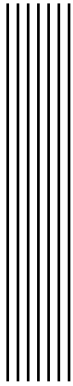
head of phrase *The word that governs the characteristics of the other elements of the phrase and which “connects” the phrase to the verb. The **HEAD**’s part of speech determines what kind of relation to the verb the phrase will have. Thus the **NOUN** is the head of **ADJECTIVES** in a noun phrase. Similarly, the **PREPOSITION** is the head of the **PREPOSITIONAL PHRASE***

independent clause *A “main” clause that can stand by itself in the broader flow of text.*

juxtaposition *The placing of two linguistic constructs next to each other with no intervening word or morpheme.*

sentence *A unit of thought containing one or more verbs or predicates.*

DRAFT



Bibliography

- Chomsky, Noam (1957). *Syntactic Structures*. Janua Linguarum, series minor 4. The Hague: Mouton. ISBN: 9027933855.
- Elson, Benjamin F. and Velma B. Pickett (1964). *Introduction to Morphology and Syntax*. Dallas, TX: Summer Institute of Linguistics.
- Radford, Andrew (1997). *Syntactic Theory and the Structure of English: A Minimalist Approach*. Cambridge Textbooks in Linguistics. Cambridge University Press. ISBN: 0521477077.
- Saussure, Ferdinand de (1998). *Course in General Linguistics*. Chicago, IL: Open Court. ISBN: 0812690230.