Cambridge Semitic Languages and Cultures

The Intertwined World of the Oral and Written Transmission of Sacred Traditions in the Middle East

Edited by Alba Fedeli, Geoffrey Khan and Johan Lundberg

UNIVERSITY OF CAMBRIDGE
Faculty of Asian and Middle Eastern Studies

https://www.openbookpublishers.com

Semitic Languages and Cultures 40

ISSN (print): 2632-6906
ISSN (digital): 2632-6914

ISBN Paperback: 978-1-80511-740-7
ISBN Hardback: 978-1-80511-741-4
ISBN Digital (PDF): 978-1-80511-742-1

DOI: 10.11647/OBP.0498

Cover images: Karaite Arabic transcription of the Hebrew Bible into Arabic letters: British Library Or 2540, fol. 12r, Exodus 4.15–4.19. Hebrew Bible: Aleppo Codex, fol. 122r, Isaiah 44.16–45.11 (Courtesy of the Ben-Zvi Institute, Jerusalem. Photographer: Ardon Bar Hama). Arabic Qurʾān: Gotha Research Library of the University of Erfurt, Ms. orient. A 432, fol. 1r, Public Domain Mark 1.0, Q.3:51–3:52. Syriac Gospel: Staatsbibliothek zu Berlin, Phillips 1388, fol 30v, Matt. 12.33–41.
Cover design: Jeevanjot Kaur Nagpal

The fonts used in this volume are Charis SIL, SBL Hebrew, SBL Greek, Scheherazade New, Estrangelo Edessa, Serto Jerusalem and Allan.

MAIN CLAUSE VERBS ARE PROSODICALLY WEAKER THAN NOUNS IN THE TIBERIAN CANTILLATION OF BIBLICAL HEBREW PROSE BOOKS

Robert S. D. Crellin

1.0. Foundations

1.1. Word Classes

Morphemes in the world's languages are typically separated into classes, including verbs, nouns, adjectives, adverbs, prepositions and (e.g., discourse) particles. These classes are known as 'word classes' (and 'parts of speech', 'grammatical categories', or 'lexical categories', see Payne 1997, 32–70). Depending on the language, these categories can overlap: thus in Homeric Greek, for instance, there is considerable overlap between the category of preposition and that of adverb (cf. Haug 2009). The present study is concerned with the relationship between a morpheme's prosodic status and its word class.

 https://doi.org/10.11647/OBP.0498.05

1.2. Functional vs Lexical Classes

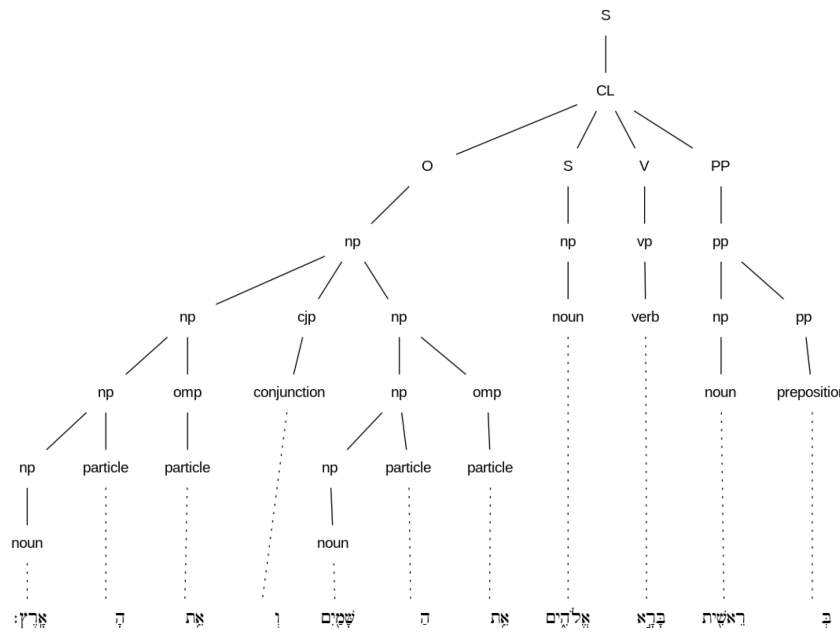
Word classes may themselves be categorised into 'functional' (or 'grammatical') and 'content' (or 'lexical') classes (for the distinction see Sapir 1921; Fries 1952). Distinguishing function words from lexical (content) words is the fact that they relate entities in the discourse to one another, whether that be lexical words, clauses or sentences etc. By contrast, content words refer to entities outside the discourse (cf. Crellin 2022a, 11–12). For instance, the preposition 'on', as in 'The cat is on the sofa', relates 'the cat' to 'the sofa', which are named participants in the discourse. By contrast, 'cat' and 'sofa' give labels in the discourse to entities outside it.

Functional word classes are typically closed (Gelderen 2004). This is to say that new members cannot readily be added (Caink 2008, 491). By contrast, lexical word classes are open, meaning that new members may be readily generated by the community of language users.

1.3. Syntactic Structure: Phrase Structure Grammar

Words may be combined to form larger units. It has been a convention for several decades to analyse the surface structure of a sentence as a tree, where each node dominates a contiguous sequence of words. Figure 1 provides a phrase structure representation of Gen. 1:1, as analysed in the Macula treebank.

Figure 1: Phrase structure representation of Gen. 1.1 (Macula)



The central insight in phrase structure representations of syntax is that a constituent—viz. a string dominated by a single node in the tree—may be exchanged with another string with the same class as the first string. In Figure 1, the sentence is analysed into prepositional phrase (PP), verb (V), subject (S) and object (O) phrases. By far the longest of these is O: אֵת הַשָּׁמַיִם וְאֵת הָאָרֶץ׃. In principle this string could be exchanged for an alternative object constituent consisting of a single word or morpheme, e.g., a pronoun.

The tree is hierarchical: in בְּרֵאשִׁית the node for the phrase as a whole (PP) stands above the nodes for each of the constituents (P and NP) respectively.

1.4. Prosodic Structure

The utterance is structured in phrases in a way parallel to the syntax of a sentence. Various levels have been identified (for a helpful overview see Goldstein 2016, 44–46). For our purposes here, it is sufficient to identify three levels: (1) the utterance itself; (2) the prosodic phrase; (3) the prosodic word.

1.4.1. Utterance

This is the highest level of the prosodic hierarchy. Nespor and Vogel define it as follows (2007, 221):

> The last phonological constituent we will consider is the phonological utterance (*U*), the largest constituent in the prosodic hierarchy. A *U* consists of one or more intonational phrases, the category just below it in the hierarchy, and usually extends the length of the string dominated by the highest node of a syntactic tree, which we will refer to as X^n.

They are at pains to point out (2007, 221),

> This is not to say, however, that *U* is simply the phonological counterpart of X^n, a phonological constituent introduced only to avoid making direct reference to syntax in the formulation of phonological rules. In fact, X^n and *U* are not always the same, as will be demonstrated on the basis of several phonological rules that operate within the U domain, but not within the confines of X^n.

For the purpose of the analysis of the Tiberian cantillation tradition of Biblical Hebrew, I take the utterance to correspond to the *verse*, since this is the highest level of the phonological

hierarchy. One consequence of this decision is that an utterance will potentially contain more than one sentence.

1.4.2. Prosodic Word

The prosodic word is identified cross-linguistically as the domain of the single main stress or word accent (cf. Truckenbrodt 1999, 220; Crellin 2022a, 12–16 for application in the context of North-west Semitic, including Biblical Hebrew, and Greek; cf. Fortson 2008, 261 and Crellin 2022b for Latin; see also references in the cited works).

1.4.3. Prosodic Phrase

Prosodic phrases are identified *inter alia* by pauses of varying lengths existing between sequences of prosodic words as they are pronounced (Dresher 1994, 23–25). Prosodic phrases can be nested in a hierarchy. The longer the pause relative to the other pauses in the sentence, the greater the strength of the division, and the higher in the hierarchy the position of the two phrases adjacent to the pause.

1.4.4. Phonological Clisis

Not all morphemes are able to stand as prosodic words in their own right. This arises because the morpheme in question is too short in prosodic terms (where prosodic length may be construed either in terms of syllables or morae, depending on the language, and the prosodic word is minimally binary either in syllables or morae). Such morphemes are incorporated into the prosodic

structure of the utterance by being incorporated into a neighbouring prosodic word (the host). Such morphemes are accentually dependent on their host.

There are at least three kinds of phonological clitics: 'free clitics', 'internal clitics', and 'affixal clitics' (for details see Anderson 2005, 46; Goldstein 2016, 48). These are represented in Figure 2, Figure 3 and Figure 4 respectively.

Figure 2: Free clitic

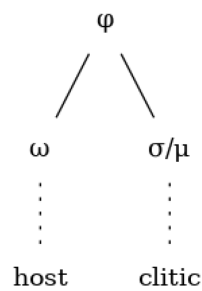


Figure 3: Internal clitic

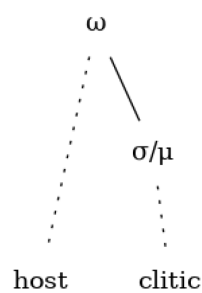
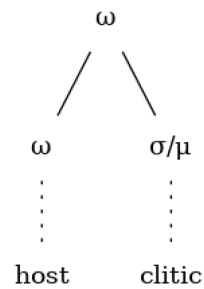


Figure 4: Affixal clitic



What these all have in common is that the clitic is not a prosodic word in its own right. The prepositions *b-*, *k-* and *l-* are always clitics in Biblical Hebrew, since, standing at not more than one mora in length, they always fall short of the two-mora threshold for prosodic wordhood.

1.5. Relationship Between Word Class and Prosodic Status

A relationship has been observed to exist between the functional status of a morpheme and its prosodic strength: function words are more likely to be prosodically 'weak' or 'deficient' than content words (see Gelderen 2004, Evertz 2018, 140, and references therein). This is to say, function words are more likely to be incorporated into the prosodic structure of neighbouring morphemes at the level of the prosodic word.

Content words are held to be incapable of prosodic deficiency, unless they are members of a special grammaticalised subclass, e.g., auxiliary verbs. Caink is explicit on this point (2008, 491):

> Items that are clitic vary across languages but are always grammatical (or functional) words and thus members of closed classes in that they cannot be coined (Emonds, 1985, Chap. 4); they may include auxiliary verbs, pronouns (as in many Indo-European languages), question particles (Slavic *-li* in (9) or Finnish *-ko* in Nevis, 1988, 9), negative particles (Slavic verbal negation ne -in (9)), and conjunctions (Latin *-que* 'and'). There are no clitic forms for open-class items such as the lexical noun *wood*, despite the homophony with the modal auxiliary *would* (cf. I'd really like that).

While clitics must be function words, according to this view, functional status is not enough to guarantee prosodic deficiency: some function words—such as 'too' and 'off' in English, and 'weil' and 'aus' in German—carry their own word accent despite their functional status (see Evertz 2018, 140, with references).

1.6. Prosody and Syntax Above the Prosodic Word

Just as prosody interacts with morphosyntax at the level of the word, so it does at the level of syntax. This relationship is not, however, isomorphic (Truckenbrodt 1999, 220). It is lexical rather than functional items that are viewed as critical for the alignment of prosodic structure with syntactic. Truckenbrodt identifies a 'lexical category condition' under which "[c]onstraints relating syntactic and prosodic categories apply to lexical syntactic elements and their projections, but not to functional elements and their projections, or to empty syntactic elements and their projections" (1999, 226).

Selkirk is categorical (1996, 191):

> *[T]he set of constraints governing the interface between morphosyntactic and prosodic structure makes no reference to functional categories at all.* Rather, it is only lexical categories and their phrasal projections which would figure in the statement of morphosyntactic constraints on prosodic structure; GCat would stand only for "LexCat" in any constraint of the Align (GCat PCat variety).

The phenomenon under which the phrasal projections of lexical categories align with prosodic phrase boundaries is known as 'edge alignment' (for further explanation and references see Truckenbrodt 2007, 437–438).

2.0. Accessing Syntactic and Phonological Structure in Biblical Hebrew: The TanakhML Treebank

In the preceding section we gave an overview of some foundational concepts in syntax and phonological structure. In the present section I consider the means of accessing syntax and phonological structure in Biblical Hebrew.

Very unusually for a document, or set of documents, with as deep a history as the Hebrew Bible, not only have we received the 'text', that is, the characters corresponding (in the case of Hebrew) to the consonant (and some vowel) phonemes of the spoken language, but also more than one system of 'accents'. The present study is concerned with the Tiberian system, as recorded and described by the medieval community of scholars known as the Masoretes.

These accents were designed to record the system of chanting the Hebrew Bible, recording both the position of the primary

accent on prosodic words, and the musical contour of the chant (Yeivin 1980, §178; Aronoff 1985, 33; Dresher 1994, 5–6; Khan 2020, 51).

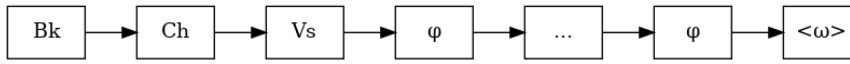
The accents' connection with the reading tradition or practice of chanting the Hebrew Bible is not disputed. Also not disputed is that the accents have a purpose beyond the marking of the contours of the reading tradition, and that this further purpose has to do with the sense of the text. This is suggested, not least, by the meaning of the Hebrew term טעמים 'senses' (Aronoff 1985, 35; Janis 1987, 4; Dresher 1994, 6). Where there is not agreement is on the linguistic domain of the units demarcated by the accents. It has been proposed that the accents demarcate 'semantic' (Yeivin 1980, §178), syntactic (Aronoff 1985) and prosodic (Dresher 1994) units. For the purposes of the present study, I follow Dresher in taking the accents as denoting prosodic structure.[1]

2.1. Structure Overview

The prosodic structure of Biblical Hebrew for this investigation is provided by the TanakhML treebank (https://www.tanakhml.org/). The structure of the tree is a direct representation of the prosodic structure of the Tiberian accents. As with any tree, a TanakhML tree comprises two fundamental types of node: terminal (= leaves) and non-terminal. Figure 5 gives the hierarchy of dominance in a TanakhML tree.

[1] Space does not permit me to go into the details of why this analysis is preferable. Relevant issues are discussed in Crellin (2022a).

Figure 5: Hierarchy of dominance in a TanakhML tree



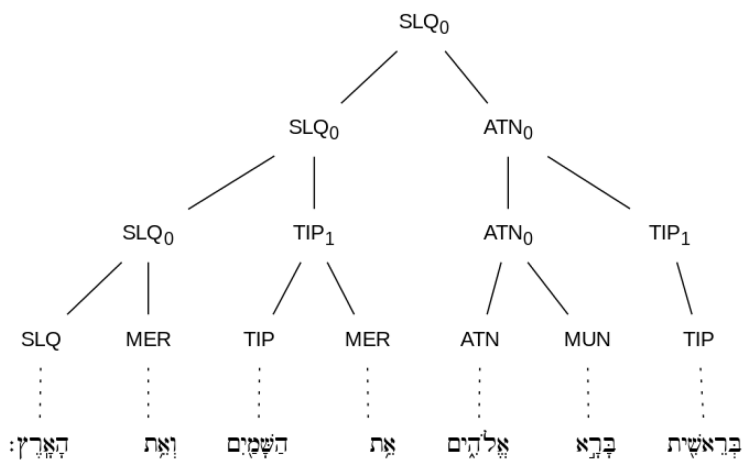
2.2. Terminal Nodes

Terminal nodes correspond to graphematic words (<ω> in Figure 5). Graphematic words are marked up for their accent. The accent may be of one of the following three types: (a) *Maqqef*; (b) Conjunctive; and (c) Disjunctive. Disjunctives are always phrase-final, whereas conjunctives and *maqqef* are never phrase-final.

In Figure 6, terminal nodes are labelled with their accents:

tifḥa, disjunctive (TIP)
munaḥ, conjunctive (MUN)
atnaḥ, disjunctive (ATN)
merka, conjunctive (MER)
tifḥa, disjunctive (TIP)
merka, conjunctive (MER)
silluq, disjunctive (SLQ)

Figure 6: Representation of Gen. 1.1 (TanakhML)



2.3. Non-terminal Nodes

Non-terminal nodes in TanakhML are phrase-level nodes (φ in Figure 5). Non-terminal nodes are labelled for the final disjunctive descendant. The subscript number indicates the position of the accent within the accent hierarchy, to which I now turn.

2.4. Prosodic Strength and the Hierarchy of Accents

Disjunctive accents in Tiberian Hebrew may be divided into four categories according to the strength of the prosodic division which follows (Dresher (1994, 4; Park 2020, 14) ranging from D0 (the strongest) to D3 (the weakest).[2] In purely numerical terms, therefore, the values range from 0 to 3, in reverse order of disjunctive strength.

In this study the conjunctive accents, which indicate that the following item belongs to the same prosodic phrase, are accorded a reverse strength score of 4. In turn, *maqqef*, which indicates that the following item belongs to the same prosodic word, is accorded a reverse strength score of 5.[3]

Since it is more intuitive for prosodic strength to increase in line with numerical value, for this study the prosodic strength of an accent is calculated by subtracting the reverse prosodic strength from 5. In these terms the weakest accent (*maqqef*) obtains a prosodic strength score of 0, while *silluq* and *atnaḥ* obtain

[2] For a recent overview of the accent hierarchy in Tiberian Hebrew, see Park (2020, 9–22).

[3] This is the practice of the TanakhML treebank, where words carrying conjunctive and *maqqef* in TanakhML are given reverse strength scores of 4 and 5 respectively.

a score of 5. In these terms the weakest disjunctive accent has strength 2.

2.5. Contextual Strength of the Disjunctive Accents

Whilst in principle the higher the position of a disjunctive accent is in the hierarchy, the stronger the prosodic pause associated with it, this is not always the case in practice. Specifically, when a series of disjunctives of the same hierarchical level appear in sequence—with or without intervening conjunctives—the strength of the break decreases successively. Park puts the matter as follows (2020, 18):

> [T]he first D_{i+1} level accent presents the greatest relative pause, and the following D_{i+1} level accents indicate progressively lesser pauses. This case presents a unique situation in that the last D_{i+1} level accent appears just before the next level domain (i.e., the D_i level) even though it itself carries the weakest pause.

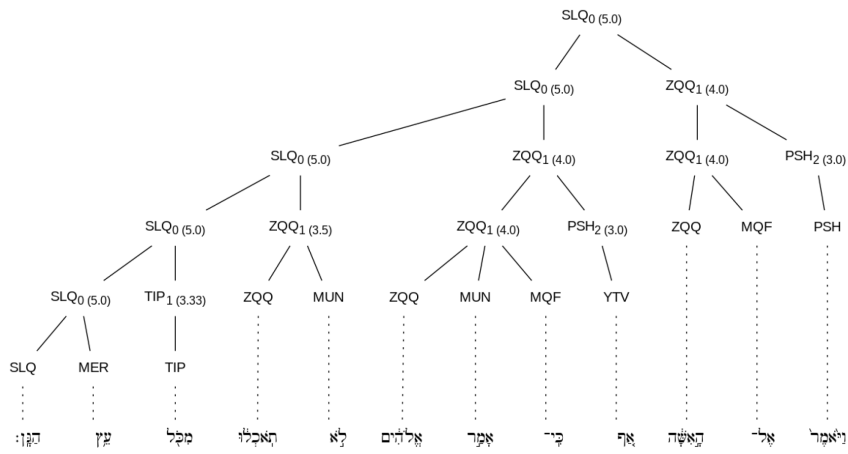
In the present study, the strength of the *nth* disjunctive in a series of disjunctives of level *i*, without interruption by a disjunctive of a different level, is given by the following formula:

$$s' = s - 1 + 1 / n$$

where s' is the modified disjunct strength, s is disjunct strength and n is the number of consecutive disjunctive accents at level *i*.

The formula ensures that the contextual disjunctive strength will always be between the nominal disjunctive strength and the disjunctive strength of the next weakest accent class, but never weaker than the latter.

Figure 7: Representation of Gen. 3.1b in TanakhML



To illustrate how this works, an analysis of Gen. 3.1b is given in Figure 7.[4] The phrase לֹא תֹאכְלוּ terminates in a disjunctive, *little zaqeph*. This is at place D1 in the hierarchy, and is therefore initially accorded disjunct strength 5 − 1 = 4 on my scale. However, it follows immediately on from the phrase אַף כִּי־אָמַר אֱלֹהִים, which also carries *little zaqeph*. Since לֹא תֹאכְלוּ contains no additional disjunctives, we have a case of two disjunctive phrases at the same level in the hierarchy following on from one another. The second of these, לֹא תֹאכְלוּ, therefore carries a lower disjunctive strength than would be expected based on its nominal position in the hierarchy. Specifically its disjunctive strength becomes 4 − 1 + ½ = 3.5.

After this follows the single-word phrase מִכֹּל, carrying the disjunctive accent *tifḥa*. Like *little zaqeph*, *tifḥa* is at place D1 in

[4] The suffixes a, b, c, etc. after verse references distinguish main clauses in the verse.

the hierarchy. Here again then we have a sequence of D1 accents. Since מִכֹּל is the third in the sequence, its adjusted disjunctive strength becomes $4 - 1 + ⅓ \approx 3.33$.

3.0. Analysing Syntax in Biblical Hebrew

3.1. Introduction

If (as we will take to be the case henceforth) the Tiberian Biblical Hebrew accents provide an analysis of prosodic structure rather than syntactic structure, it is important to address briefly the question of how the syntax of Biblical Hebrew is accessed and analysed.

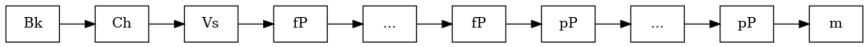
Syntax is a more abstract notion than prosody: prosody can be measured empirically in terms of length of pause, pitch contours etc. Syntax, by contrast, is an abstraction from the surface order of the text into a structure that exposes directly the relationship between the elements. This structure is not provided by (instances of) the language itself, but must be derived by the application of a syntactic theory. The resulting syntactic structure is therefore dependent on the syntactic theory one applies. The syntactic analyses used in the present study are those provided by the Macula treebank.

3.2. Structure of the Macula Treebank

3.2.1. Overview

A schematic representation of the dominance hierarchy in a Macula syntax tree is given in Figure 8.

Figure 8: Hierarchy of dominance in a Macula syntax tree



As in the case of the TanakhML treebank, terminal and non-terminal nodes correspond to different functions (see Tan and Wu 2022, 2).

3.2.2. Terminal Nodes

Terminal nodes (corresponding to *m* in Figure 8) are 'word-level' units in some sense: "A terminal node is the basic unit of syntactic analysis, usually corresponding to the type of speech analysis for each *word* as provided by the OSHB morphology" (Tan and Wu 2022, 5). While Tan and Wu (2022) use the term 'word', these units are not graphematic words, as in the TanakhML treebank, but rather 'morphosyntactic words'. Thus in Figure 1 the article הָ and the conjunction וְ occupy separate terminal nodes, despite not being graphematic words. This is, of course, reasonable, since, though these morphemes are never written as separate words in our texts, they do not have the same morphosyntactic function as their graphematic hosts (see further Crellin 2022a).

3.2.3. Non-terminal Nodes: Phrase-level Syntactic Categories

Non-terminal nodes in Macula are of two kinds: clause-level and phrase-level. These are indicated in Figure 8 by the labels *fP* ('f' for 'functional') and *pP* ('p' for 'phrase'), respectively. Tan and Wu (2022, 6) describe the phrase level as follows:

> The phrase level is the intermediate level between word level and clause level. Phrase level nodes are either non-

> terminal nodes that are the immediate parent nodes of the part-of-speech terminal nodes or parent nodes of other phrase level non-terminal nodes that together form multi-word phrases. From the perspective of the clause, single words or combinations of words form phrases, which are the minimal constituents with a specific function at the clause level.

3.2.4. Non-terminal Nodes: Clause-level Syntactic Categories

The purpose of clause-level nodes is apparently to provide information on the grammatical function of its descendant nodes within the sentence. From Figure 1 it may be seen that each clause level node (S, O, V, etc.) has a single daughter phrase level node.[5]

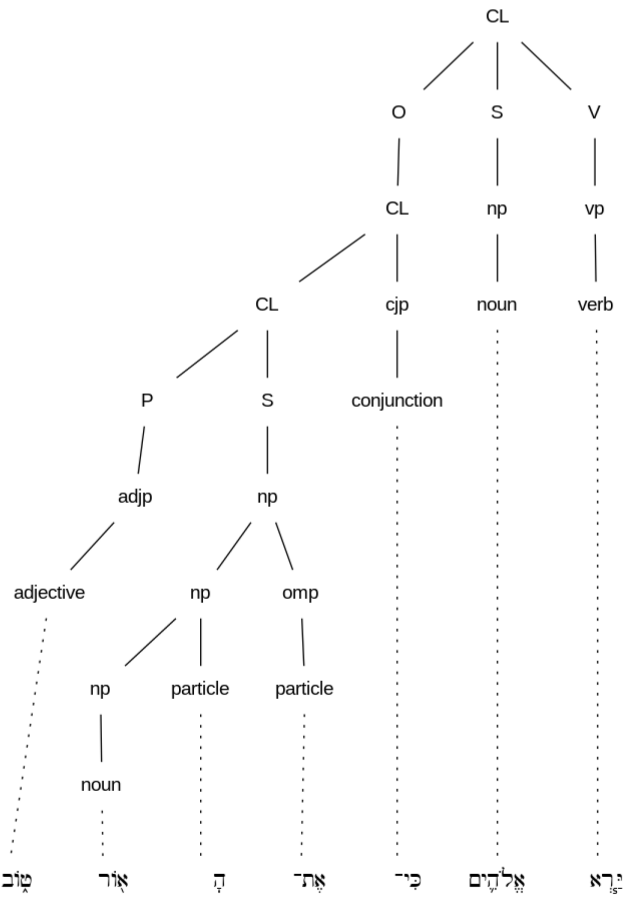
An exception to this is where a clause-level node comprises a clause in its own right, in which case a clause-level node dominates a CL node containing its own clause. In Figure 9 the object (O) constituent of the first clause (CL) immediately dominates a CL node, which in turn immediately dominates another CL node, which in turn immediately dominates a subject (S) and a predicate (P) node.[6]

---

[5] Tan and Wu (2022, 6) describe the role of clause-level nodes as follows: "The clause level differs from the phrase level by using a dependency-like structure. The terminology used to describe the functions of clause level constituents is purposely conservative for ease of understanding and to preserve a clearer link between clause level terminology and phrase and word level terminology."

[6] As explained below, clause-level nodes with nominal categories (S, O and PP) containing subclauses (CL nodes) are excluded from this study.

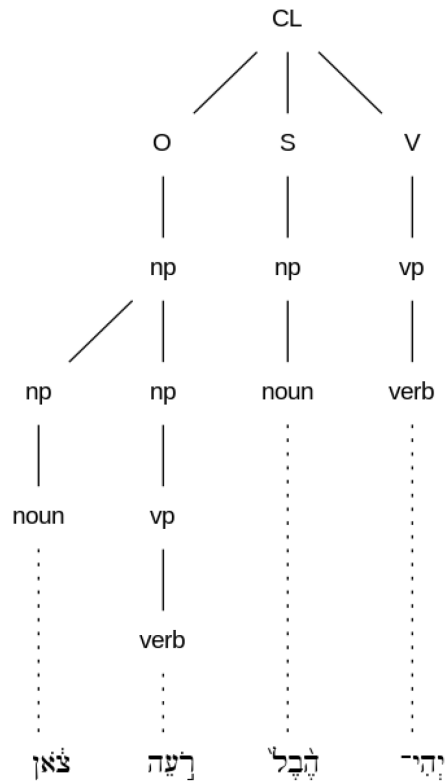
Figure 9: Graphical representation of Macula syntax tree of Gen. 1.4a



3.2.5. Predicative Sentences Involving היה

In sentences containing predications made with היה, the predicative constituent is labelled O in Macula, per Figure 10.

Figure 10: Gen. 4.2b (syntax)



3.2.6. Sentences and Verses

The Macula treebank has an idiosyncratic feature when viewed from a purely syntactic perspective: the root node of a syntactic tree in Macula corresponds to a verse of the Biblical text rather than a sentence: "[T]he Greek trees are sentence-based and frequently cross verse boundaries, whereas the Hebrew trees are verse-based" (Tan and Wu 2022, 2).

Syntactically Gen. 1.1 (Figure 1) comprises a single sentence (= main clause) whereas Gen. 1.4 (Figure 9) comprises two

sentences (= main clauses). The alignment of root nodes with the verse level is shown in the fact that both verses have a single S node as its root: the separate sentences of Gen. 1.4 are headed by separate CL nodes joined by means of conjunctions. In the purely syntactic sense, therefore, sentences correspond to top-level CL nodes.

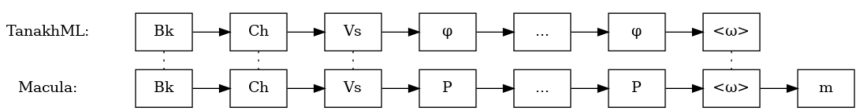
4.0. Integrating Syntactic and Prosodic Treebanks

4.1. Indexing and Minimal Word-level Units

Crucial to integrating two treebanks that were not specifically designed with one another's structure in mind is referencing and indexing. This is to say, it is critical that, when a given word-unit is identified in one treebank, its counterpart in the other treebank can be unambiguously identified.

The Macula and TanakhML treebanks are aligned at the levels of the book, chapter and verse, and at the level of the graphematic word (<ω>) (see Figure 11). The two are not aligned at the levels of the phonological phrase (φ) and the syntactic phrase (P), since the purpose of the treebank is to investigate the relationship between these two.

Figure 11: Alignment of Macula and TanakhML treebanks



Alignment at the level of the verse is rendered straightforward by the fact that the root node of a Macula tree corresponds to its root node corresponding to the same root node in TanakhML, namely, the verse (see above).

4.2. Alignment at the Level of the Graphematic Word

As we have seen, the minimal unit (= terminal node) in a TanakhML tree is the graphematic word, while the minimal unit in a Macula tree is the morphosyntactic word. This is indicated in Figure 11 by the domination by the terminal syntactic phrase node (P) of the morphosyntactic word node (m).

The alignment at the level of the graphematic word is crucial for analysing the relationship between morphosyntax and prosody in Biblical Hebrew. This is because the graphematic word may be said to correspond to the minimal prosodic word (Crellin 2022a). This means that phonological phrase divisions cannot intervene between elements of the graphematic word, even if morphosyntactic divisions do. A prosodic division will, therefore, never occur within a graphematic word, and the minimal unit for the purpose of this study is therefore the graphematic word.

While the graphematic word is a category native to TanakhML, it is not to the Macula treebank. In order to align the treebanks at the level of the graphematic word, therefore, it was necessary to generate this category secondarily in the Macula object model.

4.3. Corpus

The corpus chosen for the present analysis was that of three (Classical) Biblical Hebrew books consisting largely of prose: Genesis, Exodus and 1 Kings. Poetic books were excluded on the grounds

that it is not implausible that the relationship between syntax and prosody might be different in Biblical Hebrew poetry.[7]

4.4. Textual Differences

A significant issue in aligning the treebanks is that the underlying texts of Macula and TanakhML are not identical. For the most part the differences are minor: the vast majority of differences pertain to the order of the unicode encoding of the diacritics in unicode. There are, however, some more significant differences: for example, at Gen. 36.5, TanakhML is lacking the object marker אֶת before יְעוּשׁ.

For the purposes of the present study, if the consonantal texts of a verse are the same in both Macula and TanakhML, the verse's graphematic words are analysed; if the consonantal texts differ, or the analysis is lacking in one or both treebanks, the verse is excluded.[8]

At first sight, the use of the consonantal text as a yardstick of identity might be surprising for a (combination of) treebanks fundamentally concerned with phonology and, therefore, with the vocalised text. However, the use of the consonantal text as a

[7] My thanks to James Cuénod and Ryder Wishart (personal communication, March 2023) for highlighting the importance of considering genre at this point.

[8] In practice, the following verses are excluded on these grounds: Gen. 5.29, 8.18, 14.17, 18.18, 35.22, 36.5, 36.14 and 39.20; Exod. 4.10, 9.22, 10.13, 20.2, 20.3, 20.4, 20.5, 20.6, 20.7, 20.8, 20.9, 20.10, 20.13, 20.15, 21.8, 22.4, 28.1 and 38.12; 1 Kgs 1.27, 4.7, 6.2, 6.3, 6.10, 6.38, 8.11, 8.48, 9.18, 14.2, 16.19, 16.33, 17.15, 19.11, 20.5, 20.29 20.25, 21.29 and 22.49.

yardstick of similarity is justified since, although the purpose of combining the treebanks is to provide the possibility of querying on the syntax-phonology interface, it is not a requirement that the vocalised text with accents be the same in both treebanks, only that it be visible in one of them.

5.0. Prosodic Strength by Word Class in Biblical Hebrew: Quantitative Analysis of the Depth of Prosodic Division at the Level of the Graphematic Word

Figure 12: Gen. 1–50: Depth of prosodic division at the level of the graphematic word

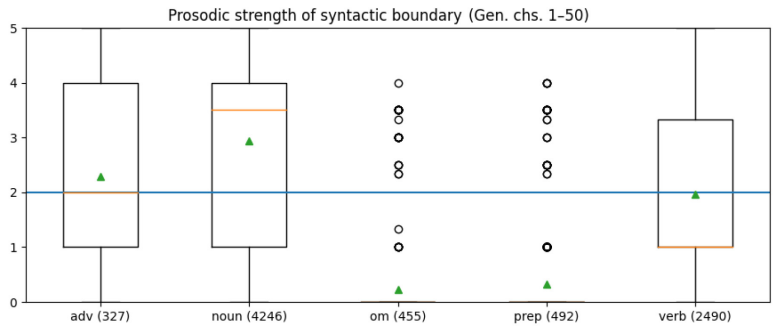


Figure 13: Exod. 1–40: Depth of prosodic division at the level of the graphematic word

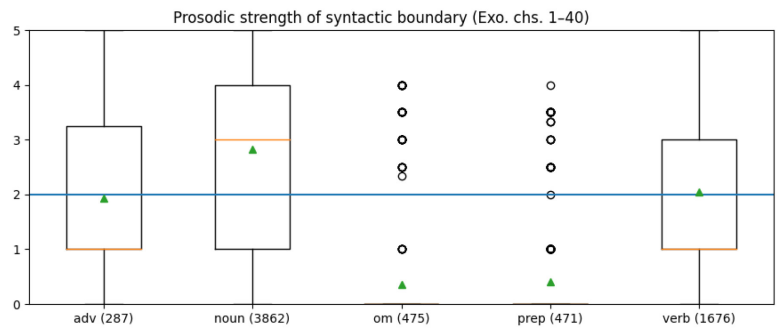
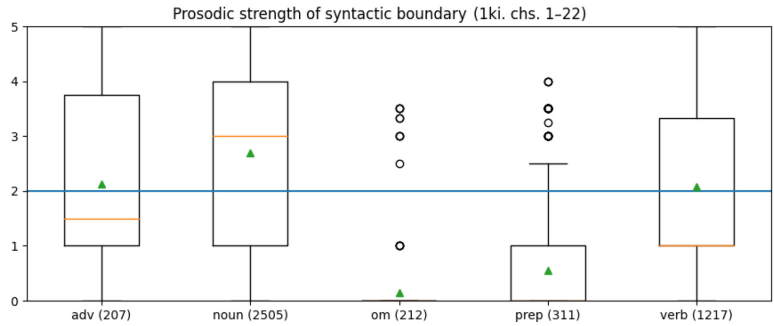


Figure 14: 1 Kgs 1–22: Depth of prosodic division at the level of the graphematic word



Figures 12–14 give the distribution of the strength of the prosodic boundary after instances of each word class (noun, verb, object marker [*om*], and preposition [*prep*]) in Genesis, Exodus and 1 Kings. Only sentences consisting exclusively of a main clause were included (i.e., subordinate clauses were excluded).

The figures show that the median accent on verbal forms is conjunctive (strength 1).[9] By contrast, the median accent on the noun is disjunctive (> 2). In terms of mean the verb also has a lower disjunctive strength than the noun: ≈ 2 vs ≈ 3. On both

[9] The results of this part of the investigation are here presented as 'box plots', plotted using *Matplotlib* (Hunter 2007). Box plots show the distribution of the data. Distributions are represented as boxes, where the lower edge of the box gives the first quartile (= median of the lower half of the dataset), the upper edge gives the third quartile. The box as a whole, therefore, provides a visualisation of the interquartile range. The lower and upper whiskers of the box plot identify the minimum and maximum respectively. The green triangle marks the mean, while the orange line marks the median. For further information on box plots, see https://en.wikipedia.org/wiki/Box_plot, last accessed 23 September 2022.

measures, therefore, the verb shows a lower propensity to bound prosodic phrases than nominal forms.

Prosodically weakest are the object marker (*om*) and the preposition, with a median boundary strength of 0 (*maqqef*), while the distribution of the adverb generally parallels that of the verb (see further below) although its disjunctive strength is perhaps marginally greater than the verb (see especially distributions in Genesis and 1 Kings).

From the average strength of the following prosodic division, a cline of prosodic status can be inferred:

noun > adverb > verb > object marker, preposition

The prosodic weakness of the object marker and the preposition is to be expected: both are closed word classes, and their members are highly grammaticalised. The prosodic weakness of adverbs is similarly expected, since (in the Macula scheme) important grammatical words such as the negative adverb are included in their number. By contrast, verbs are prosodically weaker than nouns. This is *a priori* unexpected, since the verb in Biblical Hebrew is an open lexical class, just like the noun, and would therefore be expected to behave prosodically like the noun. It is the goal of the rest of this chapter to ascertain a reason for this unexpected difference.

6.0. Prosody and Syntax in Tiberian Biblical Hebrew

6.1. Introduction

I turn first to the interface between prosody and syntax: in many syntactic models the verb forms a syntactic constituent, at some level of analysis, with the object (see e.g., 'Verb phrase', https://en.wikipedia.org/wiki/Verb_phrase, last accessed 28 March 2023). A related point is that many models expect the deepest prosodic divide in the sentence to be between the subject and the predicate (Dresher 1994, 25; Devine and Stephens 1994, 386;). Given the relationship that is generally found to exist between syntax and prosody, one might expect to find the strongest phonological divide to be between the subject and the predicate. The corollary of this is that the subject and the verb would be "unlikely to form a phonological phrase" (Devine and Stephens 1994, 386). It could be, therefore, that the relatively lower prosodic strength of the verb relative to the noun is due to the verb's tendency to form a prosodic phrase with a following object or non-subject phrase.[10]

[10] Constituent order in Biblical Hebrew has received considerable attention in recent years (see e.g., Holmstedt 2009; Hornkohl 2018; Khan and van der Merwe 2020). Indeed, there has been a growing appreciation of the importance of pragmatics and semantics, as well as syntax, in providing a context for understanding the surface word order of Biblical Hebrew sentences.

6.2. Quantitative Analysis of Prosodic Depth and Constituent Order

6.2.1. Constraints

The prosodic strength at the boundaries of major constituents in the same corpus was analysed, namely: Genesis, Exodus and 1 Kings. The constraints adopted were as follows: (a) sentences included consist only of the elements labelled, i.e., an SVO sentence starts with the subject constituent and ends with an O constituent; (b) constituents do not contain subclauses, i.e., an O constituent contains no subclause; (c) functional categories are calculated on the final element of the graphematic word. An important consequence of this is that VO sequences involving a verb and suffix pronoun, which will always manifest as O, were excluded.[11] This is helpful for the analysis, since, as enclitics, there is no prosodic boundary after V in such sentences, and these could skew the results to make the verbs look prosodically weaker than they in fact are; (d) non-verbal predicates (denoted P in Macula) are excluded.

To summarise: the sentences included in this study are main clauses containing no subordinate clauses. Consequently all verbs are main verbs. The reasons for restricting the corpus of verses in this way are: (a) since we are interested in the boundaries of verbal and non-verbal forms, it is important that we exclude the possibility of a boundary between a verb and non-verb

[11] This is because, from the perspective of the query engine such sentences will appear to have no verb, and so will not feature in the analysis, since only sentences including V constituents are included.

occurring at the right edge of a subordinate clause; and (b) the prosody of verbs in subordinate clauses is known to differ from that of main clause verbs in some languages (see below in this chapter for an example from Indo-European).[12]

6.2.2. Prosodic Boundary after First Syntactic Constituent

The distribution of prosodic strengths after the first functional phrase are given in the form of box plots in Figures 15–17.

Figure 15: Gen. 1–50 X|XX

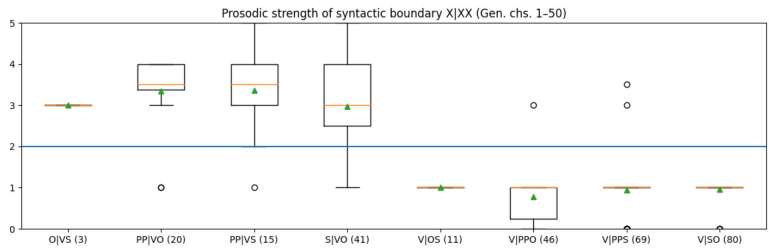
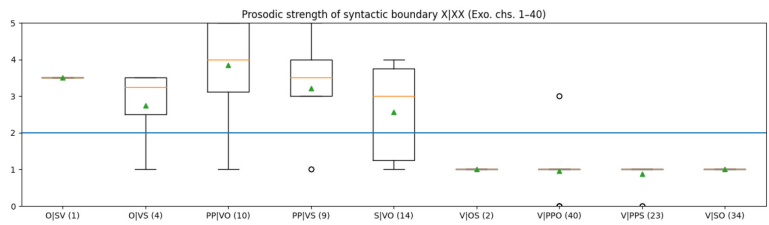
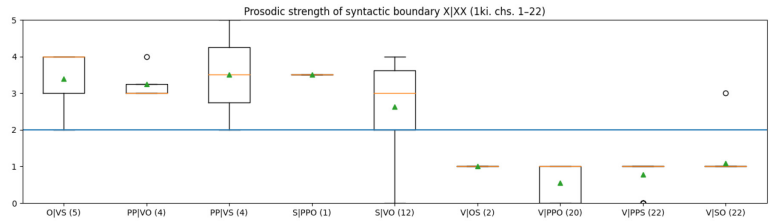


Figure 16: Exod. 1–40 X|XX



[12] Sequences included were: SVO, OVS, VOS, OVS, OSV, PPVS, SOV, SPPO, VPPS, VPPO, PPVS and PPVO. However, not all sequences were found in all books.

Figure 17: 1 Kgs 1–22 X|XX



All three plots show a remarkable distinction in the prosodic structure of the first two constituents in VXX sentences, versus XVX or XPPX sentences: in VXX sentences a prosodic phrase boundary is rarely if ever present between V and the constituent that follows it; by contrast there is almost always a prosodic phrase boundary after the first constituent of XVX sentences.

6.2.3. Prosodic Boundary after Second Syntactic Constituent

The distribution of prosodic strengths after the second functional phrase are given in the form of box plots in Figures 18–20.

Figure 18: Gen. 1–50 XX|X

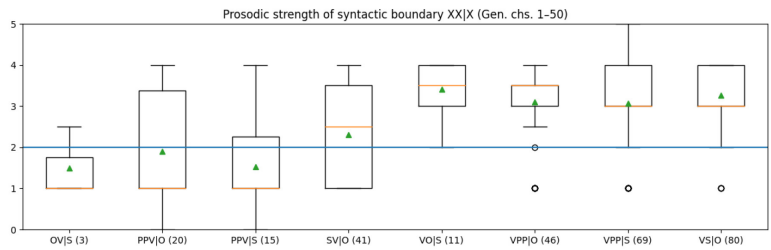


Figure 19: Exod. 1–40 XX|X

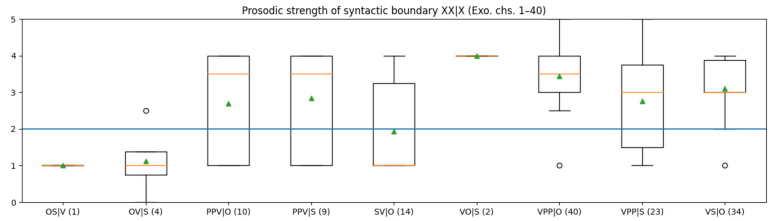
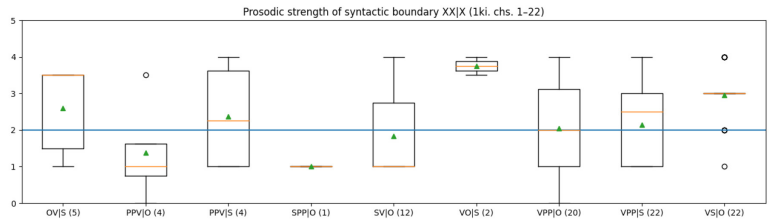


Figure 20: 1 Kgs 1–22 XX|X



At the boundary between constituents two and three we find almost the reverse distribution of that found in the prosodic structure of the boundary between constituents one and two, although the distinction between VXX and XVX sentences is not quite as clear as in the case of the boundary after the first constituent. In VXX sentences, there is usually a prosodic phrase boundary between X and X, whilst there is much more variation in the nature of the boundary between V and X in XVX sentences: the median accent is often conjunctive, i.e., the boundary is characterised by the lack of a prosodic boundary. However, the interquartile range often straddles both disjunctive and conjunctive accents, showing that in an important number of cases a prosodic phrase boundary occurs between V and X.

6.2.4. Differences between Books

It is regrettably beyond the scope of the present contribution to go into detail regarding the similarities and differences between Bible books. While the broad picture is remarkably consistent across the corpus analysed, it is worth noting some interesting differences, to be pursued in future work. In particular the boundary between V and X in XVX sentences in Exodus is much stronger than in the other books considered, although we also see median disjunctive boundaries in Genesis (SVO) and 1 Kings (OVS, PPVS). The number of instances in each case is small: further investigation is needed.

6.3. Discussion

In what follows the particular behaviour of SVO, VSO, V + X and ADV + X sequences are discussed. The guiding question at issue is to determine what might help explain the apparent prosodic weakness of the Biblical Hebrew main verb. Owing to the limited scope of the present contribution, examples are furnished from Genesis only.

6.3.1. SVO Sentences

While not statistically predominant, SVO sentences have been argued to be the basic word order in (Classical) Biblical Hebrew (Holmstedt 2009). Consider the SVO sentence in Figure 21 (syntax) and Figure 22 (prosody) of Gen. 4.1a.

Figure 21: Gen. 4.1a (syntax)

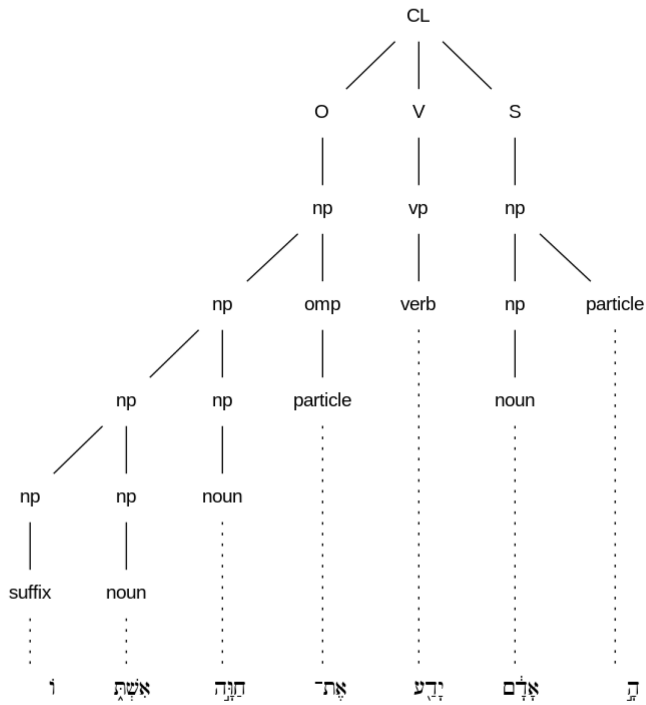
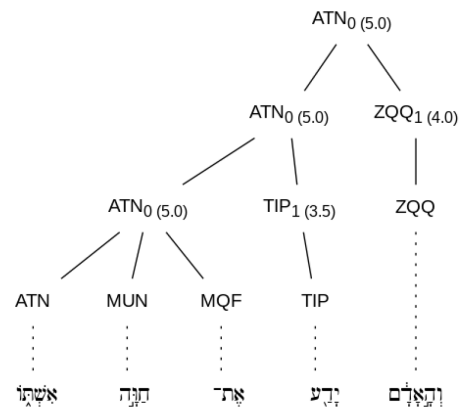


Figure 22: Gen. 4.1a (prosody)



Comparison of the prosodic and syntactic trees shows that the prosodic structure closely mirrors the syntactic, in that prosodic

phrase boundaries occur at major syntactic boundaries, namely, after S, הָאָדָ֔ם and after V, יָדַ֣ע.

Furthermore, although the disjunctive strength of the two accents, *little zaqeph* and *tifḥa* respectively, is the same (4 out of 5 on the scale used here), because disjunct strength is contextual (see above), the strength of the pause after the verb יָדַ֣ע is smaller than that after the subject הָאָדָ֔ם. This means that the deepest prosodic break in the sentence is after the subject, exactly where it is expected to be on syntactic grounds.

From the plots, it can be seen that there is a greater prosodic separation between subject and predicate than between verb and object. In 1 Kings and Exodus the median accent after the subject is a disjunctive, whereas between the verb and the object the median accent is conjunctive. In Genesis, in both cases, the median accent is a disjunctive, but the median strength of the disjunctive is lower. Here is evidence, therefore, that in Biblical Hebrew SVO sentences, syntax and prosody align as they would be expected to from a cross-/general-linguistic perspective. However, the difference is not very great, and in both cases the median accent is disjunctive.

6.3.2. VSO

As is well known, the statistically predominant word order in Biblical Hebrew is not SVO but VS(O) (Khan and van der Merwe 2020; for discussion of the 'basic' word order in Biblical Hebrew, see the discussion in Khan and van der Merwe 2020 and Holmstedt 2009).

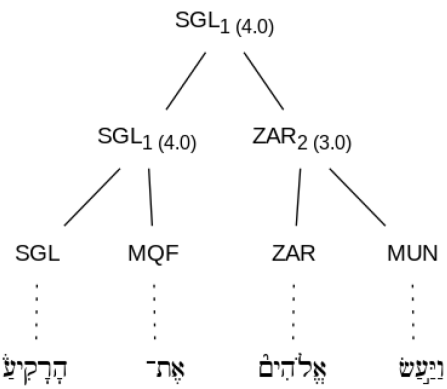
The syntactically significant point is that the subject in VSO sentences stands between the verb and the predicate. On the level

of surface syntax (leaving prosody aside briefly) the sentence cannot therefore be analysed straightforwardly into subject and predicate in cases where the predicate consists of more than a verb constituent.

The prosodic corollary of this is that in Biblical Hebrew we should obtain one or both of the following patterns: either (a) a prosodic divide occurs between the verb and the following constituents (S and O); or (b) a prosodic separation occurs between verb and subject, on the one hand, and the object, on the other.[13] Crucially, a prosodic boundary between subject and predicate (verb phrase) cannot be obtained, because the verb phrase is interrupted by the subject.

Figures 15–17 show that, in all three books considered, the first pattern, with a phrase boundary between V and S, is almost non-existent. Instead, all the examples are with a phrase boundary between S and O, e.g., Gen. 1.7a (Figure 23).

Figure 23: Gen. 1.7a (syntax)



[13] The entire sentence could, of course, form a prosodic phrase. However, this is unlikely except in the very shortest of sentences, e.g., Gen. 23.9a.

6.3.3. V + X

A verb is not restricted to forming a prosodic phrase with a following subject constituent. When initial, V forms a prosodic phrase with *any* following nominal constituent, whether S, O or PP. This emerges clearly from perusal of the final four columns of Figure 15 and Figure 17, which show that only a handful of instances of (mostly V + PP sequences) are separated by a disjunctive; otherwise V forms a prosodic phrase with the following constituent.

When V is second, there is also a tendency for it to form a prosodic phrase with the following constituent, although, as already noted, the plots show that there is a greater range of attested behaviour.

6.3.4. ADV + X

The prosodic behaviour of adverb phrases in relation to following syntactic constituents shows similar features to that of verbs (see Figure 24 and Figure 25).[14] This is to say that adverbs frequently form a prosodic phrase (or word) with a following item.

---

[14] The search included ADVVS, ADVOV, ADVSV, SADVV, VADVS, ADVVS, SADVV, OADVV, ADVPPV, VSADV, ADVPPS, ADVPPO. Not all these sequences were found in Genesis.

Figure 24: Gen. 1–50 X|XX: Depth of prosodic division between constituents in sentences containing adverbs

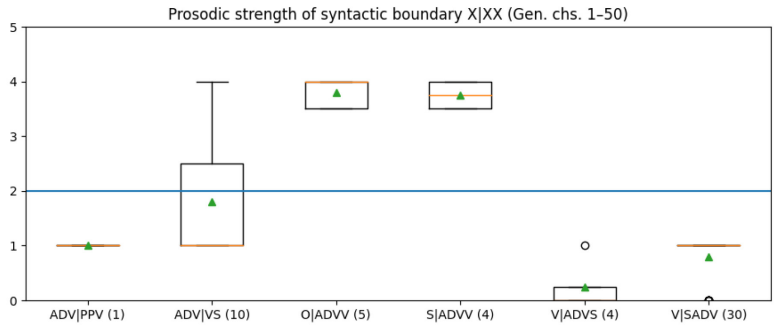
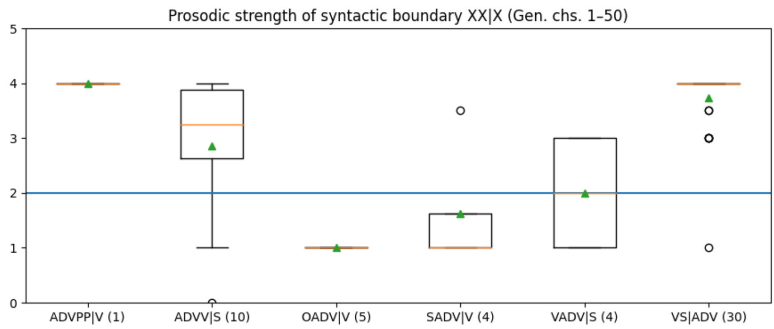


Figure 25: Gen. 1–50 XX|X: Depth of prosodic division between constituents in sentences containing adverbs



Closer inspection of the box plots reveals that in most cases the adverb forms a prosodic phrase with a following verb. This is perhaps not surprising, since adverbs typically modify verbs (hence the name).

6.3.5. (ADV + V) + X / (V + ADV) + X

Where an adverb is followed by a word of a class other than a verb, e.g., a noun phrase, there is typically a prosodic boundary, e.g., Gen. 45.21a (VADVS) in Figure 26 (syntax) and Figure 27 (prosody).

Figure 26: Gen. 45.21a (syntax)

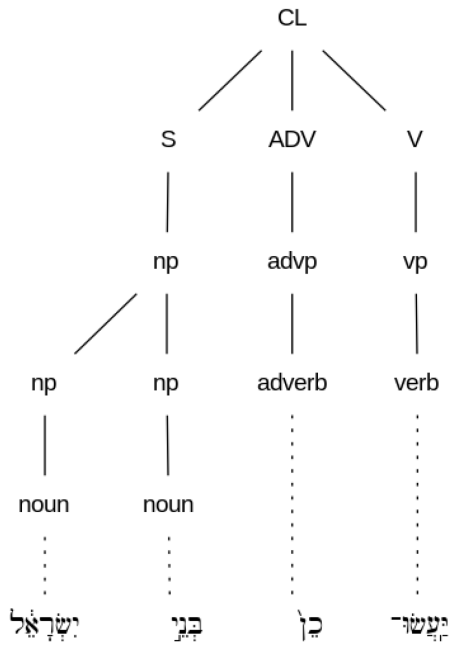
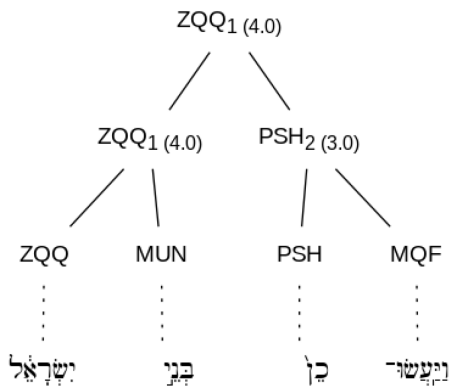


Figure 27: Gen. 45.21a (prosody)



Note, however, that here the verb forms a prosodic word with the adverb. It is then after this verb-adverb complex that the prosodic phrase boundary falls.

Figure 25 shows that a prosodic phrase boundary also tends to fall after a verb if it is preceded by an adverb. This is an interesting exception to the general rule, observed above, that a verb tends to form a prosodic phrase with a following item. Consider Gen. 15.5e (ADVVS) in Figure 28 (syntax) and Figure 29 (prosody).

Figure 28: Gen. 15.5e (syntax)

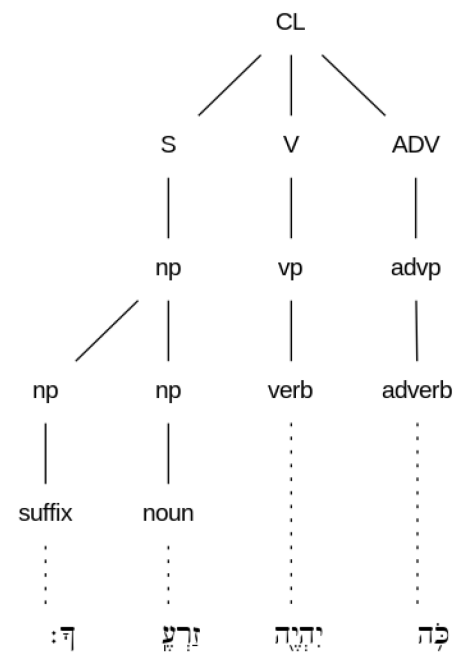
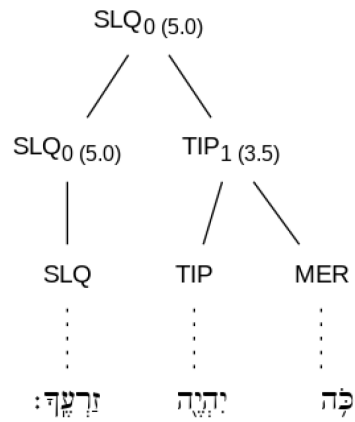


Figure 29: Gen. 15.5e (prosody)



The important point is that, both in this case and in the previous one, the verb forms a (syntactic) phrase with the adverb; the boundary between the adverb-verb complex and the following item therefore follows the last word of this complex, whether it is the verb or the adverb. (Note that this point is not revealed in the Macula syntax tree, since adverbs, or adverb phrases, are presented as top level constituents.)

7.0. Accounting for the Prosodic Integration of Tiberian Biblical Hebrew Verbs

7.1. Are Biblical Hebrew Verbs Clitics?

The verb clearly shows a propensity to form a single prosodic phrase with a following constituent, especially when sentence-initial. We have seen that this phenomenon is largely unrelated to the syntactic role of the constituent following the verb: S, O and PP all show the same interaction with V in this position; the only exception to this is in SVO sentences, where a prosodic

phrase boundary is present after V. In most syntactic contexts, therefore, it appears to be the fact of being a verb that drives the prosodic weakening of V. (I will return to the issue of SVO sentences at the end of the chapter.)

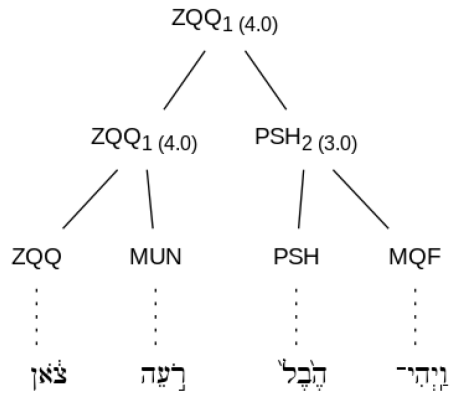
The propensity of V to become incorporated into the prosodic phrase of a following morpheme might be considered analogous to cliticization: in both cases a morpheme becomes prosodically integrated with a neighbouring one. Another similarity with cliticization is that the prosodic integration of V has a clear polarity (for a discussion of clitic polarity in connection with Ancient Greek, see Goldstein 2016, 60–68): it always forms a prosodic phrase with the following constituent, but rarely if ever with the previous one.

Indeed, the lack of accent on Hebrew verb forms, except *in pausa*, was already proposed by Kuryłowicz. In a discussion of the details of the system of Hebrew accentuation within the word (in comparison with Vedic), he states, "The above data suggest as their simplest explanation an original distribution of stressed and unstressed forms of simple verbs, viz. stressed forms *in pausa*, unstressed forms before another member of the sentence" (1959, 128).

However, the proposal that verbs are clitics in Biblical Hebrew suffers from two principal difficulties. First, the prosodic dependence of V differs from (prototypical) clitics in that the unit so formed is (for the most part) a prosodic *phrase*, not a prosodic *word*: phonological clitics (see introduction) do not carry their own accent (see e.g., Caink 2008, 491), whereas verbs in (Tiberian) Biblical Hebrew do. Exceptional are cases where the verb

does form a prosodic word with the following item, such as Gen. 4.2b (see Figure 30):

Figure 30: Gen. 4.2b (prosody)



A further objection is that the verb is an open class, whereas, as we have seen, clitics are required (at least by some scholars) to belong to closed (functional) classes (Caink 2008, 491) In the following sections I set out to address these concerns in turn.

7.2. Excursus: A Parallel from Indo-European

The prosodic weakening of V in Tiberian Biblical Hebrew has a striking parallel in a feature of early Indo-European, namely the lack of accent on main clause verbs in Vedic Sanskrit (Kuryłowicz 1959; details and exceptions see Kuryłowicz 1959, 123; Hock 1982, 1; Hock 2014, 154; Hock 2015, 69). “As is well known, (simple) Vedic finite verbs are unaccented in main clauses (unless initial in the clause or poetic line), but accented in dependent structures” (Hock 2015, 69).

Wackernagel offered an explanation of this behaviour in terms of cliticization. "[W]e must expect that in the ancestor language the verb in the main clause was placed immediately after the first word in the clause because, and insofar as, it was enclitic" (quote 2020, 327; see also Hock 2015, 70). Wackernagel himself was equivocal (2020, 327):

> I do not wish to deny that the proposal put forward here could be made less general. For the law regarding the placement of enclitics (disregarding e.g., vocatives) we have only been able to adduce examples in which the enclitic is no larger than two syllables. It could therefore be said that the law was only valid for monosyllabic and disyllabic enclitics, and that those of more than two syllables remained in the position that the constituent in question would otherwise receive—or at least, to express the idea more carefully, that above a certain size threshold an enclitic was not bound by the positional law of the enclitics. Applying this to the verb would lead to the assumption that monosyllabic and disyllabic verb forms, or shorter verbal forms below a certain threshold, moved to second position in main clauses, and that the other verbal forms in main clauses kept to the position that was dominant in subordinate clauses.

Nevertheless, it was accepted for a number of years (see Hock 1982, 3 and Hock 2015, 52), although it has fallen out of favour more recently and other explanations adduced (Hock 2015, 69–71). For present purposes it is sufficient to note that a key part of the rejection of the verb-as-clitic proposal is that verbs are prototypically too prosodically heavy to have clitic status (Hock 1982, 1–4). Thus Fortson denies clitic status to the verb, except in limited circumstances (2008, 266):

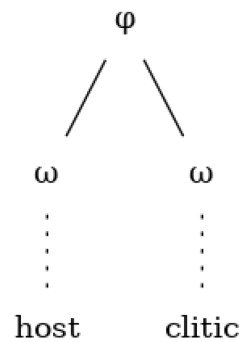
> It should also be emphasized that for the other Indo-European languages, even if verbs were atonic or lower in pitch, there is no evidence that they were clitics except for special instances, and, as in Latin, the strongest evidence of cliticization and/or weaker accentuation comes from the mono- and dissyllabic forms. In Greek, the only truly enclitic verbs are monosyllabic and dissyllabic forms of εἰμί and φημί.

7.3. Prosodic Word Clitics

A major difficulty with positing prosodic incorporation at the level of the prosodic phrase is, therefore, that clitics are usually defined as morphemes that are phonologically deficient at the level of the prosodic word (cf. Anderson 2005, 45–46). However, while most discussion of prosodic deficiency/dependency focuses on these sub-prosodic word clitics, the existence of sub-prosodic phrase clitics has also been observed (Anderson 2005, 32, 46; Goldstein 2016, 48–52). Anderson gives the examples of Italian *loro* 'to them' and Tagalog *tayo* 'we (dual)' as cases in point (2005, 32). These clitics, sometimes termed 'special clitics', are incorporated with their host to project a prosodic phrase (Anderson 2005, 46; Goldstein 2016, 48; see Figure 31).

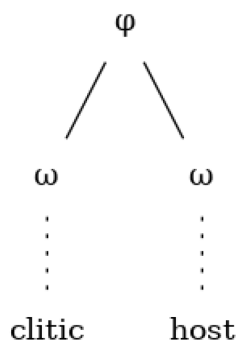
Ancient Greek 'postpositives' have been analysed in these terms (Goldstein 2016, 48–52): these carry their own accent, but nevertheless have syntactic distributions that demonstrate dependence on a host (Goldstein 2016, 48–52). As the name suggests, they are placed immediately after their prosodic host (Figure 31). Morphemes in this category are all functional particles, such as γάρ, μέν, δέ (for further information see Goldstein 2016, 48, 51).

Figure 31: Prosodic word clitic: Ancient Greek postpositive (per Goldstein 2016, 51, fig. 3.16)



An analysis of Biblical Hebrew verbs and adverbs as prosodic word clitics can account for the distribution we have observed: in our case, the verb is placed prior to the host, per Figure 32.

Figure 32: Prosodic word clitic: Hebrew verb



However, while this analysis works at the prosodic level, it remains the case that the examples from Italian, Tagalog and Greek discussed in this section involve unambiguously functional morphemes: in neither case are open class morphemes, such as verbs

or nouns, involved. It is to the functional/lexical distinction that I therefore now turn.

7.4. The Functional/Lexical Distinction is a Continuum

The distinction between function and content classes is often presented as a hard-and-fast binary distinction. Such a distinction is presupposed in statements such as that of Caink, that "[i]tems that are clitic vary across languages but are always grammatical (or functional) words" (2008, 491). If clitics are always function words, it must be possible to be categorical about whether or not a word is a function word in the first place.

However, it has been argued that the lexical/functional distinction is better seen as a continuum. Klammer, Schulz and Della Volpe point out that some lexical items have very clear semantic content, giving the examples 'finger', 'tree', 'swim', 'humid', while the semantic content of items such as 'thing' and 'do' is much less clear (2007, 96).[15] Gelderen adduces the case of English prepositions: some, such as 'behind' and 'toward' have features of both function and lexical categories (2004). Developing this, let us consider the preposition 'behind'. This has a grammatical function, namely to relate two elements in the discourse to one another: in 'The cat sat behind the sofa', 'behind' grammatically relates 'the cat' to 'the sofa' in the discourse. However, 'behind' also carries semantic content: it refers to a relationship in

[15] Indeed, 'do' is (in part) an auxiliary verb in English.

the world outside the discourse, namely the relative geographical location of the cat and the sofa.

Much more grammatical is the preposition 'of', e.g., 'The house of friendship', where the grammatical function is clear, relating 'house' and 'friendship'. However, it is hard to detect any semantic content.

In this study we have found that verbs and adverbs show prosodic dependence at the level of the prosodic phrase. It is therefore relevant to point out that verbs and adverbs have characteristics that align with function words over against lexical words.

The verb is often seen as a prototypically open lexical class. (The major exception to this are auxiliary verbs, if present in the language under consideration, whose semantic content has typically been bleached to the extent that only the grammatical function is left.) However, it is less commonly observed that verbs by their nature are more grammatical than the other prototypically open lexical class, namely nouns. This is trivially the case with predicative verbs such as 'to be'. In 'The cat is on the sofa', the verb 'is' structures the predication. It does not refer to anything concrete outside the discourse.

However, verbs with much greater semantic content, such as 'walk', are also more grammatical than a noun, such as 'station'. Consider the sentence 'I walked to the station'. Here the verb 'walk' has semantic content: it refers to the activity of stepping so as to make progress relative to the ground. Language users will recognise this activity based on their experience of living in the world. However, in addition to its semantic role, 'walk'

here has the function of describing a relationship of movement between the subject and the object: 'I walked to the station' predicates a 'walking to the station' event of the subject. This is to say, 'walk' structures the predication, in addition to its lexical function.

Adverbs are similar. Like verbs, adverbs vary in the degree of their lexical content. Some adverbs, such as the negative adverb 'not', are entirely functional. By contrast, an adverb like 'slowly' has both lexical and grammatical content: lexically it refers to 'slowness'; grammatically, it attributes 'slowness' to the event described in the sentence.

Nouns fundamentally do not have the same potential for grammaticality that verbs and adverbs do: even semantically bleached nouns, such as 'thing'—as in 'I saw an interesting thing yesterday'—do not relate items within the discourse to one another, even though the referent in the world outside the discourse is very ill-defined. Word classes can therefore be arranged on a grammaticalisation cline:

nouns < verbs, adverbs < prepositions, particles

The order of the elements of word classes on the grammaticalisation cline closely parallels that of the prosodic strength cline (see above). I therefore suggest that a relationship exists between the prosodic status of a morpheme, and its position on the lexical/functional cline, a connection that extends beyond the prosodic word to the prosodic phrase, enabling, in principle, morphemes that meet the requirements for prosodic wordhood to be prosodically dependent on a host at the level of the prosodic phrase.

7.5. What Motivates the Cliticization of V?

In the previous section I argued that there are grounds for seeing the verb as more functional than is often recognised. The fact, remains, however, that in the literature prosodic word clitics are words that belong unambiguously in the functional category, rather than verbs, which share both lexical and functional characteristics. It is therefore worth considering what might be motivating the cliticization of verbs in Biblical Hebrew.

We observed that verbs in Biblical Hebrew have a strong tendency to become incorporated into a following prosodic phrase in all word orders except SVO. V is also prosodically weakest when fronted to first position in the sentence, i.e., VXX. We also saw that SVO has been proposed as the basic word order in Biblical Hebrew, albeit not attested as frequently as VSO (Holmstedt 2009). One route to the observed distribution is, therefore, the following: Biblical Hebrew underlying word order is SVO, but in contexts of movement, and especially where V moves to first position, V loses prosodic independence and cliticises to a following element, whatever that may be. Note that a reversed explanation, where the basic word order is VSO, cannot explain why V should also be prosodically weak before all other elements in that position, i.e., O and PP.

The implication of this is that verb-first word orders in Biblical Hebrew may be in part at least phonologically driven. Compare the proposal that non-SVO word orders are cases of phonological movement in Ancient Greek (Agbayani and Golston 2010).

This hypothesis is presented tentatively: it needs to be tested by looking in detail at individual cases, and developed further in the light of crosslinguistic evidence. Whatever the correct explanation turns out to be, it is clear that there is something special about SVO sentences that leads V to be prosodically stronger in that position.

8.0. Conclusion: Biblical Hebrew Verbs are Prosodic Word Clitics

We have seen that verbs have a marked propensity to form a prosodic phrase with a following element, regardless of the syntactic role of that element. We observed that this behaviour is analogous to cliticization: both cases involve the incorporation of an element (the verb or the clitic) into a neighbouring prosodic unit (prosodic phrase or prosodic word, respectively). I raised two objections to this analysis: (1) cliticization is generally regarded as a phenomenon operating at the level of the prosodic word, rather than at the level of the prosodic phrase; (2) clitics are generally required to belong to closed functional classes, whereas the Hebrew verb (and the verb more generally) is an open lexical class;

I addressed the first objection by pointing out that, while most discussion of cliticization focuses on its manifestation at the level of the prosodic word, cliticization at the level of the prosodic phrase has also been observed to exist. It is therefore typologically acceptable for a prosodic word to depend prosodically on a neighbouring prosodic word to form a prosodic phrase, in a way analogous to a sub-prosodic word morpheme becoming integrated into a neighbouring prosodic word.

This analysis necessitates a loosening of the requirement for clitics to belong to a closed, viz. functional, word class. Further, I have argued that such a loosening is justified: the functional/lexical distinction is itself a continuum, not binary; verbs (and adverbs) occupy a position on this continuum somewhere between the (prototypically lexical) nouns and the (prototypically grammatical) object marker and prepositions. It is consistent with this that verbs (and adverbs) should sit between these two groups on the continuum of prosodic strength as well, that is, able to constitute their own prosodic words, but dependent on a host at the level of the prosodic phrase.

Biblical Hebrew verbs have sufficient prosodic weight (for the most part) to be able to stand as their own prosodic word, viz. carry their own accent. They do not in general have sufficient prosodic weight, however, to bound a prosodic phrase on the right, that is, to carry a disjunctive accent. The major exception to this is in SVO sentences, where a disjunctive is much more liable to occur at the boundary between V and O than between V and X in other word orders. I suggested that this is because in all word orders other than SVO, the verb is fronted and thereby loses prosodic independence. I leave it to further research to pursue this suggestion.

Acknowledgements

The present contribution was completed as part of ongoing research on the CROSSREADS project. This project has received funding from the European Research Council (ERC) under the

European Union's Horizon 2020 research and innovation programme (CROSSREADS: Grant agreement No. 885040).

I wish to thank the other participants of the InterSAME workshop, the editors and anonymous reviewers, as well as Randall Tan, James Cuénod, and Ryder Wishart, for their helpful questions and comments. As ever, any deficiencies remain my responsibility.

Trees are drawn on the basis of analyses in the Macula and TanakhML trees. The syntactic data was taken from the MACULA Hebrew Linguistic Datasets, available at https://github.com/Clear-Bible/macula-hebrew. The trees used were those as of 22nd May 2022. The TanakhML Project (https://www.tanakhml.org/) was founded by Alain Verboomen and is hosted by Emmanuel Dyan. The trees were downloaded between August 2022 and March 2023. The analyses are conducted using software written by the present author in Python. The software converts the XML of these treebanks into DOT (see https://graphviz.org/doc/info/lang.html). The resulting DOT file is compiled to PNG format using the Graphviz compiler in *Dependency2Tree* (https://github.com/boberle/dependency2tree/, last commit 2020) and *Graphviz* (https://graphviz.org/).

References

Agbayani, Brian. and Chris Golston. 2010. 'Phonological Movement in Classical Greek'. *Language* 86: 133–167.

Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. New York: Oxford University Press.

Aronoff, Mark. 1985. 'Orthography and Linguistic Theory: The Syntactic Basis of Masoretic Hebrew Punctuation'. *Language* 61: 28–72.

Caink, Andrew D. 2008. 'Clitics'. In *Encyclopedia of Language and Linguistics*, edited by Keith Brown, 491–495. Oxford: Elsevier Science.

Crellin, Robert S. D. 2022a. *The Semantics of Word Division in Northwest Semitic Writing Systems: Ugaritic, Phoenician, Hebrew, Moabite and Greek*. Oxford: Oxbow.

———. 2022b. 'Word-level Punctuation in Latin and Greek Inscriptions from Sicily of the Imperial Period'. In *Writing Around the Ancient Mediterranean: Practices and Adaptations*, edited by Philippa M. Steele and Philip Boyes, 195–219. Oxford: Oxbow.

Devine, Andrew M., and Laurence D. Stephens. 1994. *The Prosody of Greek Speech*. Oxford: Oxford University Press.

Dresher, B. E. 1994. 'The Prosodic Basis of the Tiberian Hebrew System of Accents'. *Language* 70: 1–52.

Evertz, Martin. 2018. *Visual Prosody: The Graphematic Foot in English and German*. Berlin, Boston: De Gruyter.

Fortson, Benjamin. 2008. *Language and Rhythm in Plautus: Synchronic and Diachronic Studies*, vol. 3. 1st ed. Berlin, Boston: De Gruyter.

Fries, Charles C. 1952. *The Structure of English: An Introduction to the Construction of English Sentences*. London: Harcourt, Brace and World.

Gelderen, Elly V. 2004. 'Function Words'. In *Encyclopedia of Linguistics*, edited by Philipp Strazny. New York: Taylor and

Francis. https://www.proquest.com/encyclopedias-reference-works/function-words/docview/2137931488/se-2.

Goldstein, David M. 2016. *Classical Greek Syntax: Wackernagel's Law in Herodotus*. Leiden, Boston: Brill.

Haug, Dag T. T. 2009. 'Does Homeric Greek have Prepositions? Or Local Adverbs? (And What's the Difference Anyway?)'. In *Grammatical Change in Indo-European Languages*, edited by Vit Bubenik, John Hewson and Sarah Rose, 103–120. Amsterdam: John Benjamins.

Hock, Hans H. 1982. 'Clitic Verbs in PIE or Discourse-based Verb Fronting? Sanskrit Sá Hovaca Gárgyaḥ and Congeners in Avestan and Homeric Greek'. *Studies in the Linguistic Sciences* 12 (2): 1–38.

———. 2014. 'Vedic Verb Accent Revisited'. In *Vedic and Sanskrit Historical Linguistics*, edited by Jared Klein and Elizabeth Tucker, 153–178. Delhi: Motilal Banarsidass.

———. 2015. 'Proto-Indo-European Verb-finality: Reconstruction, Typology, Validation'. In *Proto-Indo-European Syntax and its Development*, edited by L. I. Kulikov and Nikolaos Lavidas, 51–78. Amsterdam, Philadelphia: John Benjamins.

Holmstedt, Robert D. 2009. 'Word Order and Information Structure in Ruth and Jonah: A Generative-typological Analysis'. *Journal of Semitic Studies* 54 (1): 111–139.

Hornkohl, Aaron D. 2018. 'Biblical Hebrew Tense-aspect-mood, Word Order and Pragmatics: Some Observations on Recent Approaches'. In *Studies in Semitic Linguistics and Manuscripts: A Liber Discipulorum in Honour of Professor Geoffrey*

Khan, edited by Nadia Vidro, Ronny Vollandt, Esther-Miriam Wager and Judith Olszowy-Schlanger, 27–56. Uppsala: Uppsala Universitet.

Hunter, Jonathan D. 2007. 'Matplotlib: A 2D Graphics Environment'. *Computing in Science & Engineering* 9: 90–95. https://doi.org/10.1109/MCSE.2007.55

Janis, Norman. 1987. 'A Grammar of the Biblical Accents'. PhD dissertation, Harvard University.

Khan, Geoffrey. 2020. *The Tiberian Pronunciation Tradition of Biblical Hebrew*. Cambridge: University of Cambridge and Open Book.

Khan, Geoffrey, and Christo H. J. van der Merwe. 2020. 'Towards a Comprehensive Model for Interpreting Word Order in Classical Biblical Hebrew'. *Journal of Semitic Studies* 65 (2): 347–390. https://doi.org/10.1093/jss/fgaa025.

Klammer, Thomas P., Muriel Schulz, and Angela Della Volpe. 2007. *Analyzing English Grammar*. 5th ed. New York: Pearson, Longman.

Kuryłowicz, Jerzy. 1959. 'The Accentuation of the Verb in Indo-European and in Hebrew'. *WORD* 15: 123–129. https://doi.org/10.1080/00437956.1959.11659688.

Nespor, Marina, and Marina Vogel. 2007. *Prosodic Phonology: with a New Foreword*. Berlin: Mouton De Gruyter.

Park, Sung J. 2020. *The Fundamentals of Hebrew Accents: Divisions and Exegetical Roles Beyond Syntax*. Cambridge: Cambridge University Press.
https://doi.org/10.1017/9781108801782.

Payne, Thomas E. 1997. *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: Cambridge University Press.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. London: Oxford University Press.

Selkirk, Elisabeth. 1996. 'The Prosodic Structure of Function Words'. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, edited by J.L. Morgan and K. Demuth, 187–214. New York, NY: Psychology Press.

Tan, Randall, and Andi Wu. 2022. *MACULA Hebrew Treebank for Open Scriptures Hebrew Bible (OSHB) Initial Release Documentation*. Biblica. https://github.com/Clear-Bible/macula-hebrew/blob/main/doc/MACULA%20Hebrew%20Treebank%20for%20Open%20Scriptures%20Hebrew%20Bible.pdf.

Truckenbrodt, Hubert. 1999. 'On the Relation between Syntactic Phrases and Phonological Phrases'. *Linguistic Inquiry* 30: 219–255.

———. 2007. 'The Syntax–phonology Interface'. In *The Cambridge Handbook of Phonology*, edited by Paul d. Lacy, 435–456. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486371.019.

Wackernagel, Jacob. 2020 [1892]. *On a Law of Indo-European Word Order: Über ein Gesetz der Indogermanischen Wortstellung*. Translated by George Walkden, Christina Sevdali, and Morgan Macleod. Berlin: Language Science Press. 10.5281/zenodo.3978908.

Yeivin, Israel. 1980. *Introduction to the Tiberian Masorah*. Missoula: Scholars Press.