The Masoretes and the Punctuation of Biblical Hebrew

Biblical Hebrew was the first classical language to be systematically punctuated. In comparison with modern languages the system was very complex: there are more commonly used punctuation marks in Hebrew than there are letters in the English alphabet, and they are applied not only between phrases but to virtually every word of the scriptures, indicating very precisely each word's relationship with others in a phrase as well as dividing each sentence into hierarchically ordered word-groups. This complexity, coupled with sometimes substantial orthographic variations between printed editions, has led many students of the language to neglect Hebrew punctuation entirely.

The advent of computer typesetting, and in particular the inclusion of these marks in the magnetic text of *Biblia Hebraica Stuttgartensia*, challenges scholars in linguistics and computation to cooperate in a more detailed analysis of this punctuation than has ever before been possible. But even without further research it provides the opportunity to analyse the logical structure of the Hebrew text by computer, and to apply the results to Bible translation and to translation checking. This paper is a preliminary attempt to stimulate such a dialogue, first by addressing in detail the historical background which gave rise to the distinctively Hebrew system of punctuation, and secondly by providing a brief account of the most common elements of the system.

a. Historical Background

1. The development of punctuation in the West

In general, ancient languages were not punctuated. In some Egyptian poetic texts line and stanza marks exist as aids to the reader, and some early Greek inscriptions use a vertical ellipsis as a phrase or sentence divider, but the only punctuation apparently known to Aristotle was the *paragraphos*, an underscore to indicate the beginning of a new topic.

The invention of punctuation to mark the grammatical or logical structure of texts is generally attributed to Aristophanes of Byzantium. As chief librarian of Alexandria (c.195 BCE) Aristophanes was responsible for copying and annotating many works from classical Greek literature. In doing so he added breathings and accents and devised a simple three part system of punctuation by which a text was divided by dots placed at the end of words: a short clause (*Komma*) had the dot in the middle, a phrase (*Kôlon*) at the bottom and a sentence (*Periodos*) at the top. In spite of its obvious advantages, this system was scarcely used after Aristophanes except to clarify a few obscure poetic texts.

Latin inscriptions, unlike Greek ones, marked word divisions by a small dot. At the beginning of the Common Era these points were also used in books and a paragraph system akin to that of Greek came into being: a new subject was introduced by

starting a new line and changing the indentation, much as in modern European languages (except that instead of being indented the line was extended into the left margin as a 'hanging paragraph'). This practice was particularly used in texts which were composed to be read aloud in public (Cicero used it to mark natural breaks in his Latin orations as had Demosthenes in Greek) and was comparatively short-lived. In spite of occasional attempts to revive Aristophanes' system for Latin (most notably by Aelius Donatus who was Jerome's tutor), from the fourth century to the ninth the only consistently punctuated extant Latin text (and that just by *paragraphos*) is Jerome's Vulgate.

It was only under the influence of Charlemagne in the ninth century that punctuation systems developed in Western Europe. By this time the churches were beginning to use *Neumes* as an early musical notation for chanting and some of these, along with Aristophanes' Komma and Periodos, were gradually added to the punctuation schema, notably the *punctus elevatus* (!) and the *punctus interrogativus* (?). These had a dual function: they indicated both a change of pitch and a grammatical nuance. Although this system was fully developed by the twelfth century it was largely confined to chanting, and it was not until after the invention of printing that it was consistently used. In the work of Caxton there is no very clear relationship between punctuation and syntax; the consistent modern use of punctuation in the service of clarity dates back only to the fifteenth century (probably to the Venetian printer Aldus Manutius, d. 1515 or possibly to his grandson of the same name).

2. Punctuation in Hebrew

Masoretic Hebrew punctuation is similar to that based on the neumes in that it is very closely related to the musical pitch used in incantation. It is different in two major respects: on the one hand, it is far more comprehensive, and on the other, it became a norm for all Jews who read their scriptures as soon as it was introduced. Both of these differences were caused by political, religious, social and economic forces which culminated in the appearance of the *Masoretic Text* in the tenth century.

2a. The Jewish Academies – Mishna and Talmud

After the destruction of the Temple Galilee became the administrative centre of Judaism. The Sanhedrin moved to the new city of Tiberias and the Jewish religious academies (called *yeshivot*) moved from Jerusalem to Tiberias and Caesarea. These academies served two purposes. They were centres for training rabbis in what were to become known as the *Miqra* (the *written* scriptural Law) and the oral teachings of the great scholars, the *Mishna* (*repeated i.e.* from memory). They were also institutes of advanced scholarship where highly trained rabbis working under the direction of an officially appointed scholar called the *Nasi* (literally *prince*) devoted themselves to producing a unified and normative interpretation of the oral Law. The first fruits of this study were the tractates of the Mishna which appeared in the third century CE under the direction of Juda ha-Nasi. These were almost entirely produced in the academy at Tiberias.

Meanwhile, similar academies had been established in Sura and Nehardea in Babylon. The Babylonian academies were placed under the direct control of the *resh galuta* (*head of exile* or exiliarch), a fact which was later to have serious repercussions. The quality of scholarship produced by the Tiberian and Babylonian academies can scarcely be overstated. In them thousands of students were trained rigorously for many years, and those who graduated and left the academies were required to return twice a year for a month's 'refresher' course called *kalla* (*completion* of study). There was comparatively free movement, and therefore exchange of information, between the academies.

Now that Mishna was regarded as virtually canonical, scholars from the academies in both Palestine and Babylon undertook the task of expanding it to include all the main teachings of the greatest Jewish scholars – a work completed in the fifth century. This expansion, the *Gemara* (another word for *completion*) took different forms in Tiberias and Sura. Palestine fell under the control of Constantinople, and persecution under successive Christian emperors severely disrupted Jewish learning there, but Judaism continued to flourish in Babylon. It is probably for this reason that the Palestinian gemara is much shorter than the Babylonian one (only about a third of its length) and does not cover all the tractates of Mishna. The Mishna and the Gemara, together with some further writings, form the *Talmud* (*Learning*). The Babylonian recension is called the *Talmud Bavli* or Babylonian Talmud, as opposed to the inappropriately named *Talmud Yerushalmi* or Jerusalem Talmud.

It should be stressed that, from a Rabbinic perspective, the oral Law of Mishna and Gemara, although assembled in the academies over centuries, was no less authoritative than the Miqra. Like the text of the Torah it had been given as a promise by Moses. A Rabbinic exegesis of Ex. 24:12 says that the Tables are the Decalogue, the Law is the Scripture, the Commandment is the Mishna, the "that I have written" is the Prophets and Writings and the "to teach them" is the Gemara – "...so telling us that all of these were given to Moses from Sinai."

2b. From Talmud to Masoretic Text

The *Torah* or Pentateuch had long been the pivot of the Jewish scriptures and as such had been transmitted with the utmost care. The remainder of the scriptures, the *prophets* and *writings*, had not been so scrupulously preserved. It was not until the synod of Jamnia in the last decade of the first century CE that final agreement was reached on the canon of the Jewish scriptures, and by that time several different Hebrew (and Aramaic) texts were in circulation. Although the academies had attempted to produce an agreed text, it was not until a consensus of interpretation was reached that the rabbis were able to decide on a definitive text of the whole of scripture. The Talmud was the embodiment of this consensus.

But between the compilation of the Mishna and that of the Gemara the language of the Jews in Palestine had changed. The Mishna is written in Hebrew: not the Hebrew of the scriptures but a relatively similar language. It is sometimes asserted that Hebrew had been replaced by Aramaic as a spoken language before the Common Era. This is certainly not true: the Jewish revolutionary Bar Kokhba wrote letters to

his family and friends, some in Aramaic and some in Hebrew, between the years 132 and 135 CE. There is at least a possibility that even in the third century CE the language of the Mishna was spoken by ordinary people and not just by Hebrew scholars. But the Gemara is written in Aramaic. It is clear that by the fifth century both Biblical and Mishnaic Hebrew were 'dead' languages known only to scholars.

Of course, Hebrew continued to be taught. At a time when only a few highly educated Europeans could read and write, the Jewish educational system was very advanced. The education of male children was a religious obligation, and primary schools existed in all towns and many villages. Boys started learning scripture by six. Several went on to study Mishna by ten and the most able studied Talmud by fourteen. But of every thousand boys who started school only one could be expected to become a proficient scholar. If the scriptures were to survive, new types of help to readers needed to be devised.

The earliest Hebrew manuscripts, in common with many ancient languages, had no punctuation system (except for starting a new line to indicate a new topic) and Hebrew had no vowels in its alphabet. Neither of these omissions was important as long as Hebrew was a spoken language. Cases where a vowel could readily be mistaken were to some extent catered for at an early period by inserting consonants to help the reader: *h* for an *a* sound, *y* for *e* or *i* and *w* for *u* (although many ambiguities remained). Full stops were inserted to divide the text into sentences or verses, possibly as early as the second century BCE, and most verses were marked to indicate the most important pause to make when reading aloud. Before the Talmud was completed the rabbis had emended these for consistency and had added other marks to clarify the text. A vertical bar | was written between words when it was felt necessary to clarify where one ended and another began (inter-word spaces were not regularly used), and stress markers were inserted. In Hebrew most words are stressed on either the last or the last-but-one syllable (a very small number are stressed on the last-but-two, and several longer words also have a secondary tone), and it is very important to know where this stress is placed. Pronouncing the accent on the wrong syllable can drastically change the meaning of a word. These tasks had been completed and agreed upon by the sixth century. Unfortunately no manuscripts from this period have survived, but we know from later writers that a *Codex Hilleli*, ascribed to Rabbi Hillel ben Moses ben Hillel around the year 600 CE, and a *Codex Muga* were regarded as authoritative. After some 500 years of scholarly labour a definitive text of the Hebrew Bible had at last been produced.

This should have been the end of the Academies' work. But the demise of Hebrew as a spoken language provided a new challenge: without vowels in the text it was very hard, perhaps impossible, for anyone other than a rabbi to understand the Scripture or pronounce it correctly. A system of vocalisation had to be made and (since the text itself was now regarded as definitive) it had to be inserted without making changes to the letters in the text. To achieve this, the rabbis began to devise a system of dots to represent the different vowels. The Hebrew word for a dot or spot is *niqqud* (it is the word used to describe Joseph's *speckled* cattle in Gen. 30:39). A related and very

similar word *noqed* means *a shepherd*, and we can imagine a typically rabbinic pun that the *dot* is a *shepherd* to guide the reader.

Because the task of the rabbis from the sixth century onward was no longer to establish a consensus but to clarify and preserve the *traditions* which had already been agreed in the academies, they are usually referred to as *Masoretes* (*traditionists* – the English form can equally be written as *Massoretes*). 'Pointed' texts (texts with all the vowels represented by dots) first came into existence in Tiberias in the sixth century, but this earliest form of pointing (sometimes called the proto-Palestinian niqqud) was complex and took different forms in different academies.

During the sixth and seventh centuries the Babylonian rabbis developed a simpler and more co-ordinated system, the *Niqqud Bavli*, with vowel points placed *over* the consonants, while those at Tiberias began to produce a different system which placed the vowels predominantly *under* the consonants. This *Tiberian Niqqud* eventually came to dominate and is the basis of what we now call the Masoretic pointing. By the eighth or ninth century the aims of Rabbis and Masoretes had apparently been achieved: there was full consensus among Rabbinic scholars on the text of all the Scriptures and its correct interpretation

The ninth century, however, saw the rise of several threats to the traditions of Judaism, threats which substantially changed the perceived role of the Masoretes. Although at this time Classical Greek literature was unknown in Europe (with perhaps the sole exception of the Irish monasteries) it had long been kept alive under the Caliphate. By the ninth century it was being circulated in Arabic translation. Jewish communities were becoming increasingly urbanised and Arabic was an everyday language of trade in towns and cities. Tiberias itself, which had fallen under Byzantine control, was taken by the Arabs in 636. Greek philosophical ideas, in particular Aristotelian, Platonic and neo-Platonist philosophy, became increasingly popular and in many cases eroded the ancient Hebrew traditions. Many Jews were attracted to Gnosticism, and works such as the *Sefer Yeçira* (the *Book of Creation* which claimed to be written by Abraham himself) came to be widely circulated, retelling the Biblical creation accounts in neo-Platonist or Gnostic terms. Among many of the new urban elite, Aristotelean ethics challenged traditional Jewish beliefs. A Persian Jew, Hivi al-Balkhi, who had 'converted' to Gnosticism even produced an expurgated text of Scripture from which, as from Bowdler's Shakespeare, all 'offensive' material had been removed!

A very different but equally problematic movement from the Rabbinic point of view was the fast growing Karaite sect. The office of *resh galuta* (exiliarch) in Babylon had developed into a hereditary one, but succession to it depended upon the assent of the Caliph. In the second half of the eighth century the ruling Caliph refused to confirm the succession of Anan ben David, choosing instead his younger brother. Anan's reaction was to set up a 'new religion' which, he claimed, was a compromise between Judaism and Islam. His *Sefer haMiçvot* (Book of Precepts) written in about 770 asserted the full authority of the Hebrew Bible, but rejected both the Talmud and the authority of the Rabbinate in the interpretation of scripture, which was held to be

self-evident to all readers. In stressing this principle his followers, at first called Ananites, adopted the name *Karaites* (properly *Qaraites* from the Hebrew verb *to read*). These 'Bible readers' coined the term *Rabbinites* for those who followed the authority of the Talmud. By the mid ninth century the Karaites, under the leadership of Daniel al-Qumisi, had spread from Babylon to Palestine and attracted a considerable following there.
Given the very real possibility that large numbers of Jews would read and interpret the Bible without reference to Rabbinic orthodoxy, one final task remained for the 'Rabbinite' Masoretes: to restrict the number of ways in which the Bible *could* be interpreted by imposing upon it a system of punctuation which was so precise as to be, in almost all cases, totally unambiguous.

The method they adopted was an ingenious blend of ancient tradition with new technology. We have noted that major divisions within verses had been established for at least a thousand years and in general these were followed by the Masoretes, although some were revised in the interest of greater consistence. A musical tradition for intoning the scriptures had also been established, although it is not known how long this had existed in a stable form. In following these two systems the Masoretes were indeed 'traditionists'. Their innovation was to complete the punctuation of verses in a rigorously mathematical way.

Although the main task of the Rabbis had always been the preservation and interpretation of scripture, many of them were also noted for their learning in other fields. This was particularly true of the Babylonian rabbis. As early as the third century CE Rabbi Samuel of Nehardea, who contributed substantially to the Babylonian Talmud, was also renowned for his knowledge of astronomy, medicine and civil law. He also introduced a rule which was to change the nature of rabbinic scholarship: that Jews in exile were subject to the civil laws of the state in which they lived.

By the end of the seventh century the highly advanced mathematics of India had spread to northern Mesopotamia where it was eagerly embraced by Islamic scholars. In the early ninth century the House of Wisdom was built in Baghdad, near Nehardea. This was on the one hand a centre for translation and preservation of ancient texts (particularly the major Greek mathematical writings) but it was also here that Islamic mathematics flowered. Perhaps its most influential scholar was al-Khwarizmi whose name gave us the word *algorithm* and one of whose works coined the word *algebra* (*al-jabr*, 'restoring'). Among many innovations he developed the new mathematics of decimal fractions and binary division to calculate the time when the new moon would be first visible and the direction of Mecca, thus binding Islamic religion and science inseparably. More significantly for our present purposes, al-Khwarizmi also established mathematical models for the very complex Muslim rules of inheritance, which were also binding on Jewish exiles in Islamic territory. This applied to all the Jewish academies since even Tiberias was now under Islamic control. Mathematics, to some degree at least, was now an essential tool for both Jewish and Muslim religious scholars.

There is no direct evidence that the Masoretes were trained in mathematics, but the probability seems substantial. The binary division and subdivision of the text of scripture rapidly became the focus of their punctuation, which combined traditional musical notation with a consistent binary analysis of the constituents of every verse of scripture. This task was brought to its spectacular conclusion in two different forms at Tiberias by Rabbi ben Naftali and Rabbi ben Asher in the first half of the tenth century. Modern printed texts follow almost exclusively the text of ben Asher.

It is a measure of the rigour of these rabbis' work that following the production of the Masoretic Text all manuscripts copied it exactly, and although the first printed Hebrew Bible was not produced until 500 years later it is extraordinarily faithful to the earliest extant manuscripts. In all significant respects the Bible which ben Asher and ben Naftali completed has remained the only accepted text of the Hebrew Bible for a thousand years.

b. Hebrew Orthography and the Punctuation System

1. Pre-Masoretic signs

The Hebrew alphabet consists of 23 consonantal signs, five of which take a different form when they are the last letter of a word:

א ב ג ד ה ו ז ח ט י כ ל מ נ ס ע פ צ ק ר שׁ שׂ ת
' b g d h v z h̲ t̲ y k l m n s̲ ‘ p ç q r s š t
Final: ך ם ן ף ץ

Verses are divided by a period or *Sof Pasuq* ׃

Words may be joined by a hyphen or *Maqqef* ־ so, for example, the last three words of Gen 1.2 are ...עַל־פְּנֵי הַמָּיִם׃

Several words may be joined in this way: in Gen 25.5 we read אֶת־כָּל־אֲשֶׁר־לוֹ: when words are joined by Maqqef they effectively become a single word, and all except the last one lose their accents (but they may retain a secondary tone).

The following table gives the position of these characters in the font SIL Ezra (two digit hex) and Unicode Standard 3.0 (four digit hex):

א	27	05D0	ו	77	05D5	כ	6B	05DB	ע	76	05E2	שׁ	57	05E9+	ן	A4	05DF
ב	62	05D1	ז	7A	05D6	ל	6C	05DC	פ	70	05E4	שׂ	48	05E9+	ף	A5	05E3
ג	67	05D2	ח	78	05D7	מ	6D	05DE	צ	63	05E6	ת	74	05EA	ץ	A6	05E5
ד	64	05D3	ט	58	05D8	נ	6E	05E0	ק	71	05E7	ך	A2	05DA	׃	3A	05C3
ה	68	05D4	י	79	05D9	ס	73	05E1	ר	72	05E8	ם	A3	05DD	־	2D	05BE

In Unicode the character שׁ is a composite of 05E9 + 05C2 and שׂ is 05E9 + 05C1. The neutral form ש (53 05E9) can also occur as in the name יִשָּׂשכָר *Issachar* (Gen 30.18 etc).

One character, the vertical bar | (7C 05C0) serves for both a pre-Masoretic and a Masoretic function. As a pre-Masoretic sign it is called *Paseq* and served many different functions. The most common are:

a. to prevent ambiguity in word division
b. to separate identical letters at the end and beginning of adjacent words
c. to separate words which are identical or very similar
d. to separate words which are contradictory in meaning (e.g. *God* and *sinner*)
e. to mark abbreviations.
Its use in the Masoretic accent system will be noted below.

2. Masoretic points and accents

2a. Vocalisation: Hebrew Vowels
The Niqqud dots which the Tiberian Masoretes used to mark vowels are generally written below the consonant they follow. In what follows they will be placed relative to the letter מ. It should be noted that there are several different ways of transliterating their names.

מַ

Patah (05B7) is a short *a*. Like most of the vowels it occupies several font positions to enable it to appear correctly beneath characters with different widths: 61, C2, D1 and DE.

מָ

Qameç (05B8, 41, 6F, C3, C4, D2, D3, DF, E0) is, depending on its context, either a long *a* or a short *o*.

מֶ

Segol (05B6, 65, C6, D5, E5) is a short *e*.

מֵ

Çere (05B5, 45, C5, D4, E3) is a long *e*.

מִ

Hireq (05B4, 69, C7, D6, E6) is a short *i*. (It is long if followed by י)

מֹ

Holem (05B9, 4F, C8, D7) is a form of long *o*. It is the only vowel to be written at the top of a consonant.

מֻ

Qibbuç (05BB, 75, C9, D8, E7) is a short *u*.

מְ

Šewa (05B0, CC, CD, D9, DA, E8, ED, F8, FC) is, depending on its context, a very light indeterminate vowel or a silent syllable divider.

Other forms of the sounded šewa are:

אֱ אֲ אֳ

These cannot stand under מ so here they appear under א. From right to left these are *Hatef Patah* (05B2, CF, DC, E1, F0), *Hatef Qameç* (05B3, DD, EC, F1, F3) and *Hatef Segol* (05B1, CE, DB, E9, EF).

In SIL Ezra (but not in Unicode) two vowels are made as composites of u and a niqqud:

וֹ

Holem Vav (F4) is a long *o*.

וּ

Šureq (FB) is a long *u*.

The following composite characters are also present in the font:
דֵ (A7) דֶ (A8) פֵ (A9) דִ (AA) דִ (AC) דָ (AD) ה֥ (CB) ה֫ (E4)ה֭ (EB) ה֮ (F6)

2b. Other non-punctuating Masoretic signs
Most letters of the Hebrew alphabet could be modified by inserting a dot called *dageš* in the middle as in בּ. This dot had three functions according to its context: in the letters בגדכפת it hardened the sound of the letter: without *dageš* the letter was aspirated, with *dageš* it was not. When used in this way it is called *dageš lene*. In other letters except ה it made the sound more forceful and drawn out, almost doubling the consonant. This is called *dageš forte*. In הּ (where it is called *mappiq*) it tells the reader to give full pronunciation to the *h* sound rather than just lengthening the vowel before it. The coding of *dageš* is 05BC, 43, 46, 47, 4A, 4B, 4C, B7, D0.

Mappiq is also occasionally used with א. When it is used it is positioned in the base of the letter as אּ. In *Biblia Hebraica Stuttgartensia* this occurs only four times: at Gen 43.26, Lev 23.17, Job 33.21 and Ezr 8.18. The font code for this usage is 49: it does not appear to be included in the Unicode 3.0 specifications.

Rafe is a small bar written above a consonant as כֿ. This sign was used extensively by the Masoretes as a check to show that *dageš lene* had been omitted deliberately, not by scribal error. It is rarely found in printed editions. Its codes are 05BF, 26.

The *Masora Circle* בׇּ (05AF, 7E) is the equivalent of a footnote marker, drawing the reader's attention to a variant reading or other note.

Meteg בְּ (05BD, 95, BD, BE, BF, C0, C1) is used to mark the secondary tone, reminding the reader to give its vowel full pronunciation. In *BHS* meteg is also used sometimes to help the reader distinguish a short *o* from a long *a*. An identical mark is used for the punctuation character *Silluq*. In fact several of the accents do 'double duty' in this way, but there is always a clear algorithm to distinguish their use.

2c. Masoretic punctuation signs

There are two distinct systems of punctuation in the Hebrew Bible. One is used only in the books of Job, Psalms and Proverbs. This system is sometimes referred to as *Poetic* punctuation, but the term is misleading insofar as this system is not used elsewhere in poetry. A better title is 'the accents of the three books'. The initial letters of *Job*, *Proverbs* and *Psalms* in Hebrew spell the word אמת (*emet* 'truth'), so the system is also known as 'the accents of emet'. Although the rest of the Hebrew Bible consists of thirty-six books in most Christian bibles (more-or-less following the Septuagint), it makes only twenty-one books in Hebrew: Samuel, Kings and Chronicles count as one book each, Ezra and Nehemiah form a single book, and the twelve prophets Hosea to Malachi are also counted as one book. So the other punctuation system is referred to as 'the accents of the twenty-one books'. The punctuation system of the 'three' is very much more complex than that of the 'twenty-one' and is, with one exception, not covered by this paper.

The rabbis divided the punctuation marks of each system into two distinct groups: the *melakhim* ('kings') and the *meshartim* ('servants' or 'ministers'). The Christian grammarians gave these the Latin names *Domini* and *Servi*. The 'Kings' are disjunctive accents which divide a sentence into phrases and the 'Servants' are conjunctive: they mark words which are components of phrases to show that they are connected with each other. Some of these marks are written in various forms in different editions of the Hebrew text. For simplicity we shall here use the same forms as *Biblia Hebraica Stuttgartensia*.

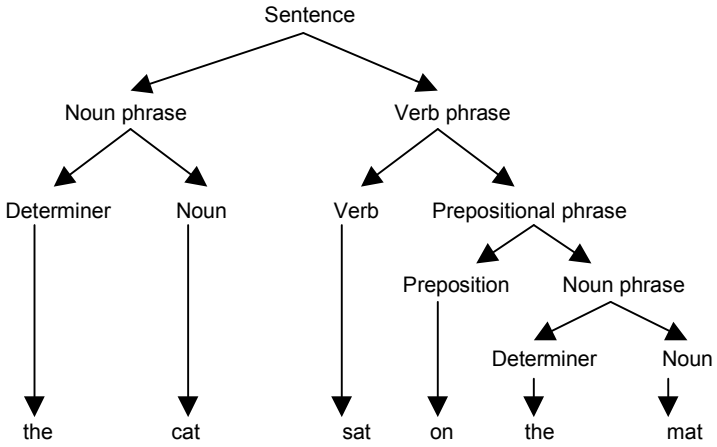
2c1. Disjunctive accents: the melakhim

Following the work of Bohlius in the Lutheran University at Rostock (*Scrutinium Sanct. Scr. ex accentibus*, 1636) it became fashionable among Christian hebraists to divide the 'kings' into a hierarchically structured 'ruling class' consisting of

Imperatores, Reges, Duces, Comites etc. Recent years have seen a resurgence of this approach among both Christian and Jewish scholars. While this system is very graphic and is a very useful teaching aid, it does not convey what actually happens in the text: Gesenius-Kautzsch describes it as “the source of manifold confusion”. In what follows the accents will be given their traditional Hebrew names.

Most of the punctuating accents serve three purposes: they indicate pitch and punctuation and they also mark the tone syllable of the word on which they stand. In some cases this last function is absent. A small number of accents are written after the end of the word on which they stand: these are called *postpositive*. Some others, called *prepositive*, are written before the beginning of the word.

The term ‘punctuation’ is perhaps confusing when applied to the Masoretic text. The system is in fact very similar to that used in marking constituent structure trees in linguistic analysis and can be illustrated by the same form of tree-diagram as:



The categories of the Masoretic system (except for Sentence) are completely different from those of phrase structure grammars, but the structure and many of the mathematical rules for manipulating it are remarkably similar. The structure and some of the rules are treated in a companion paper, *Masoretic Hebrew Cantillation and Constituent Structure Analysis.*

The main *melakhim* accents are here listed in approximately the descending order of their disjunctive strength, largely following Wickes[1].

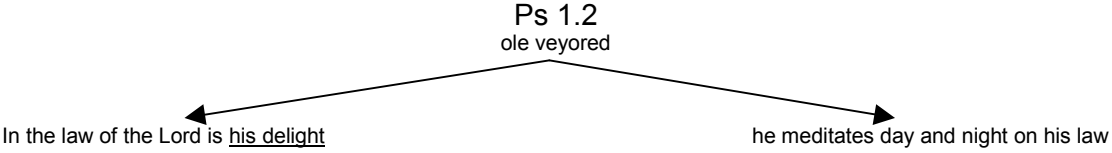
Silluq
Silluq is the strongest disjunctive accent, the equivalent of a modern full stop. It is written as a vertical bar under the tone syllable of the last word in a sentence (Gen 1.1 :הָאָרֶץ where א is the accented letter). In appearance it is exactly the same as meteg. In the vast majority of cases, silluq is written under the word immediately

[1] W. Wickes, *Ta'amey 21 sefarim (A Treatise on the Accentuation of the twenty-one so-called Prose Books of the Old Testament),* Oxford 1887

before sof passuq (׃) so it is usually redundant as a punctuation mark. But the Masoretes made good use of it in a few cases where they disagreed with the sentence divisions they had inherited from earlier rabbis. In Gen 35.22, for example, the end of the verse is doubly accented. The earlier rabbis had not placed a sof passuq between "and Israel heard it" and "the sons of Jacob were twelve", although the structure of the narrative clearly requires one – it seems likely that this was a rather delicate means of passing over an unpleasant subject by minimising its emphasis. The Masoretes were not free to insert a sof passuq, and they obediently pointed the text in the form they had received it, but *also* inserted silluq at the end of "and Israel heard it" to indicate that there *should have been* a verse division at that point. Similar emendations of the traditional verse structure are to be found in Ex 20.2ff and Deut 5.6ff. With these exceptions silluq is always the *last* accent on a word. Any mark which appears before it is to be ignored for the purposes of punctuation.

Ole veyored

This accent (strictly a pair of accents *ole* ֫ and *yored* ֥) is only used in *the three books*, but is dealt with here because it has a disjunctive force second only to silluq. It is written as two signs as in רִשְׁעִים (Ps 1.1), חֶפְצוֹ (Ps 1.2). The stroke yored עִ֥ indicates the beginning of the tone syllable. The accent is written on the last word (underlined below in English) of the first logical half of the sentence:

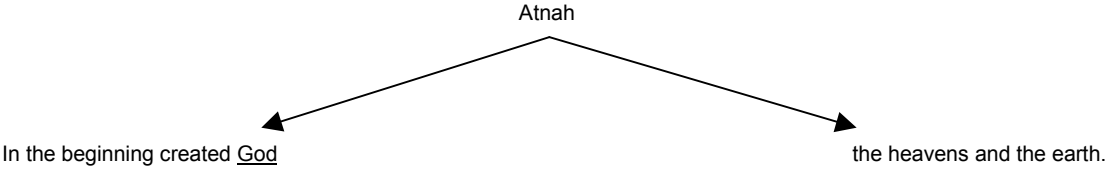


If the word carrying yored is accented on its first syllable, the ole is written on the beginning of the final syllable of the preceding word אֲמַלֵּל אֲנִי (Ps 6.3).

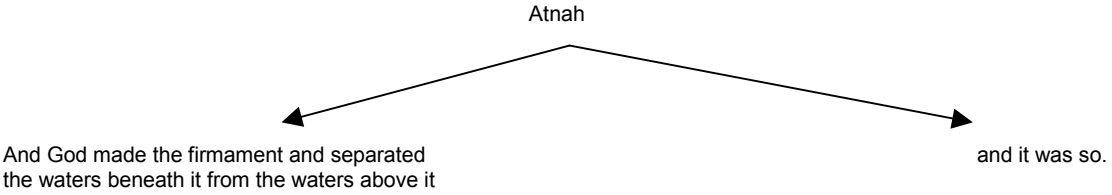
The Unicode reference for ole is 05AB, and in SIL Ezra its codes are 86 and BA. Yored shares 05A5, 42, 9C, F5 with merekha (see *Conjunctive accents* below).

Atnah

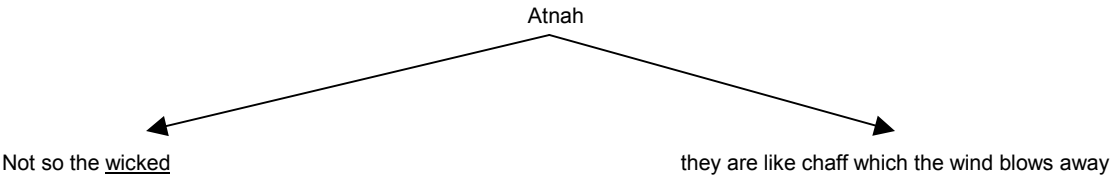
Atnah (also called Etnahta) is the main verse divider for the *twenty-one* books as ole veyored is for the *three*. It is written under the first consonant of the tone syllable as in אֱלֹהִים. It divides a verse into two logical halves as in Gen 1.1:



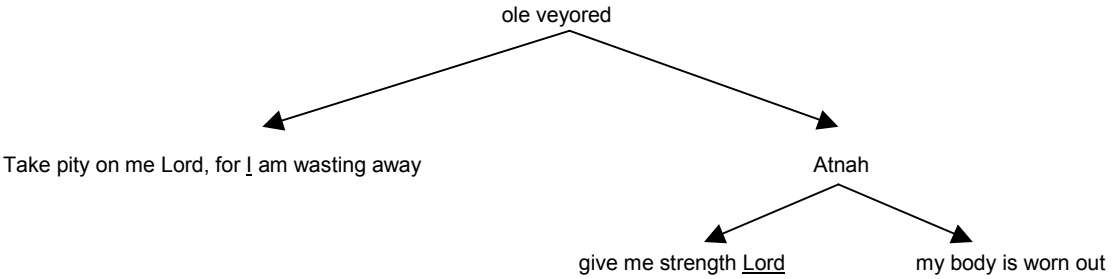
but these logical halves may be quite different in length (Gen 1.7):



In *the three books* Atnah can be the main divider in comparatively short verses (e.g. Ps 1.4):



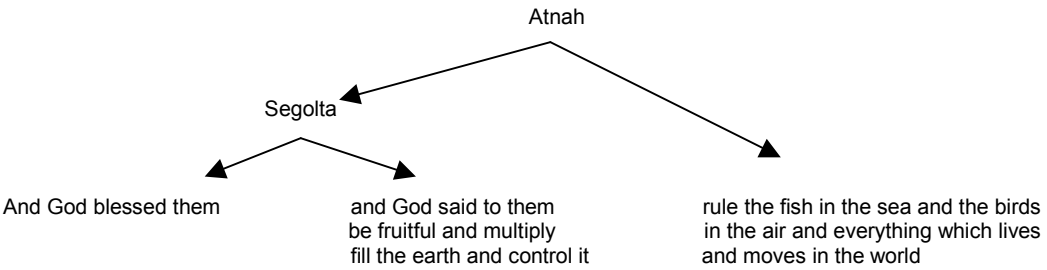
In verses divided by ole veyored, atnah is used, where necessary, to subdivide the second half of the verse (e.g. Ps 6.3 English versions 6.2):



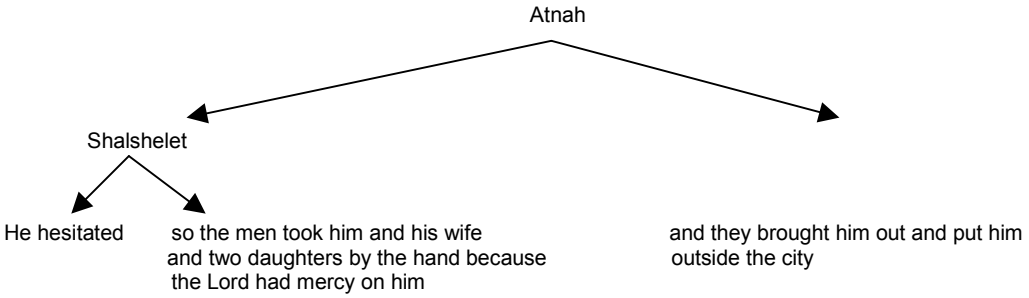
The Unicode position of Atnah is 0591, and in SIL Ezra it is A1, F2 and FD.

Segolta & Shalshelet

Segolta is used in the *twenty-one books* only as the primary subdivider of the first half-verse (the section ending with Atnah). It is written above and after (to the left of) the word which carries it: אֱלֹהִים (it is *post-positive*), so unlike most punctuation marks it does not give any indication of the tone syllable. Gen 1.28



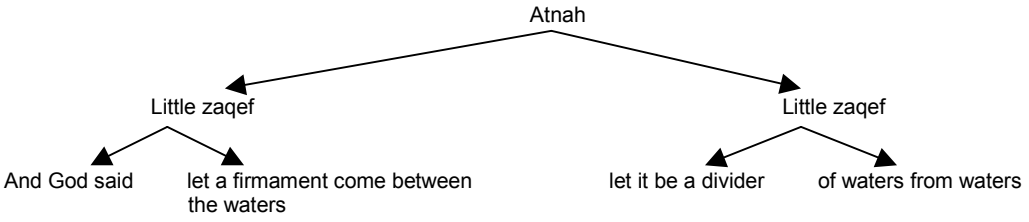
Shalshelet is a rare variant of Segolta: it is musically, but not grammatically distinct, and occurs seven times where segolta would appear on the very first word of the verse. It is made up from two signs, the vertical bar | used for Paseq (b1 above) and the symbol ֒. So the first word of Gen 19.16 is punctuated |וַיִּתְמַהְמֵּ֒ה. The verse is divided as:



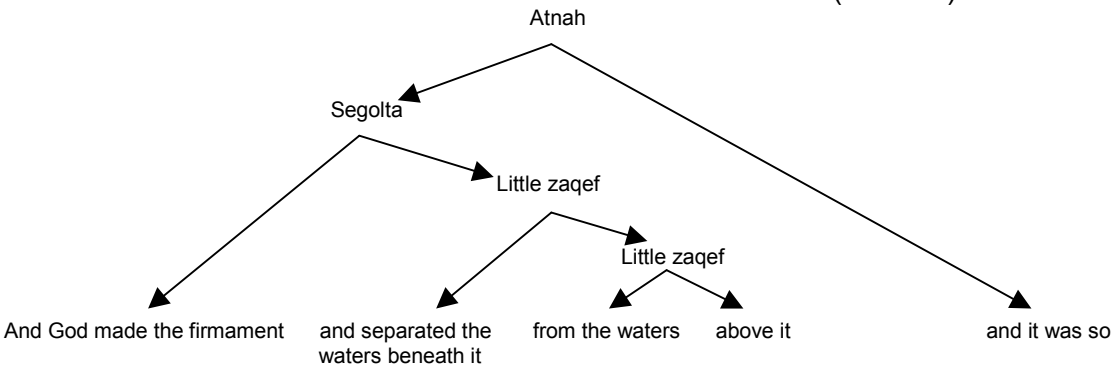
Atnah is said to *dominate* segolta or shalsheleth (the term is used by the rabbis as well as in modern mathematical descriptions of constituent structures). In structural terms one node dominates another if there is a path from the first to the second which only follows the down-ward direction of the arrows.

Segolta is coded as 0592, 82 and B1. Shalshelet is 0593, 52 and 8E.

Zaqef
Zaqef has two forms, 'Little Zaqef' and 'Great Zaqeph'.
Little zaqef is written as a small 'colon' above the beginning of the tone syllable: אֱלֹהִ֑ים. It may divide either or both halves of a verse (Gen 1.6):

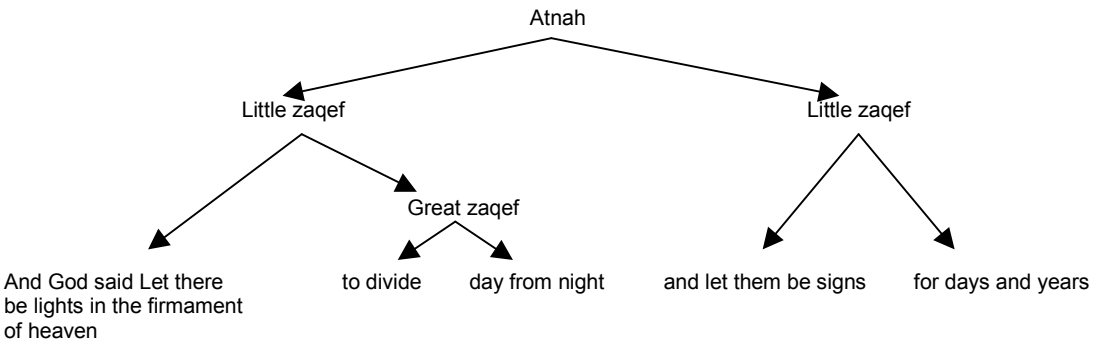


Unlike the punctuation dealt with so far, little zaqef can divide a phrase into more than two sections. in such cases the first dominates the second (Gen 1.7):



Here segolta dominates both zaqefs. It *immediately* dominates the first (is the first node up from it) and the first zaqef immediately dominates the second.

Great zaqef, in spite of its name, is generally a less forceful disjunctive than little zaqef (the names refer to musical, not grammatical value), and is frequently used to mark a sub-phrase which only consists of a single word. It is written as little zaqef followed immediately by a small bar: אֱלֹהִים. Gen 1.14 is punctuated:

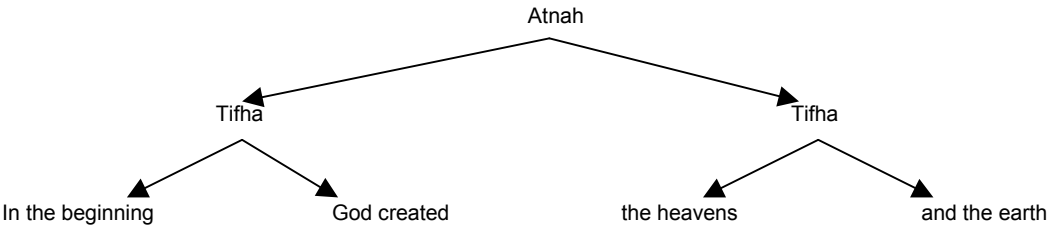


Little zaqef is coded as 0594, 81, AF and great zaqef as 0595, 44, 88. To simplify tree diagrams we will from now on use the term Zaqef to mean Little zaqef and only specify when it is Great zaqef.

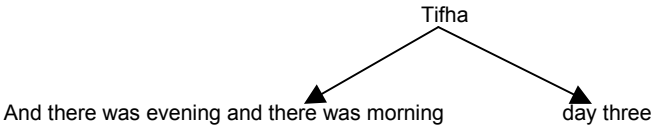
Tifha

Tifha (also called Tarha) marks the disjunction immediately before silluq or atnah. In the twenty-one books it is written below the beginning of the tone syllable: בְּרֵאשִׁית.

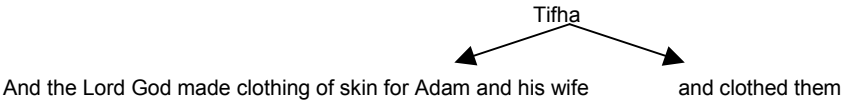
The full disjunctive pattern of Gen 1.1 is:



In short verses which do not need any strong disjunction Tifha is often the main punctuation mark. An example is Gen 1.13



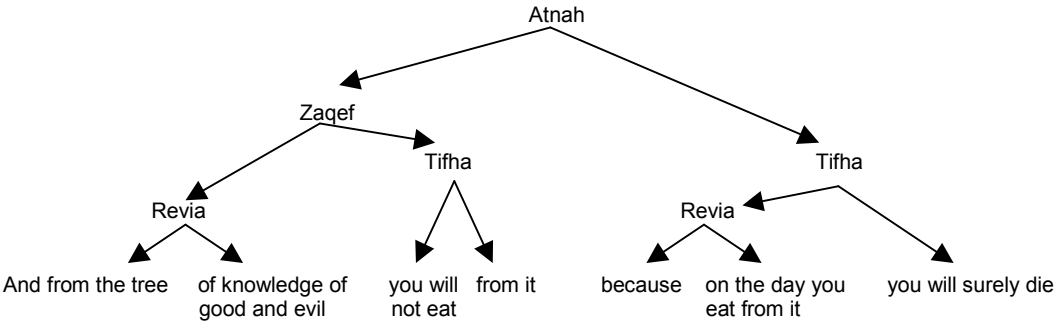
The same is sometimes true of longer verses where no strong disjunction is appropriate, e.g. Gen 3.21:



Tifha is coded as 0596, 59, 9B, F7.

Revia

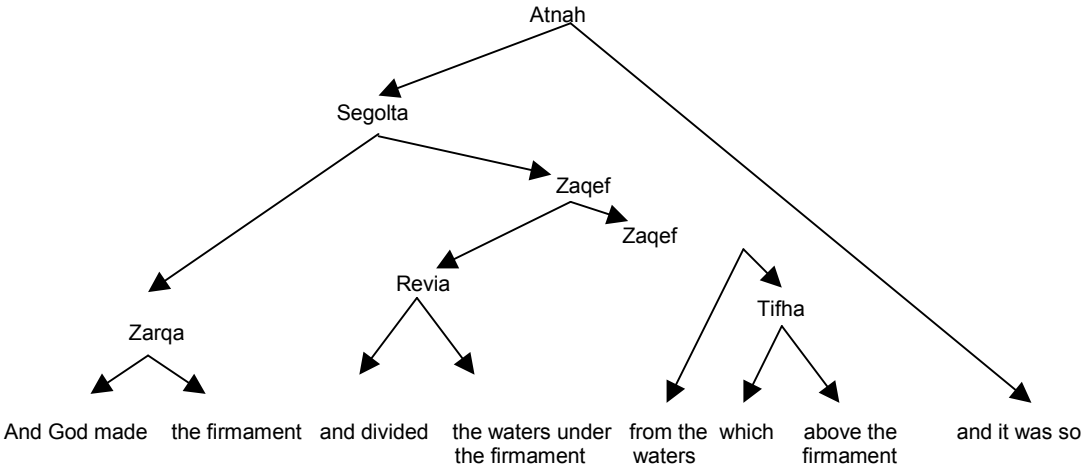
Revia is a single dot placed over the beginning of the tone syllable as in וְהָאָרֶץ. It can be distinguished from the vowel holem as being higher and larger. It is often printed as a small diamond while holem is circular. Gen 2.17:



Revia is coded as 0597, 80, AE, B2.

Zarqa

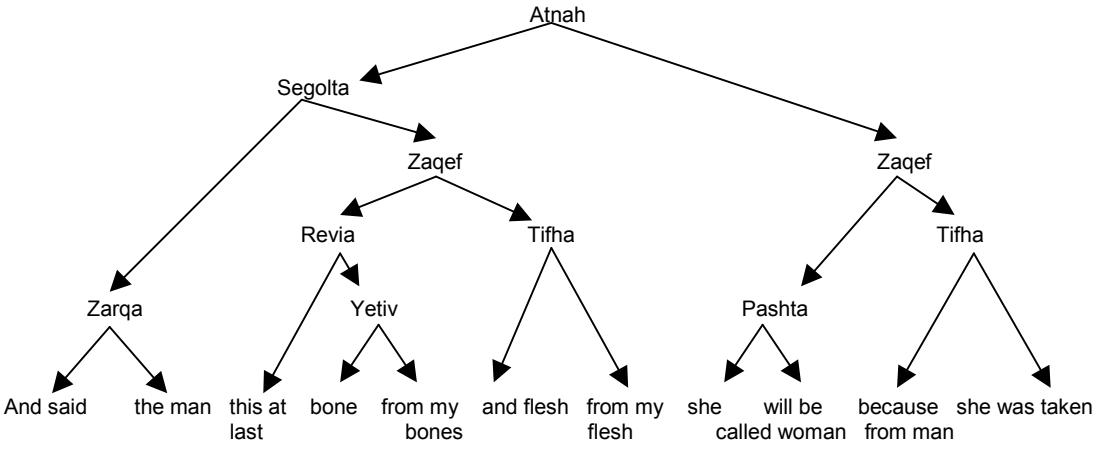
Zarqa is postpositive, written above and to the left of the final letter of the word as in אֱלֹהִים. Under *zaqef* above we have given a partial analysis of Gen 1.7. Here is a fuller (but still incomplete) one:



Zarqa is coded as 0598, 51, 56, 8D, 90.

Pashta & Yetiv

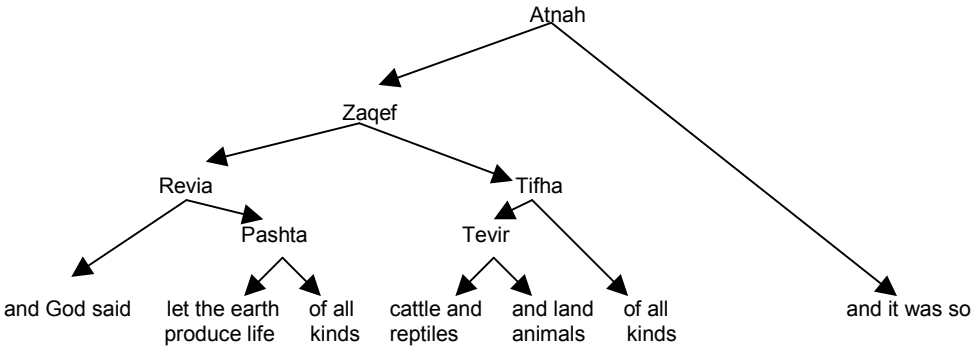
Pashta is a postpositive accent which stands over the last consonant of its word: לְאוֹר. Unlike other postpositive accents it *does* give an indication of where the tone syllable occurs, because if the tone is not on the last syllable pashta is duplicated over the tone syllable as in תּוֹהוּ. Since it is only written once on לְאוֹר it follows that this word must be accented on the final syllable. A musically distinct but grammatically equivalent variant of pashta is yetiv which is prepositive: עָשׂ֫ב. Gen 2.23 contains an example of both pashta and yetiv:



Pashta is coded as 0599, 3D, 66, 89, 8F. Yetiv is 059A, 60, 98.
In modern mathematics (but *not* in the rabbis) it is customary to refer to a node as the *daughter* of the node which immediately dominates it (which is then called its *mother*). So in the above tree *and said* is a daughter of zarqa, which is in turn a daughter of segolta. Since *and said* and *the man* are both daughters of zarqa they are referred to as *sisters*. There is a simple rule for *daughter association*: any daughter phrase may be directly associated only with the complete set of daughter phrases which are also dominated by its mother[2]. So *bone* can only be associated with *from my bones* through yetiv. After this association the phrase *bone from my bones* can only be associated with *this at last* through revia, and then the whole phrase is associated with *and flesh from my flesh* through tifha and so on. This rule is specified in greater detail in the accompanying paper *Masoretic Hebrew Cantillation and Constituent Structure Analysis.*

Tevir

Tevir is written under the tone syllable and always after its vowel as in פְּרִי. Gen 1.24 is punctuated:



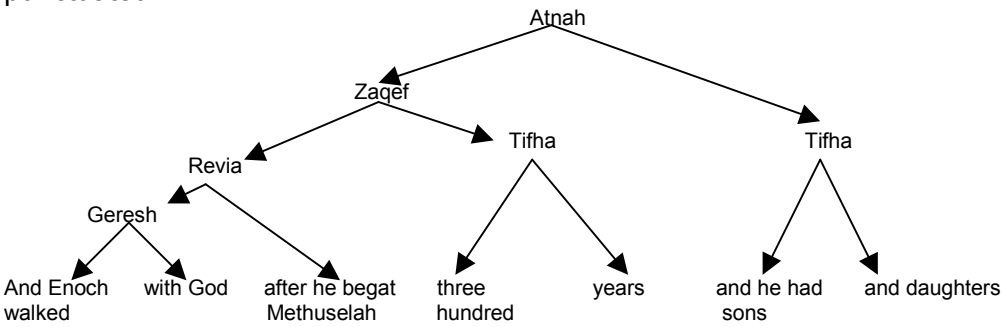
We may see from this how the punctuation is used to avoid a possible ambiguity: given the words alone it would be possible to group part of this verse as {cattle and reptiles}{and land animals of all kinds}. But, given the rule of daughter association, *of all kinds* can only be associated with the whole phrase *cattle and reptiles and land animals*.

Tevir is coded as 059B, 5F, 99.

[2] For further discussion of this see *Masoretic Hebrew Cantillation and Constituent Structure Analysis* pp 11f.

Geresh (Teres) & Gerashayim

Geresh, also called Teres, is written above the tone syllable as in הִים. Gen 5.22 is punctuated:



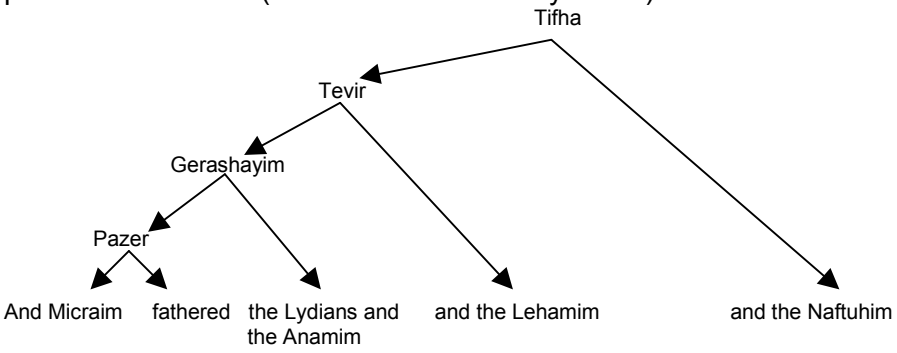
Again we may notice the disambiguation here: this passage is a genealogy of the male line of descent, and is not concerned with political correctness. The second tifha clearly forces the interpretation {and he had sons}{(and daughters as well)} rather than {and he had}{sons and daughters}.

A variant form of Geresh is Gerashayim, 'double geresh' as in Gen 5.5 ויהי״ו. As punctuation marks the two are equivalent.

Geresh is coded as 059C, 4D, 8A and gerashayim as 059E, 55, 8B.

Pazer

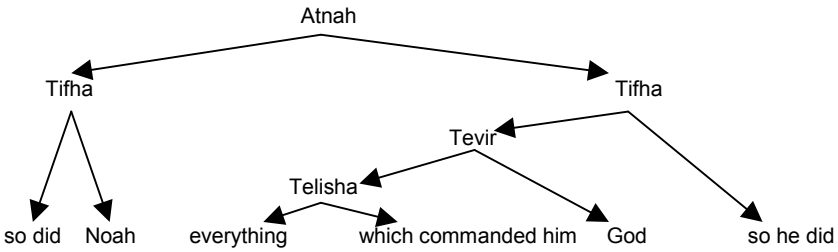
Pazer is written in several different ways in different editions of the Hebrew bible. Its form in *BHS* is ֫לְ. A rare variant is Great Pazer, written as ֬ן. This form is also known as Qarne fara, 'cow-horns' and occurs only sixteen times. A good example of pazer is Gen 10.13 (which is not divided by atnah):



Pazer is coded as 05A1, 87, BC and Great Pazer as 059F, 85, B9.

Telisha

Telisha has two forms: 'Great Telisha' which is disjunctive and 'Little Telisha' which is conjunctive and will be discussed below. Great telisha is prepositive and written as דֽשׁא. Gen 6.22 is punctuated:

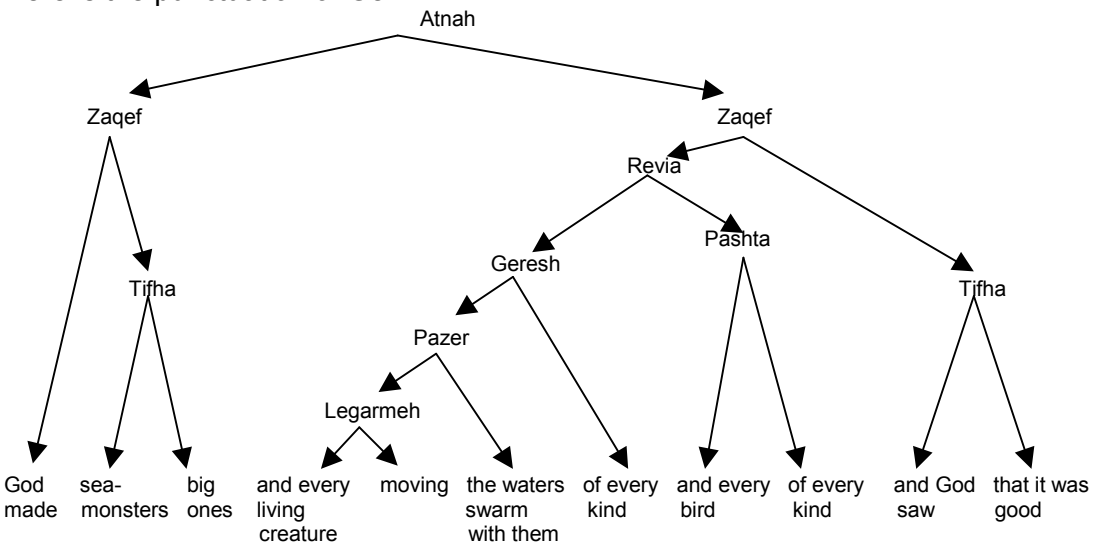


Great telisha is coded as 05A0, 84, B8

Legarmeh

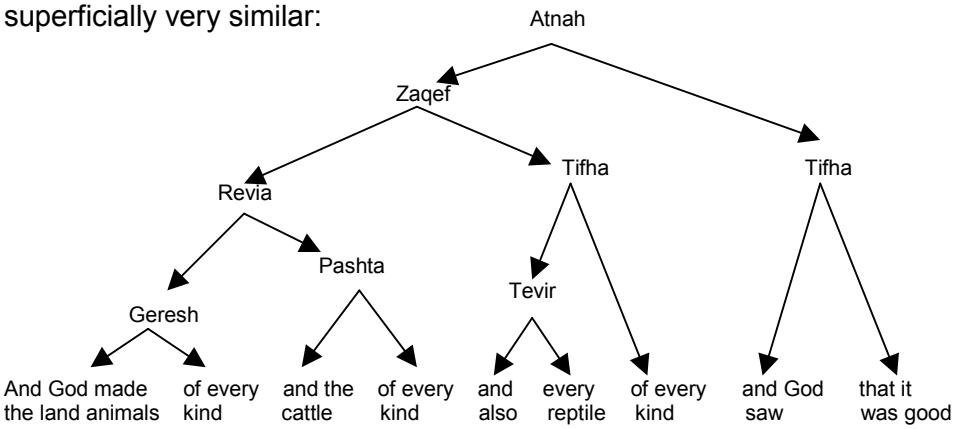
Like Shalshelet, Legarmeh is formed from both an accent and a following bar: ׀ הַחַיָּה. Also like Shalshelet, if the bar is missing the accent is conjunctive. It does not have its own coding in Unicode but is a composite of 05A3 and 05C0. In SIL Ezra it is a composite of 7C and the accents for Munah (for which see below).

Here is the punctuation of Gen 1.21:



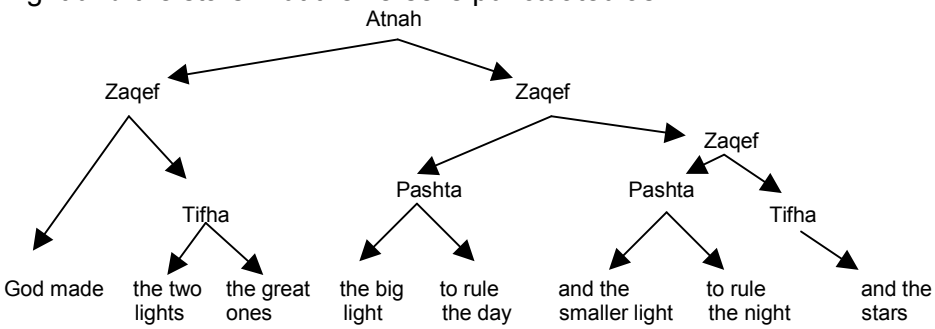
The punctuation here is subtle. In Isa 27.1 we read that the Lord will punish Leviathan and slay the dragon that lives in the sea. But these are clearly creatures which the Lord himself created. In the Masoretic punctuation the daughter association of 'and God saw that it was good' only extends backward to nodes under the second zaqef, thus avoiding the implication that the sea-monsters were good!

We may contrast the punctuation of 1.25 where the unpunctuated verse is superficially very similar:

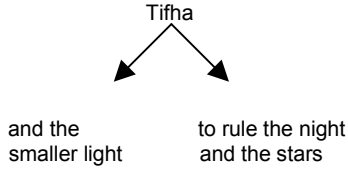


Here *land animals* and *cattle* are distinguished, but quite closely related to each other, while *reptiles* belong to a different category (this is further stressed by the tevir on *and also*). Nevertheless, it is true for all subgroups that they were seen to be good since the daughter association of *God saw that it was good* covers the whole verse.

Before turning our attention to the conjunctive accents it may be useful to consider some cases where use of the disjunctive accents can affect the way in which verses are translated. Without punctuation Gen 1.16 reads: 'God made the two great lights the big light to rule the day and the smaller light to rule the night and the stars'. It would seem quite natural to take this as meaning that the smaller light rules *both* the night *and* the stars. But the verse is punctuated as:

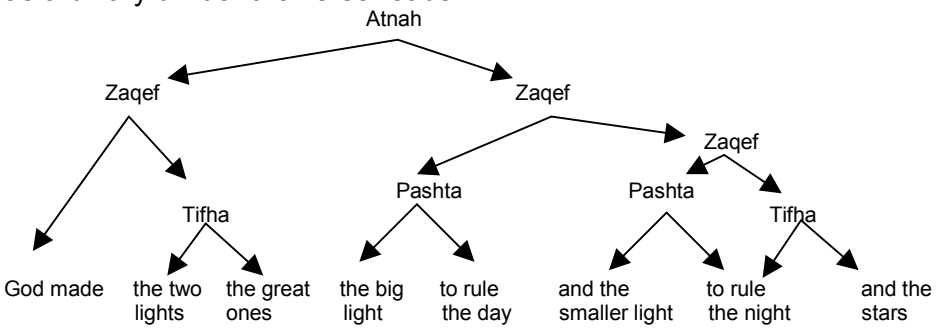


If the smaller light ruled both the night and the stars, the punctuation at the end of the verse would have to be



The actual schema forces us to divide the final section as {and the smaller light to rule the night} {and the stars} rather than {and the smaller light}{to rule the night and the stars}. Thus the King James Version translates '...the lesser light to rule the night:

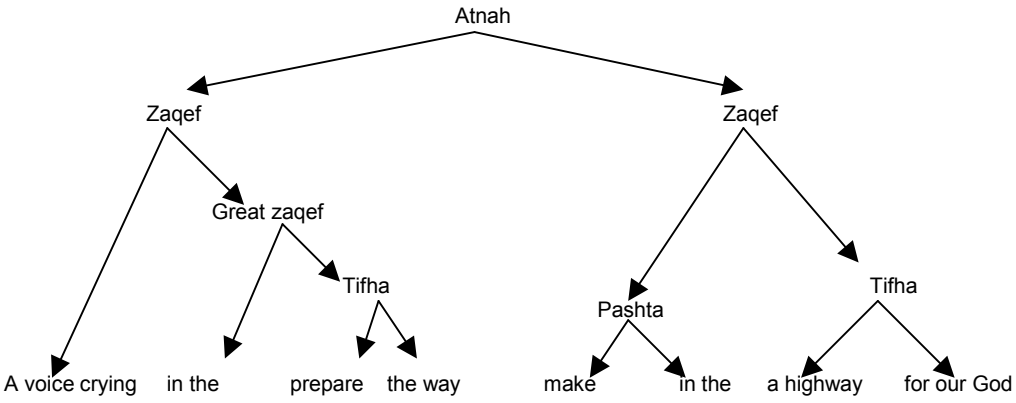
he made the stars also'. Most modern translations give a similar rendering. If this is correct, then ben Asher's punctuation is less than perfect: 'and the stars' has no possible direct daughter association with either 'God made' or with 'God made the two great lights'. But if we accept the punctuation as it stands there is a further possibility: we have noted that while the Masoretes were scrupulous in preserving the text they had inherited, they were also prepared to use punctuation to challenge what they considered to be errors[3]. If the final Tifha of the verse is given its regular status as a binary divider the verse reads:



In this case the rabbis are telling their readers that 'to rule the night' is actually to be understood *twice* in this verse: 'the smaller light to rule the night and the stars to rule the night'. It was a constant rabbinic concern to show that the Hebrew Scriptures were self-consistent and without contradiction. Read in this way, Gen 1.16 may be reconciled with Ps 136 7-9, and in particular vs 9: 'the moon and stars to rule the night'.[4] On any other reading the two passages are in conflict. More interestingly for our present purposes, the 'correctional' use of punctuation is then fully consistent with the general scheme of binary division.

Isa 40.3 is also of interest both to Jewish and Christian scholars. The King James Version translates: 'The voice of him that crieth in the wilderness, Prepare ye the way of the Lord, make straight in the desert a highway for our God'.

The Masoretic text punctuates the verse as:

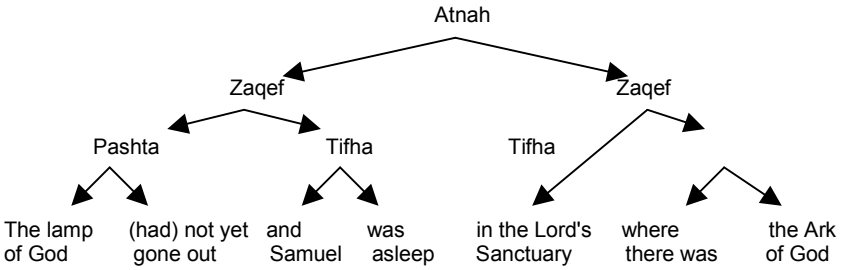


[3] v supra pp 11f. (under Silluq)
[4] cf Jer 31.34 (English versions 31.35)

wilderness of the Lord straight desert

By the rule of daughter association, ‘in the wilderness’ must first be associated with ‘prepare the way of the Lord’ before that whole phrase can be extended to include ‘A voice crying’. So the verse should be translated something like: A voice crying ‘Prepare the way of the Lord in the wilderness...’ Of the four New Testament citations of part of this verse as a self-description by John the Baptist, only that of the Fourth Gospel (Jn 1.23) is compatible with this analysis (although few translations indicate this possibility).

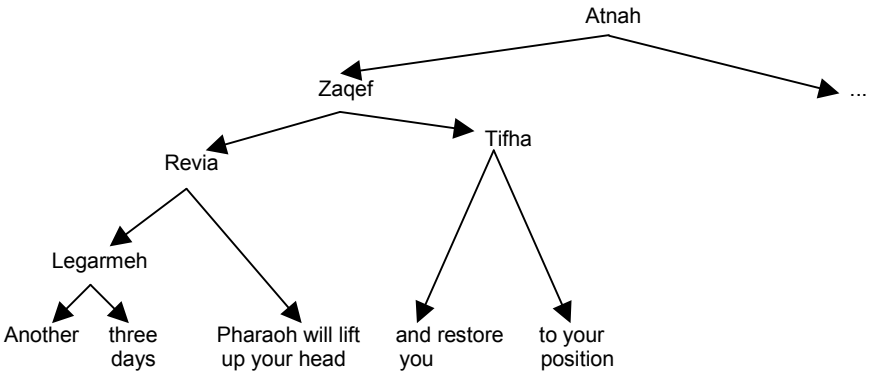
Many modern translations ignore the punctuation of 1 Sam 3.3. Its schema is



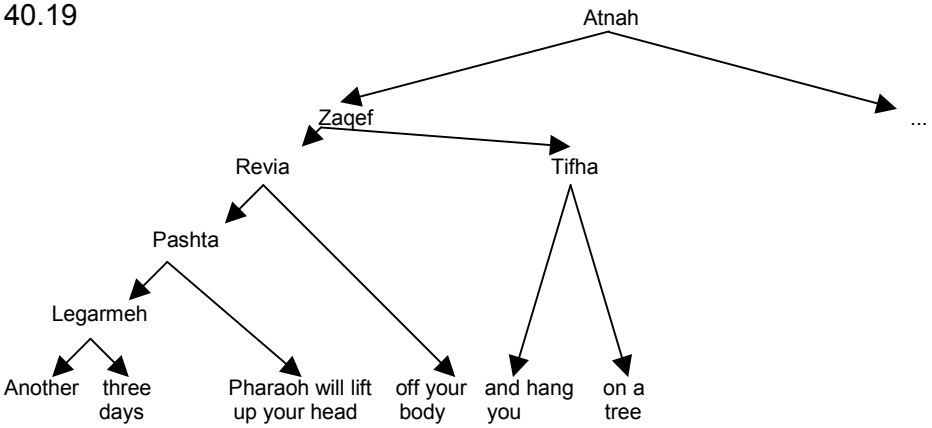
This means that ‘Samuel was asleep in the Lord’s Sanctuary’ is *not* a valid division of the verse. What the rabbis are stressing is that Samuel was quite naturally asleep because it was not yet dawn.

We give finally a nice use of punctuation which uses daughter insertion to wryly humorous effect. Compare Gen 40.13 with 40.19 (we give only the first part verse for each):

40.13



40.19



2c2. Conjunctive accents: the meshartim

The conjunctive accents serve to bind words together in much the same way as maqqef, but less strongly. The rules governing which conjunctive is used to bind a word to each of the disjunctive accents are complex, but the distinctions are of an entirely musical nature, so they do not affect the present paper and will not be dealt with in any detail.

Munah
Munah is written under the tone syllable of a word as בָרָא. It is coded as 05A3, 5A, 9E, F9. It is used, for example, in Gen 1.2 to bind *the-Spirit-of* with *God*: וְרוּחַ אֱלֹהִים.

Mehuppakh
Mehuppakh is written under the tone syllable as אֱלֹהִים. It is coded as 05A4, 60, 98. Mehuppakh is often used to bind a word to another which carries *pashta* as in Gen 1.14: יְהִי מְאֹרֹת joining *let-there-be* and *lights*.

Merekha
Merekha is written under the tone syllable as אֶת. It is coded as 05A5, 42, 9C, F5 (the same codes as are used for *yored*). It is often used to bind a word to one which carries *tifha* as in Gen 1.3 וַיֹּאמֶר אֱלֹהִים *said God*. A variant form is *double merekha* or *merekha khefula* as in Gen 27.25: לֹו יָיִן joining *wine* and *for him*. *Merekha khefula* is coded as 05A6, 54, 9D.

Darga
Darga is written under the tone syllable as in וַיִּקְרָא. It is coded as 05A7, 5E, 9F, FA.

Azla & Qadma

Azla is written above the tone syllable as in וַיִּקְרָא. It is coded as 05A8, 3D, 66, 89, 8F. In appearance it is identical to *pashta*, but can readily be distinguished because the latter is written after its word (and sometimes *also* on the tone syllable). *Azla* is often associated with the disjunctive *geresh* as in Gen 1.9: יִקָּווּ הַמַּיִם *let the waters come together*. In such cases it is known as *Qadma*.

Little Telisha
Little Telisha is written above and after its word (i.e. postpositively) as in אֹשֶׁר. It is coded as 05A9, 83, B6.

Galgal
Galgal is written under the tone syllable as in וַיְהוֹשֻׁעַ. It is coded as 05AA, 6A, 9A. It is very rarely used (16 times in all, always before *great pazer*).

Several conjunctive accents can occur consecutively, with or without the hyphen *maqqef*, to indicate that all the words so connected are to be treated as a single unit. In Ezek 48.21 the instructions for tribal allotments have been dealing thus far with two *different* areas, the sacred Temple-area and the secular city-area (separated by disjunctive accents). Now, however, the area belonging to the prince is distinguished from these areas *as a totality*, so in this verse they are joined by conjunctives as: *the-holy* <maqqef> *area* <munah> *and-the-property* <galgal> *of-the-city*. In Josh 19.51 three parties draw lots at Shiloh to divide the land: the first is *Eleazar* <munah> *the-priest*, the second is *Joshua* <galgal> *son-of* <*maqqef*> *Nun* and the third is *the-heads-of* <munah> *the-fathers-of* <munah> *the-tribes-of* <little telisha> *the-sons-of* <maqqef> *Israel*. In this case five Hebrew words are involved, but they are marked as forming a single semantic unit.

The Masoretic pointing as a whole, and the punctuation in particular, is arguably one of the greatest literary and linguistic achievments in history. Its development spanned more than a thousand years and was only possible through the co-operation of countless forgotten scholars whose dedication to accuracy was without parallel. It offers to all the 'people of the Book' a detailed explanation of how the great biblical teachers understood their sacred text. For all its antiquity, it is completely amenable to computer analysis and its mathematical and semantic structure is available to all who study or promulgate the scriptures.

David Robinson — BFBS MAT Team
Elisabeth Levy — Bible Society in Israel
February 2002

Select Bibliography

The classic account of the accentuation system was given by
Rabbi David Qimhi (died ca 1235)
Sefer Miklol (The Book of completeness).
This was translated into English by William Chamaky 1933.

The earliest non-rabbinic treatment was by
S. Bohlius
Scrutinium S.S. ex accentibus 1636

The most comprehensive study in modern times is
W. Wickes
Accents of the Poetic Books *Ta'amey emet* 1881
Accents of the Twenty-one Books *Ta'amey 21 sefarim* 1887

A recent study which takes a very different approach from this paper is
M. Broyer
Ta'amey hammiqra be-21 sefarim uvesifrey emet 1981
(Biblical accentuation in the twenty-one books and the poetic books)

Brief surveys of the system are hard to find. Probably the best is in
W. Gesenius ed. E. Kautzsch tr. A.E. Cowley
Hebrew Grammar 1813, 28th edition 1910

We understand that a computer-based concordance to the cantillation, *Tanach Cantillation Concordance*, is now available, published by Davka 1996, but have not been able to consult it.